Universidad Tecnológica Nacional
Facultad Regional Santa Fe

Doctorado en Ingeniería
Mención en Sistemas de Información

Tesis Doctoral

Una Ontología del Correo Electrónico y su Trazabilidad como Soporte para la Forensia Digital

Parra de Gallo, Herminia Beatriz

Directora: Vegetti, Marcela
Co-Director: Leone, Horacio

Santa Fe, Argentina
Año 2019

La presente publicación corresponde a una tesis presentada para cumplir con los requisitos exigidos por la Universidad Tecnológica Nacional Facultad Regional Santa Fe, para obtener el grado académico de Doctor en Ingeniería con Mención en Sistemas de Información.

Tesis Doctoral

Una Ontología del Correo Electrónico y su Trazabilidad como Soporte para la Forensia Digital

Parra de Gallo, Herminia Beatriz

Directora: Vegetti, Marcela
Co-Director: Leone, Horacio

Jurado de Tesis
Dr Alvaro Ortigosa Juarez
Dr. Pablo Becker
Dra. Mariela Rico

Diciembre, 2019

Parra, Herminia Beatriz
Una ontología del correo electrónico y su trazabilidad como soporte para la forensia digital / Herminia Beatriz Parra. - 1a ed . - Salta : Herminia Beatriz Parra, 2020.
Libro digital, PDF

Archivo Digital: descarga
ISBN 978-987-86-3560-6

1. Correo Electrónico. 2. Auditoría Forense. 3. Seguridad Informática. I. Título.
CDD 004.692

Agradecimientos

Este trabajo no hubiera sido posible sin la ayuda de un conjunto de personas que con su aporte me permitieron llegar hasta esta instancia.

En primer lugar la Dra. Marcela Vegetti, a quien le debo horas de dedicación puntillosa y acertada. Su experiencia, paciencia y disposición siempre alerta para corregir, sugerir, aportar y orientar la investigación fue destacable. El Dr. Horacio Leone también colaboró en los lineamientos generales, abriendo camino en los primeros momentos en donde quedaba muy lejos la tesis. Los investigadores del INGAR con quienes pude compartir algunos espacios, mostraron su generosidad, particularmente la Dra. Soledad Sonzini y el Ing. Álvaro Fraga.

El Consejo de Investigaciones de la UCASAL puso su confianza en mi proyecto, y no hubiera sido posible avanzar sin el aporte académico de esta institución, que en la práctica se tradujo en el otorgamiento de fondos para cumplir con los requerimientos de cursos y publicaciones exigidos. Mi ámbito natural, la Facultad de Ingeniería de la UCASAL, estuvo acompañándome en todo momento, desde la figura de su decano, el Mg. Ing. Néstor Lesser, hasta la de graduados destacados como el Ing. Enzo Notario, a quien particularmente debo agradecer su capacidad y predisposición para poner en valor mis ideas. También debo agradecer a los integrantes del Grupo de Investigación en Forensia Digital de la Facultad de Ingeniería, que me ayudaron en las discusiones técnicas sobre el análisis forense de correos electrónicos y a los colegas peritos que colaboraron en la prueba de la aplicación prototipo.

A todos ellos muchísimas gracias.

Dedicatoria

Este trabajo es para Rafael, mi compañero de toda la vida, para Pablo y Cristian que me siguieron día a día en este proyecto. Y para Bruno, Fiorella, Julieta y Benjamín para que sepan que cuando uno quiere…se puede.

Resumen

En esta Tesis se estudió la aplicación de ontologías a la Forensia Digital, tomando el caso del correo electrónico, y se formuló una ontología que permite incorporar al análisis forense de correos electrónicos la condición de ***evidencia digital no repudiable***. Así, se desarrolló ***OntoFoCE*** una ontología que representa el proceso de transmisión del correo electrónico permitiendo derivar su trazabilidad y respondiendo a 21 preguntas de competencia relativas a los puntos de pericia habituales. Y a partir de OntoFoCE se construyó ***ObE Forensic*** una aplicación web destinada al análisis forense de correos electrónicos.

Se toma como insumo el archivo de texto plano que contiene la cabecera del correo electrónico en análisis, y se identifican y representan todas las direcciones IP y nombres de dominio que se encuentran en dicha cabecera en lo que se denominó *ocurrencias* del correo (sucesivas copias del correo que se almacenan en los equipos/servidores que intervienen en la transmisión).

Para el proceso de validación de OntoFoCE se consideró una metodología integrada, que incluye métricas y herramientas semiautomáticas. Por otra parte, ObE Forensics se validó mediante cuatro escenarios diferentes de casos periciales, más la ayuda de usuarios expertos que trabajaron el prototipo según un protocolo de experimentación pre-definido.

PRÓLOGO

La Forensia Digital se ocupa del análisis de los componentes digitales que se aportan como prueba en una causa judicial. La mayoría de las veces, el objetivo que se persigue con el análisis forense es comprobar la existencia de la prueba digital, que al probarse, valida su contenido.

Con la utilización de las TIC (Tecnologías de la Informática y de las Comunicaciones) en el ámbito del delito, se hizo necesaria la incorporación de profesionales informáticos en el área de la justicia, en carácter de "peritos", es decir, auxiliares de la justicia en áreas o temáticas que no son de competencia directa del derecho. De este modo, a partir de los años 90, los informáticos comenzaron a actuar como peritos, compartiendo un espacio de trabajo con profesionales del derecho y de la criminalística, con quienes deben comunicarse, interpretar y hablar en un "lenguaje común" para todos, dejando de lado la terminología técnica propia de la informática.

En el caso particular de los correos electrónicos, el análisis forense consiste en tomar la cabecera del mismo, y obtener de allí todos los datos necesarios para establecer cual fue el camino utilizado para enviar el correo desde una cuenta emisor a una cuenta receptor. Si se logra establecer la secuencia de equipos utilizados en el proceso de transmisión, se puede validar la autenticidad del correo, sustentando así el carácter de no repudio de la evidencia digital (es decir, no puede ser negada la existencia del correo electrónico).

Durante el proceso de transmisión, el correo electrónico parte de un equipo emisor y viaja de servidor en servidor hasta que llega al equipo receptor, y los sistemas de gestión interna de dicho proceso van almacenando en la cabecera del correo los datos de identificación de cada equipo por el que pasó el correo. Al realizar una pericia, el perito

toma la cabecera del correo electrónico y la revisa identificando los equipos y servidores intervinientes mediante la *dirección IP* que individualiza cada componente que interviene en la transmisión. Por otra parte, el Derecho Procesal exige que la pericia se realice cumpliendo con criterios técnico-científicos que sustenten formalmente los resultados que se obtienen, es decir, el procedimiento que se realiza debe seguir protocolos establecidos como válidos para revisar la evidencia digital. Por ello, es de interés generar herramientas que –sostenidas en criterios científicos- ayuden al perito en su tarea.

Teniendo en cuenta esto, la presente tesis propone una ontología específica para el análisis forense de correos electrónicos, denominada ***OntoFoCE*** (***O***ntología para la ***Fo***rensia de ***C***orreos ***E***lectrónicos), que se describe en el Capítulo 3. ***OntoFoCE*** es el componente fundamental de la herramienta ***ObE Forensics*** (***O***ntology ***b***ased ***E***mail ***Forensics***), la cual se describe en el Capítulo 4 de esta tesis. ***ObE Forensics*** tiene como propósito último ayudar al perito informático en la tarea de verificar la *autenticidad* de un correo electrónico que se presenta como evidencia judicial.

La definición del dominio de una ontología parte del conjunto de preguntas de competencia utilizadas en la ingeniería ontológica para determinar los requerimientos de una ontología. Esas preguntas guían la delimitación o acotación del universo de discusión, y ayudan a decidir qué objetos son relevantes y cuáles no son representativos para el análisis. En el caso de OntoFoCE, el dominio está definido por el contexto en el que se realiza el análisis forense de correos electrónicos. Particularmente interesa representar tres elementos: el objeto de estudio (correo electrónico), el procedimiento de análisis forense (a partir de los datos de la cabecera del correo electrónico) y los puntos de pericia que usualmente actúan como fin último de realización de la pericia.

Los puntos de pericia son particulares para cada caso judicial, usualmente no se repiten en su enunciado ya que deben explicitar la tarea del perito con toda claridad y

según los detalles de la causa. Aun así, es posible considerar que existen datos comunes que habitualmente se toman como base para responder a los puntos de pericia. Para encontrar un conjunto de conceptos comunes se realizó un relevamiento sobre pericias de correos electrónicos entre un grupo de profesionales informáticos que actúan como peritos habilitados, obteniéndose un conjunto inicial de 86 puntos de pericia, que luego se concentraron en 46 puntos según criterios de similitud de contenido, y que luego derivaron en la definición de 21 preguntas de competencia que se responden utilizando los conceptos definidos en OntoFoCE.

Mediante la definición de las *ocurrencias* (sucesivas copias del correo que va almacenándose en cada equipo/servidor que interviene en la transmisión), OntoFoCE permite representar el paso del correo electrónico desde el equipo emisor hasta el equipo receptor considerando los sucesivos servidores de paso, y luego, derivar la *trazabilidad* del proceso de transmisión, validando de esta forma la procedencia u origen del correo electrónico.

Para la construcción de OntoFoCE se recurrió a la metodología Methontology, aprovechando la capacidad de iteración que propone ésta, para ajustar sucesivamente la ontología durante las fases de desarrollo, hasta llegar a una versión adecuada y suficiente que luego se utilizó como componente fundamental de la aplicación propuesta. Todo el proceso de construcción de la ontología propuesta se describe en el capítulo 3.

En el capítulo 4 se describe el proceso mediante el cual, a partir de OntoFoCE se construyó ObE Forensics, una aplicación web que toma la cabecera del correo electrónico, instancia sus datos en la ontología y pone a disposición del perito un menú de 21 preguntas que responden a los puntos de pericia solicitados por el Juez.

Por último, ambos componentes –la ontología y la herramienta construida a partir de ella- se validaron obteniéndose resultados satisfactorios. Las instancias de validación se describen en el Capítulo 5. Allí se muestra la validación de OntoFoCE mediante una metodología integrada basada en el análisis comparativo respecto de un corpus de conocimientos, y la validación de ObE Forensics realizada por usuarios expertos a los que se invitó a probar la aplicación con casos reales, obteniéndose valiosa información del protocolo de experimentación que se les entregó para que lo completen con los datos relevantes del caso de estudio procesado.

En el contexto de la Forensia Digital, existen muchas y variadas herramientas para realizar el análisis forense de correos electrónicos, pero todas ellas se agotan en la obtención de los datos técnicos que luego el perito utiliza para responder los puntos de pericia. Si bien la mayoría de ellas parten del análisis de la cabecera del correo electrónico para realizar el análisis forense, no consideran la trazabilidad del proceso de transmisión para sustentar la autenticidad del correo electrónico ni incluyen respuestas concretas a los puntos de pericia. Estas consideraciones se sustentan en el detallado estudio que se realizó para definir el marco teórico de esta tesis, desarrollado en el capítulo 2, en el que se consideraron las investigaciones realizadas en los últimos cinco años acerca de Forensia Digital, y particularmente el uso de las ontologías para representar herramientas y/o artefactos forenses.

Fue necesario orientar el estudio y la investigación a la utilización de las tecnologías semánticas en el contexto de la forensia digital, abordando el tema desde varias ópticas.

Por una parte, se circunscribió el contexto de aplicación y experimentación del tema en estudio, según criterios de alcance, profundidad, oportunidad y acceso a problemáticas reales de la Forensia Digital, identificando los problemas básicos al momento de realizar

el análisis forense del correo electrónico. Es así que se trabajó en base a los siguientes criterios:

- Se toma el concepto de *trazabilidad* del correo electrónico, como hilo conductor del análisis forense, para abordar la propiedad de existencia indicada.
- Los puntos de pericia, que actúan como guía para el análisis forense del correo electrónico, se representan en la ontología en términos de las *preguntas de competencia,* de modo que, al ser respondidas, sea posible responder a los puntos de pericia enunciados por el Juez.
- La ontología construida debe tener la capacidad de representar el dominio, cualquiera sea la estructura del correo electrónico que se está analizando, aún en aquellos casos en los que dicha estructura no se ajusta a las normas de rigor.

En esta tesis se resalta la utilización de la ingeniería ontológica en el desarrollo de una herramienta de soporte para el análisis forense, que permita sostener científicamente el carácter de no repudio de la evidencia digital, cuando ésta es un correo electrónico. Dando lugar a los dos aportes más importantes de la tesis:

- La ontología OntoFoCE, que permite la representación de la trazabilidad de un correo electrónico a fin de comprobar la autenticidad del mismo como prueba digital y en consecuencia la condición de no repudiabilidad de la prueba.
- La aplicación ObE Forensics que, basado en OntoFoCE, constituye una herramienta de soporte para la tarea del perito informático, agregando un conjunto de beneficios a su tarea, entre los que se destacan la agilidad en la actividad forense, así como mayor precisión en el análisis debido al procesamiento automático de la evidencia que le permite evitar errores humanos resultantes del análisis visual.

Los resultados que se fueron obteniendo durante esta investigación, realizada desde 2014 a la fecha, se fueron publicando en eventos académicos y publicaciones especializadas, y aquí se detallan:

- B. Parra, M. Vegetti, H. Leone "Advances in the application of Ontologies in the area of Digital Forensic Electronic Mail", IEEE LATIN AMERICA TRANSACTIONS, Vol. 17 (10), pág. 1694-1705, diciembre 2019.
- Notario Enzo, Parra de Gallo Beatriz, Vegetti Marcela, Leone Horacio, "Herramienta para el Análisis Forense de Correos Electrónicos", Revista Ibérica de Sistemas y Tecnología de Información, RISTI, N° 32, pág. 17-32, DOI: 10.17013/risti.32.17–32, junio de 2019.
- Notario Enzo, Parra de Gallo Beatriz, Vegetti Marcela, "ObE Forensics: Una herramienta para el Análisis Forense de Correos Electrónicos", presentado en IX CIIDDI, Montevideo (Uruguay), mayo de 2019.
- Notario Enzo, Rivetti Esteban, Parra de Gallo Beatriz. "Desarrollo de herramienta para análisis de cabeceras de correos electrónicos", VIII CIIDDI 2018, Salta, Mayo de 2018.
- Parra de Gallo Beatriz, Vegetti Marcela, Leone Horacio, "Hacía una Ontología para el soporte de la trazabilidad del correo electrónico en la Forensia Digital", presentado al VII CIIDDI (Congreso Iberoamericano de Investigadores y Docentes de Derecho e Informática), realizado en La Habana (Cuba), del 15 al 20 de mayo del 2017.
- Rivetti Esteban, Aráoz Fleming José, Parra de Gallo Beatriz, Leone Horacio "Análisis de los Documentos Oficiales sobre Obtención, Tratamiento y Preservación de la Evidencia Digital, Aportes para el Tratamiento del Correo Electrónico como Evidencia Digital", presentado en el IV CoNaIISI, desarrollado en Salta, del 17 al 18 de noviembre de 2016.

- Parra de Gallo Beatriz, Leone Horacio, “Aplicación de la Ingeniería Ontológica para representar la trazabilidad de un Correo Electrónico”, presentado en la 45 JAIIO, realizadas en la ciudad de Buenos Aires, del 5 al 9 de setiembre de 2016.
- Parra de Gallo Beatriz, Vegetti Marcela, Leone Horacio, “Avances en la Construcción de una Ontología para el Análisis Forense de Correo Electrónico”, presentado al VI CIIDDI (Congreso Iberoamericano de Investigadores y Docentes de Derecho e Informática), realizado en Santa Fe (Sta. Fe), del 5 al 7 de mayo del 2016.
- Parra de Gallo Beatriz, Leone Horacio “Aplicación de tecnologías semánticas a la Forensia Digital ETAPA 1: Estudio y Diseño de una Ontología Semántica”, presentado en el WICC - Workshop de Investigadores en Ciencias de la Computación, UNL, Santa Fe, 2016 (marzo 2016).
- Beatriz P. de Gallo, Vegetti Marcela, Leone Horacio, “Población de ontologías con datos no estructurados utilizando herramientas de minería de datos”, CoNaIISI 2015 Actas del 3° Congreso Nacional de Ingeniería Informática/Sistemas de Información, Buenos Aires, Argentina, ISBN: 978-987-1896-47-9, noviembre 2015.
- Beatriz P. de Gallo, Marcela Vegetti, Horacio Leone, “Ontología para el Análisis Forense de Correo Electrónico”, CoNaIISI 2014 - ISSN: 2346-9927 - Página 1008-1018, noviembre 2014.

Otros antecedentes de investigación sobre Forensia Digital se indican a continuación:

- Luz Clara Bibiana, Rivetti Esteban, Gamarra Álvaro, Aráoz Fleming José, Parra de Gallo Beatriz, “Forensia de Internet de las Cosas (IoT), avances en la investigación”, presentado en IX CIIDDI, Montevideo (Uruguay), mayo de 2019.
- Aráoz Fleming José, Luz Clara Bibiana, Parra de Gallo Beatriz, “La importancia de las pericias informáticas en el ámbito del Derecho”, FODERTICS, Salamanca (España), Abril 2018.

- Notario Enzo, Luz Clara Bibiana, Parra de Gallo Beatriz, “Análisis de Herramientas para protección de datos personales”, CONAIISI 2018, Mar del Plata, Noviembre de 2018.
- Luz Clara Bibiana, Rivetti Esteban, Gamarra Álvaro, Aráoz Fleming José, Parra de Gallo Beatriz, “Propuesta de Criterios Técnicos y Legales para responder a la Vulnerabilidad de Internet de las Cosas”, XXII Congreso Iberoamericano de Derecho e Informática, Panamá, Setiembre 2018.

TABLA DE CONTENIDO

INDICE DE TABLAS

INDICE DE FIGURAS

LISTA DE ABREVIACIONES

ABAC: Attribute Based Access Control

ARPANET: Advanced Research Projects Agency Network

CPCyC: Código Procesal Civil y Comercial de la Nación Argentina

DFRWS: Digital Forensic Research Conference

DNS: Domain Name System

IMAP: Internet Message Access Protocol

IP: Internet Protocol

IRI: International Resource Identification

ISP: Internet Service Provider

MTA: Message Transfer Agents

NIST: Instituto Nacional de Estándares y Tecnología

ORSD: Ontology Requirements Specification Document

OWL: Web Ontology Language

POP3: Post Office Protocol

PURI: Proceso Unificado de Recuperación de Información

RDF: Resource Description Framework

RDFS: Resource Description Framework Schema

RFC: Request for Comments

SGSI: Sistema de Gestión de Seguridad de la Información

SMTP: Simple Mail Transfer Protocol

SPARQL: Protocol and RDF Query Language

SWRL: Semantic Web Rules Language

TDB: Triple Database

UML: Unified Modeling Language

URI: Uniform Resource Idenfication

URL: Uniform Resource Locator

URN: Uniform Resource Name

W3C: World Wide Web Consortium

XML: eXtensible Markup Language

CAPÍTULO 1. INTRODUCCIÓN

1.1 Escenario actual y justificación del tema

La integración de la tecnología en todos los contextos sociales ha impactado grandemente en el desarrollo de las actividades y las comunicaciones interpersonales. Si bien con beneficios conocidos acerca de los resultados logrados a través de la tecnificación de procesos, la búsqueda de la eficiencia, la salud de las personas y el bien común, como ejemplos de un buen uso de las tecnologías, también se observa una utilización masiva de éstas en actividades criminales.

Los diferentes estamentos de seguridad –tanto militares como judiciales y políticos- se preocupan por encarar la lucha contra el crimen desde la óptica tecnológica, es decir, con una mirada cada vez más preocupante sobre el uso de la tecnología para delinquir. Esto dio lugar a la generación de un espacio propio, dentro de la Informática, denominada Forensia Digital. En 2001 la Digital Forensic Research Conference (DFRWS) definió la "Forensia Digital" como "*El uso de métodos científicamente derivados y probados para la preservación, recolección, validación, identificación, análisis, interpretación, documentación y presentación de la evidencia digital derivada de fuentes digitales para el propósito de facilitar o favorecer la reconstrucción de los hechos criminales o para la prevención de acciones no autorizadas que se estima como perjudiciales para operaciones planificadas*" (Ce, Parlamento, & Del, 2008).

En sus inicios la Forensia Digital hizo su aparición en el ámbito de la justicia con el surgimiento de las primeras pruebas digitales, requiriendo mayormente la presencia de peritos para constatar la veracidad de esas pruebas que técnicamente no representaban grandes desafíos (constatar la existencia de un correo del cual se adjuntaba su impreso, o verificar la funcionalidad de un software o el contenido de un archivo por ejemplo), pero a partir de la popularización de internet la Forensia Digital ha entrado en una crisis producto del impacto de dos elementos que marcan la época actual de la tecnología informática: la masividad de los datos y la multiplicidad de plataformas tecnológicas. (Garfinkel, 2010) presenta varios desafíos, involucrando no solo los modelos de "visibilidad y búsqueda" que

proponen las herramientas forenses de uso actual sino también la falta de integración de las estrategias (como la ingeniería reversa) con dichas herramientas para reducir tiempos y costos. Cita este autor como próximos desafíos a resolver:

- Diseño de las herramientas orientadas a la evidencia: las primeras herramientas forenses se orientaron a la búsqueda de elementos digitales (evidencia) pero no a la presentación, resumen o análisis de correlaciones entre los datos encontrados. Este es un requisito necesario, dada la masividad de datos que se pueden encontrar en una evidencia digital.

- Modelo de visibilidad, filtro e informe: las herramientas utilizan interfaces de comunicación con el experto forense que habitualmente no permiten establecer vínculos o relaciones de prioridad entre los datos encontrados. Incluso algunas herramientas se basan en algoritmos computacionales costosos en tiempo y pueden faltarle características de usabilidad para el usuario final. La automatización o generación de scripts para búsqueda y filtro no siempre resultan, lo cual se complica aún más ante el avance continuo de las tecnologías (procesamiento paralelo, virtualización, deep web, etc.)

- Problemas estructurales en las herramientas forenses: en muchos casos se recurre a software desarrollado para el contexto de negocios, o para sistemas transaccionales, que no responden exactamente a las necesidades puntuales de la búsqueda de evidencia digital. Ocurre lo mismo con tecnologías integradas, tales como las aplicaciones monolíticas.

- Abstracción y modularización: debido al volumen de datos que se procesan en la búsqueda de la evidencia digital, se requiere fijar estándares para la identificación, transmisión e intercambio de los datos; igualmente es importante generar arquitecturas de procesamiento que superen los conflictos del software abierto y propietario.

- Enfoque en la identidad del individuo: tomando como atributos todos aquellos datos que puedan generar una "imagen" de la persona (datos de identificación, datos bancarios, correos, vínculos de las redes sociales, etc.).

En el contexto forense, es de suma importancia vincular los datos a partir del significado de cada cosa. No se trata solo de "encontrar la evidencia digital", sino de interpretarla en el contexto de la situación, vinculándola con el resto de los componentes de la investigación (pruebas físicas, interrogatorios, marco legal y procedimental del caso, etc.). De modo que es indispensable avanzar en la forensia

digital desde la óptica de la semántica –como elemento vinculante de todos los componentes del sistema- así como desde un marco referencial que pueda interpretarlo –una ontología-.

Si bien la definición más referenciada en la literatura es la de (Gruber, 1993a) que expresa que "*una ontología es una especificación explícita de una conceptualización*", vale detallar un poco más el concepto, tomando lo dicho por (de Reuver & Haaker, 2009), "*Una ontología es la descripción conceptual y terminológica de un conocimiento compartido acerca de un dominio específico. Dejando de lado la formalización e interoperabilidad de aplicaciones, esto no es más que la principal competencia del término: hacer mejoras en la comunicación utilizando un mismo sistema en lo terminológico y conceptual*".

Aunque la Forensia Digital avanzó en concordancia con la tecnología, es necesario aún trabajar un aspecto que no es propiamente del ámbito tecnológico y que genera un conjunto de interrogantes que impactan en gran medida en los resultados que se obtienen, es decir, la interpretación de los resultados. (Harichandran, Breitinger, Baggili, & Marrington, 2016) señalan la importancia de mejorar las instancias de comunicación entre los técnicos y los profesionales del derecho, incrementando la accesibilidad y usabilidad de las herramientas de análisis forense para facilitar su interpretación por parte de los profesionales no informáticos. El volumen de datos que se obtiene al realizar el análisis forense debe ser interpretado a la luz de la pesquisa. Cualquiera sea el componente sobre el cual se realiza el análisis forense (celulares, correo electrónico, discos, etc.), es habitual que se genere una cantidad de información técnica que es necesario insertar en el conjunto de pruebas documentales de la causa judicial, colocándolo en un estadio de lectura que facilite la interpretación de esos datos técnicos por parte de los profesionales de la criminalística y el derecho.

Por otra parte, la fuerza probatoria del dictamen pericial se sustenta en *"... los principios científicos o técnicas en que (se) funda, en concordancia de su aplicación con las reglas de la sana crítica..."*[1], es decir, es importante sostener la actividad pericial mediante la aplicación de herramientas y técnicas probadas científicamente. En el caso de las evidencias digitales, es habitual que el perito recurra a una revisión manual de la misma, o a la utilización de herramientas informáticas ad-hoc, pero de

[1] Art. 477 del CPCyC (Código Procesal Civil y Comercial de la Nación Argentina).

las cuales no se conoce el marco científico que sustenta la funcionalidad de las mismas. De allí la necesidad de generar herramientas construidas desde un entorno metodológico y científico que sustente los resultados que se obtienen.

En el contexto de estos requerimientos "no técnicos", se encuentra la motivación de este trabajo. Resulta necesario contar con un marco de referencia basado en la conceptualización formal del universo de discusión que permita extraer la información técnica y establecer una correspondencia unívoca con una descripción de evidencia digital. En particular, las ontologías resultan una herramienta universal o pluridisciplinar para facilitar el análisis de la prueba documental, por parte de todos los actores (abogados, jueces, investigadores y peritos).

Los documentos digitales que se pueden proponer como "prueba" en una causa judicial son muy diversos, en cuanto a estructura, formato y origen. A los fines de acotar el estudio de la aplicación de las tecnologías semánticas a la Forensia Digital, se propone considerar el correo electrónico como objeto de estudio, y en particular, aplicar estas tecnologías para formular una ontología que permita incorporar el análisis forense de correos electrónicos como evidencia digital no repudiable.

1.2 Alcance del Estudio

La Forensia Digital es lo suficientemente amplia como para abarcar todo aquello vinculado al resguardo, transmisión electrónica de información y su recuperación en un documento digital que luego se considere una *evidencia digital*[2].

Adentrados ya en el tema, se debe definir el alcance del estudio y acotarlo a un área de interés que resulte factible de desarrollar.

Del conjunto de pruebas digitales que más se presentan en las causas judiciales (archivos, imágenes, videos, etc.), el correo electrónico es el más habitual, debido principalmente a que formaliza *por escrito* la interacción entre dos partes.

Respondiendo a criterios de alcance y profundidad del trabajo de investigación, se selecciona y se aborda el correo electrónico de entre todos los diferentes tipos de documentos digitales que pueden analizarse en una pericia. Esto se sustenta en las características propias de la correspondencia epistolar, que no se modifican. Con idéntica naturaleza que el correo postal tradicional, el correo electrónico se encuentra

[2] Cabe aclarar que la Forensia Digital actúa también en otros ámbitos además de la justicia, abocándose a la revisión y análisis de fallas de seguridad principalmente, pero no se considerarán esos contextos en este trabajo.

protegido por las leyes que regulan la correspondencia epistolar. Tanto los datos recibidos cuanto los enviados desde la cuenta de correo, constituyen elementos protegidos bajo el principio de inviolabilidad de las comunicaciones, y de por sí, es el documento digital más cercano a una prueba documental.

Y para este componente –de masiva utilización para la comunicación interpersonal- es de interés identificar su *validez como documento digital y su capacidad como prueba no repudiable*, para lo cual se propone definir los componentes de trabajo que aseguren estas condiciones del correo electrónico, recurriendo a herramientas y estructuras que garanticen la obtención de resultados científica y técnicamente aceptados como válidos.

Desde el punto de vista legal, el correo electrónico tiene interés como documento probatorio en un juicio, por lo que resulta importante introducirlo con la fuerza y el rigor técnico suficiente para que actúe en el proceso judicial de igual manera que lo hiciera cualquier otra prueba material.

Cualquier análisis forense se debe trabajar desde la formulación inicial de la cosa, o sea los puntos de pericia. Su ofrecimiento permitirá al Juez determinar la procedencia de la prueba, es decir, la congruencia entre los aspectos a conocer y la necesidad de un técnico para que lo asesore. Usualmente los puntos de pericia referidos a correos electrónicos abordan cuestiones relacionadas con la verificación de la autenticidad y existencia de un correo electrónico.

Se propone trabajar el análisis forense de correos electrónicos, desde el requisito de legitimidad de la prueba digital, entendiendo con ello que si se puede probar esta propiedad, el correo electrónico será no repudiable. Esta es la esencia de la presente investigación.

Los elementos que permiten verificar la veracidad de un correo electrónico son los siguientes:

- la identificación de los datos del remitente (cuenta de correo y dirección IP),
- la trazabilidad del mismo (identificación del camino recorrido por el correo electrónico desde el equipo emisor, los servidores intermedios y el equipo receptor), y
- los datos del destinatario (cuenta de correo y dirección IP).

Es posible encontrar herramientas disponibles para el análisis forense de correos electrónicos que permiten procesar un conjunto de cabeceras de correos electrónicos, pero la mayoría de ellas se agotan en mostrar los datos de la cabecera, que permiten

responder puntos de pericia simples (como por ejemplo cuales son los datos de emisión/recepción del correo), dejando a consideración del perito la respuesta a los puntos periciales complejos como por ejemplo, establecer la trazabilidad del correo, identificar correos enviados/recibidos en rangos de fechas, o buscar información de correos asociadas entre un conjunto de cuentas de correo, entre otros.

Ninguna de las propuestas analizadas durante la investigación responde totalmente a los requerimientos de contener en un único espacio, toda la información necesaria para el análisis forense de correos electrónicos.

Por ello, esta tesis plantea la utilización de la ingeniería ontológica para desarrollar una herramienta que cumpla con dos condiciones esenciales:

a) garantizar la obtención de resultados científica y técnicamente válidos ajustados a las normas procesales de la justicia; y

b) servir de soporte para verificar la condición de autenticidad del correo electrónico, es decir, sostener la condición de no repudio de este tipo de evidencia digital.

Así, se describe la investigación realizada para obtener ***OntoFoCE,*** una ontología para el análisis forense de correos electrónicos, y ***ObE Forensics,*** la aplicación web basada en dicha ontología; ambos componentes construidos con el propósito de servir como herramientas que permitan comprobar la autenticidad del correo electrónico como evidencia digital y en consecuencia la condición de no repudiabilidad de la prueba.

1.3 Organización de la Tesis

El presente trabajo se estructura de la siguiente manera: el Capítulo 2 describe el marco teórico de la investigación, detallando los aspectos más destacados de la Forensia Digital, las ontologías y el correo electrónico como objeto de estudio. El Capítulo 3 describe la ontología denominada ***OntoFoCE*** (Ontología para Forensia de Correos Electrónicos), tomando Methontology como guía para llegar desde los requerimientos a la formulación del modelo lógico. El Capítulo 4 detalla el desarrollo de ***ObE Forensic*** (Ontology based Email Forensic)**,** la aplicación web basada en la ontología destinada a la realización del análisis forense de correos electrónicos. El Capítulo 5 aborda el proceso de validación realizado tanto para

OntoFoCE como para ObE Forensic. Y por último, en el Capítulo 6 se describen las conclusiones de la investigación.

El marco teórico de la investigación, que se describe en el **Capítulo 2**, incluye las bases teóricas y metodológicas que se tomaron en cuenta para el desarrollo de OntoFoCE y sustentan la aplicación informática denominada ObE Forensic que se desarrolló para el análisis forense de correos electrónicos. Durante la elaboración de esta tesis se avanzó de manera conjunta en dos líneas de conocimiento: ontologías –por una parte- y Forensia Digital–por la otra-. En ambas áreas, el punto de conexión fue el estudio del correo electrónico y las características, avances y estado del arte como objeto de estudio. Se estudió también el procedimiento de análisis forense y sus características cuando la evidencia es un correo electrónico.

El **Capítulo 3** está dedicado a describir la ontología propuesta. Se introducen algunos aspectos metodológicos y se define el alcance de la ontología, poniéndose énfasis en el modelo conceptual de OntoFoCE. Asimismo, se describe cómo se representa el correo electrónico, los conceptos y relaciones sustantivas, y el proceso de transmisión del mismo, que para una mejor comprensión se muestra en términos de los subprocesos involucrados (emisión, transmisión y recepción). Este capítulo termina con la ejemplificación de la instanciación de la ontología, considerando tres escenarios distintos de pericias informática, y la respuesta de las preguntas de competencia para cada caso.

Luego de explicar la ontología, en el **Capítulo 4** se describe ObE Forensic, la aplicación web desarrollada para el análisis forense de correos electrónicos. Se detallan las funcionalidades de la aplicación, con énfasis en el modo en que se resuelven los procesos de instanciación de los datos en OntoFoCE, las tecnologías utilizadas para implementar la aplicación web, y se explica el uso de ObE Forensic para los escenarios ya ejemplificados en el Capítulo 3, pero esta vez, se siguen desde la aplicación web. Por último, en este capítulo se agrega un cuarto escenario forense que muestra la capacidad de la herramienta para procesar una importante cantidad de cabeceras provenientes de más de una cuenta de correo electrónico.

El **Capítulo 5** describe los procesos realizados para validar OntoFoCE y ObE Forensic, según un enfoque integrado. Se presentan las técnicas y herramientas utilizadas para validar la ontología construida, buscando identificar y corregir errores en la conceptualización y diseño de la OntoFoCE, en su estructuración, uso del lenguaje, vocabulario utilizado, y respuesta a los requerimientos planteados. Además

se aprovecha el espacio de este capítulo para explicar los resultados obtenidos de la validación por usuarios expertos, realizada sobre la aplicación web ObE Forensic que utiliza dicha ontología para el análisis forense de correos electrónicos.

Por último, el **Capítulo 6** incluye las conclusiones arribadas de esta investigación, el grado de cumplimiento de los objetivos formulados para la misma y las futuras líneas de investigación que pueden derivarse.

La tesis finaliza con la inclusión de los siguientes anexos:

- ANEXO I: INVESTIGACIÓN BIBLIOGRÁFICA
- ANEXO II: PUNTOS DE PERICIA
- ANEXO III: REPRESENTACIONES INTERMEDIAS DE LA CONCEPTUALIZACIÓN DE OntoFoCE
- ANEXO IV: AUTORIZACIONES DE USO DE DATOS
- ANEXO IV: INFORME DE ANÁLISIS FORENSE PARA ESCENARIO 1
- ANEXO V: INFORME DE ANÁLISIS FORENSE PARA ESCENARIO 2
- ANEXO VI: INFORME DE ANÁLISIS FORENSE PARA ESCENARIO 3
- ANEXO VII: CÓDIGO OWL DE ONTOFOCE
- ANEXO VIII: INSTRUCTIVO DE USO DE ObE Forensics

CAPÍTULO 2. MARCO TEÓRICO

2.1 Introducción

El contenido de este capítulo describe las bases teóricas y metodológicas que se tomaron en cuenta para el desarrollo de la ***OntoFoCE*** (Ontología para Forensia de Correos Electrónicos).

Durante la elaboración de esta tesis se avanzó de manera conjunta en las dos temáticas de interés para la investigación: ontologías y forensia digital. En ambas áreas, se consideró particularmente el estudio del correo electrónico y las características, avances y estado del arte como objeto de estudio. En particular se avanzó en la construcción de la ontología para el tratamiento de la evidencia digital de los correos electrónicos. La ontología que se propone en esta tesis está dirigida al contexto judicial y forense, en el cual se desempeñan los profesionales no informáticos que interactúan en una causa judicial (abogados, jueces, policías, investigadores, profesionales forenses de otras áreas). Con ella, se pretende generar un espacio de comunicación común y entendible para todos ellos, que facilite la inclusión del correo electrónico como prueba documental digital.

Este capítulo se ha organizado de la siguiente manera: en la sección 2.2 se aborda la descripción de los aspectos referidos a las tecnologías semánticas que se consideraron para construir OntoFoCE. La sección 2.3 explica los conceptos relevantes acerca de Forensia Digital. En tanto, el Correo Electrónico como objeto de estudio, se describe en la sección 2.4. Luego, en la sección 2.5 se detalla el procedimiento que se sigue para la realización de pericias sobre correos electrónicos y en la sección 2.6 se discute acerca de las herramientas forenses más usuales para el análisis de correos electrónicos. La sección 2.7 contiene el estudio realizado sobre el estado del arte acerca la aplicación de ontologías a la forensia digital y en la sección 2.8 se describen las conclusiones de este capítulo.

2.2 Ontologías

En esta sección se introduce una breve descripción de lo que es una ontología y las posibles metodologías que pueden guiar su desarrollo. Asimismo, se da un

pantallazo general sobre las diversas herramientas que soportan el proceso de construcción de una ontología, los procesos de inferencia y consultas, así como su almacenamiento y su validación.

2.2.1 Qué es una Ontología

El término ontología no es nuevo, proviene del área de la filosofía, fue utilizado con posterioridad en el área de la Inteligencia Artificial y toma nueva relevancia en el siglo 21 en el ámbito de la Web Semántica. Existen varias definiciones de este concepto. (Alberto, Luna, López, Ingrid, & Torres, 2012) señalan las definiciones más conocidas del concepto:

- *"Una ontología define las condiciones básicas y relaciones que comprenden el vocabulario de un área del tema así como las reglas para combinar condiciones y las relaciones para definir extensiones del vocabulario"* (Neches et al., 1991).
- *"Una especificación explícita de una conceptualización, es decir, que proporciona una estructura y contenidos de forma explícita que codifica las reglas implícitas de una parte de la realidad; estas declaraciones explícitas son independientes del fin y del dominio de la aplicación en el que se usarán o reutilizarán sus definiciones"* (Gruber, 1993b).
- *"La ontología describe una cierta realidad con un vocabulario específico, usando un conjunto de premisas de acuerdo con un sentido intencional de palabras del vocabulario"* (Guarino, 1997).

También son de interés destacar otras definiciones, como por ejemplo:

- *"Una ontología es una especificación explícita y formal de una conceptualización compartida. Se entiende por i) Conceptualización: un modelo abstracto de un dominio particular, que identifica los conceptos relevantes de dicho dominio y sus relaciones; ii) Formal: la ontología debe ser definida mediante una teoría lógica, lo cual permite que la misma pueda ser procesada por agentes de software; iii) Explícita: los conceptos, sus relaciones y sus restricciones están definidos explícitamente; y iv) Compartida: refleja que una ontología debe, en el mejor de los casos, dar cuenta de conocimiento aceptado como mínimo, por el grupo de personas que deben usarla."* (Studer, Benjamins, & Fensel, 1998)

Ya en el ámbito de la Web Semántica, la W3C (World Wide Web Consortium)[3] define una ontología como *"…los términos utilizados para describir y representar un área de conocimiento. Las personas, bases de datos y aplicaciones que necesitan compartir información de dominio utilizan ontologías (un dominio es solo un área temática específica o área de conocimiento, como medicina, fabricación de herramientas, bienes raíces, reparación de automóviles, administración financiera, etc.). Las ontologías incluyen definiciones computables de conceptos básicos en el dominio y las relaciones entre ellos [...]. Codifican el conocimiento en un dominio y lo hacen extensible a varios dominios. De esta manera, hacen que ese conocimiento sea reutilizable…".*

Pero más allá de la definición conceptual, ¿Cuál es la característica destacable de una ontología que hace posible su aplicación en diversos ámbitos disciplinarios? Pues la utilidad de las ontologías se observa en la capacidad de describir una realidad en los propios términos de sus actores. Es decir, permite establecer una base de comunicación e interpretación de los conceptos referidos a un dominio que facilita la comprensión entre aquellos que comparten ese dominio, y a partir de éste, las ontologías nutren a la Web Semántica con nuevos términos o conceptos básicos optimizando el proceso de búsqueda de información relacionada temáticamente. Asimismo, la característica de estar escrita en un lenguaje formal, permite que una ontología sea interpretable por una computadora, posibilitando procesos de consultas e inferencia de conocimiento implícito.

De acuerdo a (Guizzardi, 2005) una conceptualización es el "*conjunto de conceptos utilizados para articular abstracciones del estado de cosas en un dominio dado. La abstracción de una determinada porción de la realidad articulada de acuerdo con una conceptualización se denomina modelo".* Tanto las conceptualizaciones como los modelos son entidades abstractas que solo existen en la mente de un persona o comunidad. Según este autor, es necesario capturar los modelos mediante un artefacto concreto a fin de que éstos puedan ser documentados, comunicados y analizados. Las ontologías permiten especificar estos modelos, mediante el uso de un lenguaje de modelado (o especificación). Existen muchos lenguajes para la construcción de una ontología, sin embargo, todos tiene al menos los siguientes elementos: clases, propiedades, individuos y axiomas.

[3] https://www.w3.org/2003/08/owlfaq

En una ontología, las clases son las representaciones de los conceptos del modelo, que dicha ontología especifica. Las clases pueden organizarse jerárquicamente mediante relaciones de subsunción formando lo que se conoce como taxonomía de clases. Por ejemplo: *cuenta de correo electrónico* es un concepto del domino de estudio que en la ontología que se propone se define como la clase *Cuenta* que, a su vez, tiene dos subclases: *CuentaEmisor* y *CuentaReceptor,* las cuales representan los roles que juega una cuenta de correo electrónico al enviar y recibir un correo. Las entidades que son instancias de las clases se denominan *Individuos.*

Por su parte, los conceptos tienen características particulares que son representadas en la ontología mediante *propiedades.* Una propiedad puede representar una característica (o atributo) de un concepto o la vinculación entre conceptos. Por ejemplo: la clase *Cuenta* tiene *aliasUsuario* y *cuentaCorreo* como atributos. En tanto que las relaciones, vinculan conceptos entre sí, como en el caso de *cuentaEmisorEmiteCorreo* que vincula una instancia de *Cuenta* con una instancia de *Correo*, indicando cuál es la cuenta de correo desde la cual se emitió el mismo.

Los axiomas y las reglas constituyen un mecanismo para expresar restricciones sobre la interpretación de los conceptos del dominio y para inferir conocimiento que se deriva a partir del conocimiento expresado de manera explícita por el resto de los componentes de la ontología. Las reglas y los axiomas, además, permiten verificar la consistencia de la ontología. Siguiendo el ejemplo de las cuentas, es posible formular una regla que restrinja la unicidad de la cuenta que actúa como emisora de un correo electrónico: "*Todo Correo debe tener una única Cuenta que actúa como emisora"*. Esta restricción se puede expresar en lógica de primer orden (LPO) de la siguiente manera:

$$\forall x Correo(x) \exists e_1, e_2 \,/\, CuentaEmisor(e_1) \wedge CuentaEmisor(e_2)$$

$$\wedge\, cuentaEmisorEmiteCorreo(x, e_1) \wedge cuentaEmisorEmiteCorreo(x, e_2) \Rightarrow e_1 = e_2$$

La expresión anterior indica que, para toda instancia de la clase *Correo* (que representa el concepto *Correo Electrónico*) existe una y sólo una instancia de la clase *CuentaEmisor*, vinculada con ella mediante la relación *cuentaEmisorEmiteCorreo*.

Hasta aquí se mencionaron los aspectos generales de una ontología, más adelante, en el Capítulo 3, se profundiza la explicación y utilización de los mismos al construir OntoFoCE.

Continuando con las definiciones necesarias para comprender los distintos elementos que se trabajan en la ingeniería ontológica, se detalla en el apartado siguiente las herramientas metodológicas propias de las ontologías.

2.2.2 Métodos y Metodologías para el desarrollo de Ontologías

A partir de 1990 comenzaron a definirse metodologías para el desarrollo de ontologías, la historia desde esa fecha a la actualidad muestra el avance en la propuesta de metodologías cada vez más optimizadas según diferentes agregados que hicieron los distintos autores.

En 1990, (Lenat, Guha, Pittman, Prat, & Shepherd, 1990) publicaron los pasos generales y orientaciones acerca del ciclo de desarrollo de sistemas basados en el conocimiento. Más tarde, en 1995 (Uschold & King, 1995) proponen una serie de pasos para la especificación del dominio.

Por esa misma época (Gruninger & Fox, 1995) introducen el concepto de *preguntas de competencia,* las cuales consisten en un conjunto de interrogantes –escritos en lenguaje natural- que el conocimiento capturado por la ontología debe poder responder. La definición de *preguntas de competencia,* constituye una de las técnicas más utilizadas para definir el dominio y alcance de una ontología.

Por su parte (Schreiber, Amsterdam, Wielinga, & Jansweijer, 1995) definen KACTUS, que aborda el desarrollo de la ontología en tres etapas: especificación de la aplicación, diseño preliminar basado en categorías ontológicas top-level relevantes y refinamiento y estructuración de la ontología.

Luego en 1997 (Fernández, Gómez Pérez, & Juristo, 1997) presentan Methontology incorporando conceptos de la ingeniería de software y ciclo de vida al proceso de desarrollo de la ontología. Por la misma época, surge un método basado en la ontología SENSUS (Swartout, Knight, Russ, & Rey, 1997), proponiendo un nuevo paradigma para la construcción de modelos, donde el foco está en vincular términos específicos del dominio a una ontología existente en lugar de construir un modelo desde cero.

La metodología On-To-Knowledge (Fensel et al., 2000) trabaja sobre información disponible electrónicamente para mejorar los procesos de gestión del conocimiento en las grandes organizaciones, incluyendo escenarios de casos de uso y roles de los trabajadores del conocimiento.

En el año 2002, aparece TERMINAE (Szulman & Biébow, 2002) para la construcción de ontologías a partir de textos, considerando el análisis lingüístico de los textos y herramientas para el procesamiento del lenguaje natural.

Una metodología de interés es NeOn (Suárez-Figueroa, 2010), propuesta para construir redes de ontologías a través del desarrollo colaborativo, con énfasis en las guías metodológicas para considerar distintos escenarios de la construcción de ontologías como cuando se consideran casos de reutilización de recursos ontológicos y no ontológicos que las otras metodologías no contemplan.

En el último decenio y en el afán de encontrar el método más adecuado, los estudiosos del tema desarrollaron muchas metodologías específicas (o propias de un dominio), pero en su esencia la mayoría de ellas contienen al menos cinco fases de construcción de una ontología (especificación, conceptualización, formalización, implementación y evaluación).

En particular, para desarrollar esta tesis se seleccionó *Methontology*, una metodología desarrollada por el Grupo de Ingeniería Ontológica de la Universidad Politécnica de Madrid. Mediante la incorporación de elementos propios de la Ingeniería de Software, esta metodología propone guías de actividades para la especificación, conceptualización, formalización, implementación y mantenimiento de la ontología a construir, bajo un esquema de procesos iterativos que ayudan en el ajuste del modelo a construir.

Methontology no impone pasos estrictos o métodos a seguir en cada una de las etapas, sino que indica lineamientos generales para cada fase, los que a continuación se describen.

La actividad de ***especificación*** permite determinar por qué se construye la ontología, cuál será su uso, y quiénes serán sus usuarios finales. En otras palabras, esta actividad permite establecer los requerimientos, así como el alcance de una ontología. Si bien Methontology no propone ningún método particular para desarrollar los requerimientos de la ontología, es válido considerar la propuesta de (Suárez-figueroa, Gómez-pérez, & Villazón-terrazas, 2009) para definir el documento *ORSD* (Ontology Requirements Specification Document), que incluye

los siguientes componentes: propósito de la ontología a desarrollar; usos previstos y usuarios de la ontología a desarrollar; y conjunto de requisitos que la ontología debe satisfacer después de ser formalmente implementada, definidas en términos de preguntas de competencia. Respecto de esto último, las preguntas de competencia, cabe mencionar que es habitual el uso de estas consignas para definir los requerimientos de una ontología.

El paso siguiente en la construcción de la ontología es la ***conceptualización,*** que se encarga de organizar y convertir una percepción informal del dominio en una especificación semiformal, para lo cual utiliza un conjunto de representaciones intermedias, basadas en notaciones tabulares y gráficas, que pueden ser fácilmente comprendidas por los expertos de dominio y los desarrolladores de ontologías. En esta fase se recurrió a las definiciones de tablas, diagramas UML (Unified Modeling Language), reglas y axiomas para restringir la interpretación de dichos diagramas. En la siguiente sección de este mismo capítulo se describe en detalle estas herramientas.

Para la etapa de ***formalización e implementación,*** en la que se realiza la transformación de dicho modelo conceptual en un modelo formal o semicomputable, se recurre a un conjunto de herramientas que se describen con mayor detalle en la sección siguiente, y que cabe adelantar en estos párrafos: el lenguaje de descripción de la ontología es OWL, se utiliza SWRL para definir las restricciones del modelo y las preguntas de competencia se expresan como consultas escritas en SPARQL (Protocol and RDF Query Language).

Por último, se requiere incluir en el proceso metodológico la ***validación*** de la ontología. Si bien existen diversas herramientas que ayudan en la definición y construcción de una ontología, son escasos los estudios referidos a cómo evaluar una ontología. A modo orientativo se puede considerar los modos de evaluación de ontologías que cita (Ouyang, Zou, Qu, & Zhang, 2011):

- Métodos basados en el usuario: en los que se considera la experiencia de los usuarios, con lo cual, el resultado queda sujeto a la experticia de los mismos en el conocimiento del dominio de la ontología.
- Métodos basados en la aplicación de la ontología: el inconveniente aquí es que la ontología se utilizó antes de la evaluación, por lo que los usuarios no pueden conocer los resultados de la evaluación de la ontología antes de usarla lo cual le quita confianza al modelo.

- Métodos basados en la comparación de la ontología con un "estándar de oro" (o estándar de alta calidad) que puede ser en sí mismo otra ontología. El problema aquí sigue siendo como se mide la calidad de ese “estándar de oro”.
- Métodos basados en la comparación con la fuente de datos: en estos casos el inconveniente radica en que se está comparando la ontología contra un conjunto de documentos “fuentes” haciendo que el proceso de comparación sea poco eficiente.

La evaluación de una ontología debe realizarse siguiendo las buenas prácticas de validación y control de calidad fijada para el desarrollo de software. La evaluación consiste en someter la ontología a una verificación de cumplimiento de los requerimientos que le dan origen, es decir, las preguntas de competencia. En el caso de OntoFoCE, las preguntas de competencia son interrogantes expresados en lenguaje natural que formulan los usuario expertos, es decir, los peritos informáticos, y representan lo que ellos requieren o pretenden de la ontología. De ese modo, si la ontología responde satisfactoriamente dichas preguntas, entonces se puede aseverar que la misma se ha validado.

Methontology propone la actividad de ***mantenimiento*** en la que se realiza la actualización y/o corrección de la ontología, a partir de las modificaciones y/o ajustes que surgen de la utilización de la misma por los usuarios expertos.

Methontology también identifica actividades de gestión (planificación, control y aseguramiento de la calidad), y de soporte (adquisición de conocimientos, integración, evaluación, documentación y gestión de la configuración).

La decisión de porqué utilizar este marco de desarrollo se funda en que Methontology propone una visión integrada de la gestión del proyecto, haciendo posible un trabajo organizado y formal de todas las actividades, para acompañar el proceso iterativo de las sucesivas etapas de la construcción de la ontología. En el Capítulo 3 se explica detalladamente cómo se aplicaron las distintas fases de Methontology para la construcción de OntoFoCE.

Habiendo descripto las metodologías habituales para la construcción de una ontología, y en particular, cuál de ellas se considera más adecuada para el desarrollo de la ontología objeto de esta tesis, OntoFoCE, se aborda a continuación la descripción de los instrumentos requeridos en cada fase de la metodología.

2.2.3 Herramientas para la construcción de ontologías

Existe una gran variedad de herramientas específicas para la construcción de ontologías. (Gómez-Pérez, Fernández López, & Corcho, 2004) las clasifican teniendo en cuenta la función que cumplen en el proceso de desarrollo. De esta clasificación se detallan las utilizadas en la construcción de OntoFoCE.

Por una parte están las *herramientas de desarrollo de ontologías*, se trata de aquellas utilizadas para la construcción de nuevas ontologías o para la reutilización de ontologías preexistentes. Usualmente son editores de ontologías. Cobra especial importancia los lenguajes que se utilizan para construir y formalizar la ontología. Las *herramientas de consulta ontológica y motores de inferencia* permiten realizar consultas a la ontología con facilidad, realizando inferencias. Normalmente están fuertemente relacionadas con el lenguaje utilizado para implementar la ontología. Luego están las *herramientas de almacenamiento y recuperación de información* que son aquellas orientadas a facilitar la gestión de búsquedas, consultas, almacenamiento y recuperación de información a partir de una ontología. Por último, las *herramientas de evaluación de ontologías* son los componentes de apoyo que permiten mejorar la calidad de la ontología desarrollada.

En el caso de OntoFoCE, se recurrió al uso de un conjunto de herramientas que responden a la clasificación citada en el párrafo anterior, y se describen en detalle a continuación.

2.2.3.1 Herramientas de Desarrollo de ontologías

Estas herramientas se pueden considerar en base al grado de generalidad que presentan. Por una parte, están las herramientas desarrolladas expresamente para un proyecto en particular, y por otro, están los Kits o Conjuntos Integrados de Herramientas, o plataformas de trabajo para ontologías que contienen componentes para trabajar según diversas funcionalidades.

Este segundo grupo de herramientas tienen la característica de incluir una arquitectura orientada a la gestión del conocimiento en la ontología, con

independencia de la metodología, el lenguaje y demás componentes de la solución. Las más conocidas son: *Fluent Editor*[4], *WebODE*[5], *Protégé*[6], *OntoEdit*[7] y *KAON2*[8].

La principal característica del editor de ontologías *Fluent Editor* es el uso del inglés controlado como lenguaje de modelado de conocimiento, prohibiendo la escritura de cualquier oración que sea gramatical o morfológicamente incorrecta.

Aun cuando el Grupo de Ingeniería Ontológica de la Universidad Politécnica de Madrid, desarrollador de *WebODE,* discontinuó el mantenimiento de esta herramienta, ésta se sigue utilizando. Principalmente porque es escalable, extensible e integrada abarcando la mayoría de las actividades de desarrollo de una ontología.

Protégé ha sido desarrollado por Stanford Medical Informatics (SMI) de la Universidad de Stanford. De código abierto, es una herramienta con arquitectura extensible, que consta de un editor de ontologías como componente base, y se le puede adicionar complementos de la biblioteca disponible, según la funcionalidad que se quiera introducir. El panel principal organizado en ventanas permite visualizar y trabajar ya sea de manera individual los distintos componentes, como de manera integrada cuando se necesita. Esta herramienta permite conectarse con un razonador para identificar inconsistencias en la ontología que se está construyendo.

OntoEdit desarrollado por el Institute AIFB de la Karlsruhe University, es un ambiente extensible y flexible para editar ontologías, construido teniendo en cuenta cinco objetivos principales: facilidad de uso, desarrollo guiado de la ontología, desarrollo con ayuda de inferencias, desarrollo de axiomas ontológicos y extensibilidad a través de plug-ins.

La suite de la herramienta *KAON* (Karlsruhe Ontology) ha sido desarrollada como el sucesor de OntoEdit, que después derivó en *KAON2,* mejorada en base a la incorporación de una *API* (Application Programming Interface) para gestionar ontologías de diferentes tipos, un servidor independiente para acceso a ontologías distribuidas, un motor de inferencias, una interfaz de comunicación con Protégé y un módulo para la extracción de instancias desde bases de datos relacionales.

Sea cual fuere el editor de ontologías seleccionado, cuando se trabaja con ontologías, una de las cuestiones más importantes es el *lenguaje* utilizado para

[4] http://www.cognitum.eu/semantics/FluentEditor/
[5] http://oa.upm.es/5483/
[6] www.protege.stanford.edu
[7] https://link.springer.com/chapter/10.1007/3-540-36124-3_76
[8] http://kaon2.semanticweb.org/

representar el conocimiento –más allá de la información o el dato mismo- y pasar de una expresión escrita en lenguaje natural a una expresión computable.

En el caso particular de los lenguajes para expresar ontologías, se requiere que cumpla con las siguientes características:

- Sintaxis formalizada, que permita diseñar un motor de inferencias sólido.
- Semántica bien definida para una correcta implementación del modelo.
- Expresividad suficiente para representar el conocimiento lo más acabadamente posible.
- El costo computacional del razonamiento debe ser razonable.

A mediados de la década del 80 aparecieron algunos lenguajes para la representación del conocimiento que mostraban parte de estas características. Al principio, se recurrió a los lenguajes basados en la lógica de predicados de primer orden para escribir las ontologías. En ese sentido, *Prolog*[9] fue útil por su capacidad para formalizar la sintaxis y la semántica a partir de las reglas de inferencia de la lógica de predicados de 1° orden y resolución tipo *modus ponendo ponens* (si P implica Q; y si P es verdad; entonces Q también es verdad). Y también surgieron los lenguajes basados en modelos de la psicología para representar el comportamiento de la mente humana, que aun cuando cuentan con algoritmos de razonamiento más eficientes, no tienen una sintaxis formalizada ni una definición semántica precisa. Un ejemplo de estos lenguajes son las *Redes Semánticas*[10] que utilizan el modelo de grafos para representar el conocimiento lingüístico.

Surgiendo a la par de las metodologías, en el año 1990, los lenguajes de ontologías derivaron de estos primeros lenguajes utilizados para la representación del conocimiento, tomando de éstos los componentes de lógica de primer orden, a los que se le fueron agregando características cada vez más orientadas al análisis semántico de los datos. Lo que toma mayor fuerza cuando se comienza a hablar de la *Web Semántica*, con la intención de que la web no solo almacene información y la muestre al usuario, sino que además permita la *interpretación semántica* mediante un *lenguaje de marcado*, que describe el contenido, el significado y la relación de los datos de una web.

Surge así el lenguaje *XML* (eXtensible Markup Language), que permite definir etiquetas personalizadas para descripción y organización de datos y así representar

[9] https://www.visual-prolog.com/
[10] https://www.ecured.cu/Redes_sem%C3%A1nticas

información estructurada en la web, de modo que esta información pueda ser procesada por aplicaciones que requieran de esos datos. XML proporciona una sintaxis firme para documentos estructurados, aunque sin restricciones semánticas sobre el significado de estos documentos. Derivado de este último, *XML Schema* permite definir tipos de datos y *XML Namespaces* incluye repositorios de definiciones reutilizables ayudando a conformar vocabularios para dominios completos.

Haciendo uso de XML, surge *RDF* (Resource Description Framework), el estándar definido por la W3C para representar información sobre los recursos de la web. RDF está especialmente diseñado para representar metadatos sobre recursos web, como los datos del autor de una página o la fecha de creación de la misma, pero también es posible utilizar RDF para representar información sobre cosas que se pueden identificar en la web, de este modo los datos contenidos en una web pueden ser procesados por las aplicaciones sin perder su significado además de ser mostrada solo a las personas.

RDF identifica los recursos mediante identificadores web, conocidos como URI (Uniform Resource Idenfication) y describe los recursos en términos de propiedades simples y valores de propiedades. Esto permite a RDF definir declaraciones simples sobre recursos como un gráfico de nodos y arcos que representan los recursos, sus propiedades y valores. Una declaración RDF, denominada también *tripleta,* tiene la siguiente estructura:

<*sujeto*> <*predicado*> <*objeto*>

expresando una relación (representada por el *predicado)* que vincula dos recursos (representados por el *sujeto* y el *objeto).* La naturaleza de la relación entre ambos recursos es siempre de manera direccional (es decir, del *sujeto* hacia el *objeto).*

En la web, un recurso puede ser cualquier objeto referenciable, incluidos documentos, personas, objetos físicos y conceptos abstractos, y se identifica mediante un *URI* (Uniform Resource Identifier) que a su vez consta de dos conceptos:

- Un *URN* (Uniform Resource Name) que es el identificador único que permite referenciar el recurso, y
- Un *URL* (Uniform Resource Locator) que es el identificador único mediante el cual se accede al recurso en algún lugar de la web.

En 2007, aparece el identificador *IRI* (International Resource Identification) que extiende la capacidad de identificación de un URI al utilizar los caracteres ASCII no estándar para codificar el identificador de objeto.

Como ejemplo considere las siguientes declaraciones:

"hay una Persona identificada por
<http://www.semanticweb.org/beatriz/ontologiaCorreos#usuario>,
cuyo nombre es Juan Perez, cuya dirección de correo electrónico es juan@dominio.com,
y cuyo Teléfono es 15412345. "

Estas declaraciones se representan con las siguientes tripletas RDF, en las cuales los prefijos *myOntology* y *rdf* representan los espacios de nombres identificados por los siguientes URIs: *<http://www.semanticweb.org/beatriz/ontologiaCorreos>* y *<http://www.org/1999/02/22-rdf-syntax-ns>* respectivamente*:*

<myOntology#unUsuario rdf:type myOntology:Persona>

<myOntology#unUsuario myOntology:tieneDireccionCorreo myOntology:juan@dominio.com>

<myOntology#unUsuario myOntology:tieneNombre "Juan Perez">

<myOntology#unUsuario myOntology:tieneTelefono "*15412345"*>

La representación gráfica de estas tripletas se muestra en la Figura 2-1, allí se ilustra que RDF utiliza URI para identificar:

- individuos, por ejemplo, el usuario Juan Perez, identificado por *http://www.semanticweb.org/beatriz/ontologiaCorreos#unUsuario,*
- tipos de cosas, por ejemplo, Persona, identificadas por *http://www.semanticweb.org/beatriz/ontologiaCorreos#Persona,*
- propiedades de esas cosas, por ejemplo, la dirección de correo, identificado por *http://www.semanticweb.org/beatriz/ontologiaCorreos#tieneDireccionCorreo,*
- valores de esas propiedades, por ejemplo, *http://www.semanticweb.org/beatriz/ontologiaCorreos#juan@dominio.com* como el valor de la propiedad *tieneDireccionCorreo* (RDF también usa cadenas de caracteres como "Juan Perez", y valores de otros tipos de datos como enteros y fechas, como los valores de las propiedades).

RDFS (Resource Description Framework Schema) es una extensión semántica de RDF, permite describir grupos de recursos relacionados y las relaciones entre estos recursos. Con un esquema similar al sistema de clases y propiedades de la Orientación a Objetos, RDFS permite definir un vocabulario básico en un espacio de nombres denominado informalmente como *rdfs*, bajo el siguiente URI:

http://www.w3.org/2000/01/rdf-schema# y se asocia convencionalmente con el prefijo *rdfs:*.

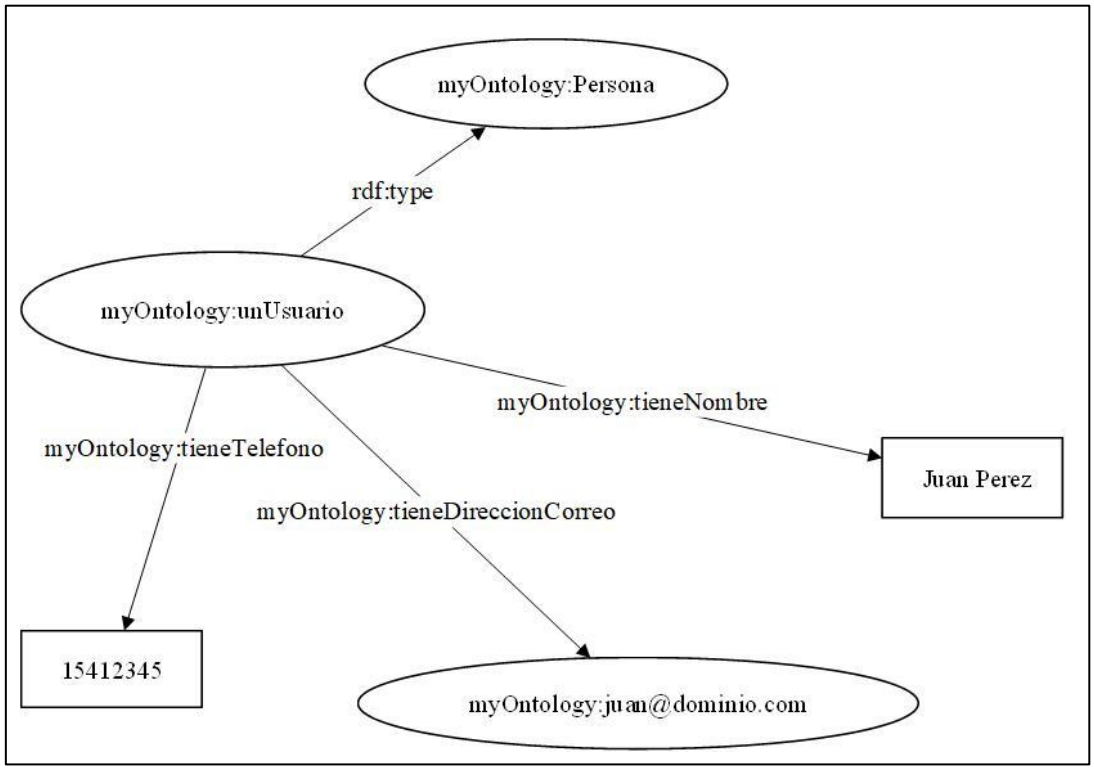


Figura 2-1: Representación gráfica de Tripletas

Tanto RDF como RDFS presentan limitaciones para representar modelos complejos, se hace necesario recurrir a un lenguaje que permita añadir expresividad al modelado de la ontología, así surge *OWL*. Este lenguaje fue desarrollado por el Grupo de Trabajo OWL del W3C en 2004 para representar el conocimiento a partir de los objetos y las relaciones que mantienen entre ellos. En el sitio de *Recomendaciones de la WC3*[11] figuran los estándares y características que definen OWL. Este lenguaje agrega la lógica descriptiva a la estructuración de documentos, mediante la ampliación del vocabulario para describir propiedades y clases en términos de relaciones entre clases, cardinalidad, enumeración de clases, entre otros elementos, de manera que los programas informáticos pueden razonar el conocimiento expresado en dicho lenguaje para verificar la consistencia de ese conocimiento o para hacer explícito el conocimiento implícito. Las ontologías implementadas en este lenguaje pueden ser publicadas en la Web y pueden ser referenciadas desde otras ontologías OWL.

[11] https://www.w3.org/TR/

OWL 2, última versión de este lenguaje, se define a partir de un conjunto de documentos de la W3C que fija las especificaciones sobre su estructura conceptual, la sintaxis de intercambio primario (entre RDF y XML), dos semánticas alternativas (Directa y basada en RDF) y los requisitos de conformidad. Existen además otros documentos auxiliares que definen las sintaxis existentes, tales como functional style[12], OWL/XML[13], RDF/XML[14], sintaxis Manchester[15], Turtle[16], N-triples[17], entre otras.

Hay dos características OWL que hace que a quienes vienen trabajando con lenguajes basados en la lógica de primer orden y/o en bases de datos, les resulte difícil el primer contacto con este lenguaje. Por un lado, OWL no respeta el *principio de unicidad de nombres*, esto significa que dos individuos que se llamen de manera diferente pueden ser el mismo individuo. Por otra parte, OWL 2 adhiere a la suposición del *mundo abierto*, esto significa que no se puede aseverar que algo no existe hasta que se afirme explícitamente que no existe, o sea, no se puede declarar "inexistente" solo porque no se haya podido probar su existencia.

(Baader & Nutt, 2003) define el concepto de "mundo abierto" a partir de la relación entre el vocabulario de una base de conocimiento y las afirmaciones sobre el mismo. Un sistema de conocimiento basado en lógica descriptiva brinda ventajas para configurar el conocimiento, permite razonar sobre su contenido y generar nuevo conocimiento. Usualmente se puede considerar dos componentes en una base de conocimientos: el *TBox* o vocabulario y el *Abox* que son el conjunto de afirmaciones sobre los individuos de ese vocabulario. Los elementos de TBox son los que denominamos conceptos, que definen un grupo de individuos con roles o relaciones entre ellos, a partir de este esquema de conceptos y relaciones se pueden construir descripciones más complejas. Según se considere los elementos de Abox, se puede hablar de "mundo cerrado", como en el caso de las bases de datos, en donde la relación entre un concepto y otro está explícitamente definida y no tiene ningún otro significado más que aquel que se estableció con esa relación, Es decir, de la base de datos no se puede obtener más conocimiento que aquel que está reflejado a través del modelo de datos de su estructura, ya que una consulta no implica *razonamiento*, sino

[12] https://www.w3.org/TR/owl2-syntax/
[13] https://www.w3.org/TR/owl2-xml-serialization/
[14] https://www.w3.org/TR/owl2-mapping-to-rdf/
[15] https://www.w3.org/TR/owl2-manchester-syntax/
[16] https://www.w3.org/TR/turtle/
[17] https://dvcs.w3.org/hg/rdf/raw-file/default/rdf-turtle/n-triples.html#

tan solo la verificación de una interpretación (aquella definida por la relación entre los conceptos). Mientras que en una base de conocimiento, que actúa basada en la lógica descriptiva, el conocimiento se genera por acción de los elementos de Abox que indican la validez de una aseveración en base a los individuos instanciados en esa base de conocimiento, pero en caso de que se incorporen nuevos individuos, no significa que la regla seguirá dando el mismo resultado solo porque está establecida, sino más bien, se parte del supuesto de que *puede haber elementos en el mundo abierto* que puedan contradecir la aseveración encontrada. De allí que las bases de conocimiento sostenidas por la lógica descriptiva representan un *mundo abierto,* que se modificará a partir de nuevos conocimientos generados por los mecanismos de razonamiento.

Según las definiciones formuladas en las *Recomendaciones de la WC3,* los constructores básicos del lenguaje OWL son las *clases* y las *propiedades.* Las clases son conjuntos de individuos que cumplen *restricciones* para ser considerados miembros de esa clase. Las propiedades son las relaciones que vinculan las clases entre sí, denominadas *ObjectProperties* en OWL, y las relaciones entre las clases y tipo de dato o valor RDF, denominadas *DatatypeProperties* en OWL.

Las clases proporcionan un mecanismo de abstracción para agrupar recursos con características similares y se definen mediante *descripciones de clase*, que se pueden combinar en *axiomas de clase*. OWL considera seis tipos de *descripciones de clase.* El primero de ellos corresponde a la descripción de la clase mediante un *identificador de clase* (o referencia URI), es decir, se describe la clase mediante un *nombre de clase*. Este tipo de clases reciben el nombre de *clases primitivas*. Los restantes tipos describen la clase *por extensión* indicando en cada caso diferentes restricciones sobre esa extensión. En este último caso, las clases que los distintos tipos describen se conocen como *clases definidas*.

El segundo tipo describe una clase mediante la *enumeración* exhaustiva de los individuos que representan las instancias de la clase, es decir, se conoce exactamente cuáles son todas las instancias de esa clase.

También se puede utilizar una *restricción de propiedad* para especificar una clase. Cuando se usa esta descripción de clase se establece una restricción mediante la cual, todos aquellos individuos que la cumplen son instancias de esa clase. Esta restricción describe una clase anónima, o sea, al conjunto de individuos que satisfacen la restricción. Existen varios tipos de restricciones de propiedad:

- *Restricciones de Valor* (*hasValue restrictions)* cuando la restricción se aplica a una clase en particular, permiten especificar relaciones entre individuos de una clase y valores específicos para una propiedad.
- *Restricciones de Cardinalidad (cardinality restrictions)* que pone condiciones al número de relaciones en las que un individuo debe participar en una propiedad determinada, es decir, limitan el número de relaciones entre individuos. Existen tres tipos de restricciones de cardinalidad: cardinalidad mínima (*at least*), cardinalidad máxima (*at most*), cardinalidad exacta (*exactly*). OWL cuenta con constructores particulares para definir cada una de las restricciones de cardinalidad: owl:ObjectMinCardinality, owl:ObjectMaxCardinality y owl: ObjectExactCardinality.
- *Restricciones de Cuantificación (quantifier restrictions)* que incluye:
    - *Restricciones existenciales* (*existential restrictions* o *some restrictions*): que en OWL se representan con el constructor owl: ObjectSomeValuesFrom, y permiten describir clases para la cual todos los individuos deben estar vinculados con al menos un individuo que sea instancia de una clase específica (la indicada en la restricción).
    - *Restricciones universales* (*universal restrictions*): en OWL se expresan mediante el constructor owl:ObjectAllValuesFrom, es análoga al cuantificador universal (para todo) de la lógica del Cálculo de Predicados de primer orden. Según esta restricción, cada instancia de la clase que se describe que tiene la propiedad P, debe estar vinculada por dicha propiedad con un individuo perteneciente a una determinada clase (la que se indica en la restricción). Cabe mencionar que este tipo de restricción amplía la funcionalidad del cuantificador universal pues incluye como individuos de la clase (aquellos que cumplen la restricción) a las instancias que no tienen esa propiedad.

Los otros tres tipos de descripciones de clase denominados *Unión*, *Intersección y Complemento* se refieren al cumplimiento de condiciones lógicas que restringen a los miembros de una clase partiendo de las otras descripciones de clases. Es decir, son clases que se describen a partir de la intersección, unión y complemento de clases, que se representan mediante los constructores owl:ObjectUnionOf, owl:ObjectIntersectionOf y owl:ObjectComplementOf.

Una clase y su definición se vinculan mediante el constructor *owl:EquivalentClass.* Como ejemplo, considere la definición de la clase *CuentaEmisor* la cual representa el conjunto de individuos que son instancias de la clase *Cuenta*, tienen al menos una relación *cuentaEmisorEmiteCorreo* con una instancia de la clase *Correo* y, además, todas las instancias de la propiedad *cuentaEmisorEmiteCorreo* las asocian a instancias de la clase *Correo*. Esta definición se expresa en OWL (utilizando la sintaxis Manchester) como:

Class:CuentaEmisor EquivalentClass Cuenta and (cuentaEmisorEmiteCorreo some Correo) and (cuentaEmisorEmiteCorreo only Correo)

Por otra parte, para la descripción de las reglas de inferencia de la ontología se utilizó el lenguaje denominado *SWRL*. Basado en la combinación de OWL DL, OWL Lite y el sublenguaje Rule Unitary Datalog RuleML[18] del lenguaje de marcado de reglas, en SWRL una regla se expresa de la siguiente manera:

$$(antecedente) \Rightarrow (consecuente)$$

De manera que si el *antecedente* es cierto, entonces el *consecuente* también lo es. El antecedente puede ser uno solo o varios asociados mediante la conjunción "y". Tanto el antecedente como el consecuente pueden representar distintas cosas:

- *Conceptos*, de la forma *C(?x)* donde *C* es una descripción de clase OWL o un rango de datos. Esta expresión indica que lo que está entre paréntesis representa una instancia de la clase C. En particular, ?x indica una variable.
- *Propiedades de datos*, con la sintaxis de *P(?x, ?y)* en la que *P* es una propiedad establecida entre ?*x e ?y*.
- Operadores predefinidos en el espacio de nombre de SWRL. Un ejemplo de este tipo de operadores es *SameAs(?x, ?y)* que evalúa verdadero si *x e y* son el mismo individuo.

Por ejemplo, se puede definir la relación *"es emisor"* mediante la siguiente regla SWRL:

$$usuario\ (?u) \wedge cuenta\ (?cta) \wedge correo(?c) \wedge$$
$$emiteCorreo\ (?cta, ?c) \wedge tieneCuenta\ (?u, ?cta)$$
$$\Rightarrow esEmisorDe\ (?u, ?c)$$

en la que se indica que los individuos representados por las variables *?u, ?cta* y *?c*, instancias de las clases *Usuario*, *Cuenta* y *Correo* respectivamente, se vinculan mediante las relaciones *emiteCorreo* y *tieneCuenta*. Entonces si es cierto que ?*cta*

[18] http://wiki.ruleml.org/index.php/RuleML_Home

emite el correo ?*c* y además que el usuario ?*u* tiene una cuenta ?*cta,* también es cierto que el usuario ?*u* es el emisor de la cuenta ?*c* (*esEmisorDe(?u, ?c)*).

Otro elemento fundamental para el desarrollo de la ontología, es el lenguaje de consulta de las tripletas. SPARQL es el lenguaje de consulta sobre estructuras RDF también definido en las recomendaciones de la W3C. A la fecha W3C aprobó SPARQL 1.1, la última versión de este componente, que integra un conjunto de especificaciones sobre lenguajes y protocolos para consultar y gestionar grafos RDF contenidos en la web o en un almacén RDF.

2.2.3.2 Herramientas de Soporte para el Desarrollo de Aplicaciones Basadas en Ontologías

Son varias las herramientas disponibles para definir la estructura de almacenamiento y recuperación de datos en las ontologías. Las más usuales son *Apache Jena*[19] y *Jena Fuseki*[20].

Apache Jena es un framework de Java de código abierto para crear aplicaciones basadas en ontologías. Es un entorno de programación cuya arquitectura incluye un conjunto de librerías que hacen posible leer y analizar tripletas RDF, navegar grafos RDF, también es posible realizar consultas mediante SPARQL y cuenta con un motor de inferencia propio para razonar ontologías OWL. La Figura 2-2 muestra la arquitectura de esta herramienta.

Allí se observa el API para trabajar las tripletas RDF, una API para ejecutar consultas en SPARQL sobre las tripletas RDF, y la API para OWL que permite manejar las ontologías escritas en ese lenguaje. Apache Jena presenta varias opciones para el almacenamiento de los datos: por una parte provee una API para el almacenamiento en memoria (Store API); también es posible almacenar las tripletas en la TDB (Triple Database); en base de datos SQL (API SQL DB) si se requiere almacenamiento persistente de los datos; o recurriendo a servidores externos a los que se accede mediante Fuseki, que hace posible presentar los datos RDF y realizar consultas SPARQL sobre el protocolo HTTP.

Jena Fuseki es un servidor de almacenamiento de tripletas RDF que utiliza el protocolo HTTP, de acceso libre y de fácil instalación. Además, al ser parte de

[19] https://jena.apache.org/
[20] https://jena.apache.org/documentation/fuseki2/

Apache Jena, se plantea como una estructura amigable para la ejecución de consultas sobre las tripletas RDF.

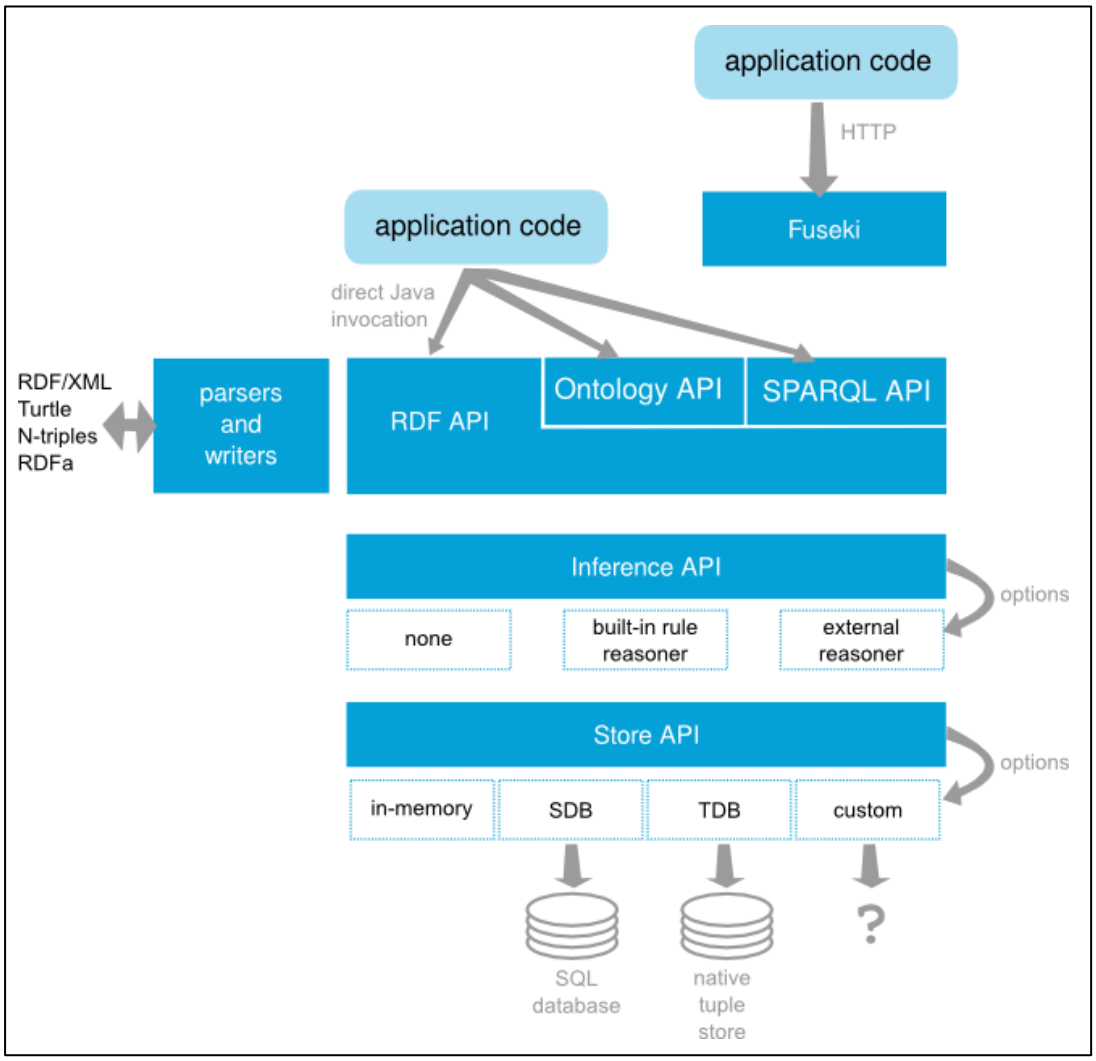


Figura 2-2: Arquitectura de Apache Jena[21]

Fuseki permite dos modos de almacenamiento: persistente (se almacena en disco) y volátil (se almacena en memoria). Esta última característica es de mucho interés en la presente tesis en la que se procesan datos provenientes de una evidencia digital, y por lo tanto, es prioritario no exponerlos a una copia indebida en dispositivos digitales agregados. También se debe mencionar las opciones de ejecución que permite Fuseki: como un servicio de un sistema operativo, como una aplicación web basada en Java o como un servidor independiente que se ejecuta en modo local.

En esta tesis, se utilizó Jena framework como arquitectura de procesamiento de la ontología propuesta, OntoFoCE, en la que se basa la aplicación ObE Forensics.

[21] https://jena.apache.org/about_jena/architecture.html

Las metodologías y herramientas introducidas en este apartado han sido utilizadas en el desarrollo de OntoFoCE y de ObE Forensics. En el Capítulo 3 de esta tesis, se describe el desarrollo de OntoFoCE guiado por la metodología Methontology y utilizando las herramientas descriptas: OWL, SWRL, SPARQL y Protégé.

En tanto, el capítulo 4 presenta ObE Forensics, la aplicación basada en OntoFoCE desarrollada utilizando Jena Framework en la cual se utiliza TDB para el almacenamiento de las tripletas. Finalmente, en el capítulo 5 se detallan los procesos de validación de la representatividad de OntoFoCE, la validación integral de la ontología, así como la validación de la herramienta ObE Forensic realizada por los usuarios expertos.

2.3 Objeto de Estudio: El Correo Electrónico

El correo electrónico es un servicio de red que permite el envío y recepción de mensajes entre un grupo de usuarios. Su nombre proviene de la analogía con el correo postal, en donde también existen remitentes y destinatarios, y se transmiten mensajes empaquetados a través de una estructura de comunicación compleja conformada por múltiples centros de distribución vinculados entre sí.

El correo electrónico (o e-mail, de su nombre en inglés electronic mail) fue creado en 1961 cuando en el MIT (Massachusetts Institute of Technology) se establece un sistema de comunicación entre los distintos usuarios que accedían a una misma computadora desde varias terminales remotas. Luego, en 1971, Ray Tomlison diseña un software de comunicación para ARPANET (Advanced Research Projects Agency Network) incorporando el símbolo "@" para identificar el dominio o servidor que administra la cuenta del usuario destinatario.

Manteniendo el modelo de un *correo postal común*, el servicio de correo electrónico consiste en el envío de una "correspondencia digital" que va pasando por diferentes "oficinas de distribución de envíos" hasta llegar al destinatario final. Esta correspondencia mantiene siempre oculto su contenido, y señala con claridad los datos del destinatario y del remitente. Durante este servicio de envío, cada oficina de distribución va agregando a la correspondencia un "sello de entrada/salida" como registro de paso.

Tomando este ejemplo como analogía, se entiende que un correo electrónico responde al mismo esquema que la correspondencia en papel en donde las oficinas de distribución de envíos son los servidores que van recibiendo el correo electrónico y lo redireccionan para que siga su camino, según los datos de remitente y destinatarios que figuran en el "encabezado" del correo, y en este último, se van registrando los pasos por los distintos servidores de distribución.

Obviamente, el correo electrónico mantiene la característica de confidencialidad del contenido de la correspondencia y de inviolabilidad hasta el momento en que llega a manos de su destinatario.

En el marco de este trabajo se define al mismo en función de los elementos necesarios para la realización del análisis forense, es decir, *un correo electrónico es un documento digital que consta de dos partes: a) una cabecera que contiene información sobre el proceso de transmisión que se desarrolla con identificación de las cuentas intervinientes y los distintos servidores en que el correo se fue almacenando durante la transmisión; y b) un cuerpo que contiene el mensaje que se transmite más los archivos adjuntos que opcionalmente integran el mensaje.*

Vale indicar que aún contra los pronósticos usuales respecto de que el correo electrónico ha caído en desuso, los estudios demuestran que en realidad, esta herramienta se está transformando para adaptarse a las nuevas necesidades que las redes sociales no pueden resolver aún.

Según indica Anthes[22] para 2019 se esperaba que circulen por la red 3.2 millones de mensajes por segundo (entre reservaciones de hotel, avisos de reuniones, saludos de amigos, diseños de productos, recibos, filtraciones, quejas, solicitudes de ayuda y, por supuesto, spam).

Por una parte, las aplicaciones clientes de correos electrónicos han ampliado sus funciones agregando calendarios, repositorios de datos y otros servicios web como el de traducción de textos, que transforman el "gestor de correos" en un "servicio de asistencia personal al usuario".

Por otra parte, la portabilidad de estas aplicaciones clientes dirigidas a dispositivos móviles, hizo que el correo electrónico ampliara su radio de influencia y de presencia en la vida cotidiana de las personas, particularmente en aquellas

[22]Anthes G., *You've Got Mail!*, se puede leer en http://delivery.acm.org/10.1145/3220000/3213776/p18-anthes.pdf?ip=190.31.218.4&id=3213776&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1549741305_aa7c6ee345bbcc267546a4f25b13d88a

actividades de tipo comercial o laboral que requieren de alguna instancia de resguardo acerca de lo tratado con otro usuario.

2.3.1 Estructura de un Correo Electrónico

Desde el punto de vista técnico, el correo electrónico se ajusta a la norma RFC 822[23] (y sus modificatorias) que contiene los estándares de formato para mensajes de texto. Sin competencia respecto del contenido del correo, esta norma señala cómo debe estructurarse la cabecera de un correo electrónico, y dispone la forma en que el servicio de envío agrega información sobre el servidor de mail en la cabecera del correo.

El correo electrónico es un servicio de red que permite a los usuarios enviar y recibir mensajes mediante redes de comunicación electrónica. Puede decirse que un correo electrónico es un paquete o bloque de datos que circula por la red, desde un origen a un destino, utilizando para ello el software y hardware necesario para el proceso de transmisión, y conteniendo todos los datos que permiten identificar el camino recorrido entre el punto de origen y el de destino.

Al igual que en el "correo postal", el correo electrónico utiliza "buzones" intermedios (o servidores de correo) cuya función es el almacenamiento y reenvío de modo que no es necesario que el origen y el destino del correo se encuentren conectados simultáneamente ni de manera directa. En Internet, existen multitud de servidores (servidores de correo corporativo o de correo gratuito, servidores de proveedores de servicios de internet) que hacen posible la comunicación entre una cuenta del emisor y una cuenta del receptor.

Además del contenido (o cuerpo) del mensaje, el correo puede contener adjunto un archivo. Cuando se realiza una pericia, interesa identificar los datos de los usuarios que enviaron y recibieron el correo, y cuál fue el proceso o camino de transmisión seguido desde el emisor al receptor, así como toda la información referente al contexto (equipos utilizados, clientes de correo utilizados, copia de los correos, asunto del correo, etc.), de modo de poder responder a las preguntas del Juez expresadas en los puntos de pericia.

[23] RFC (*Request for Comments)* es una serie de publicaciones técnicas que describen diversos aspectos del funcionamiento de internet, fijando protocolos, estándares y procedimientos avalados por el IETF (Internet Engineering Task Force) que es una organización internacional abierta de normalización de la ingeniería de Internet.

A los fines del presente trabajo, se abordará en profundidad aquellos aspectos del correo electrónico que resulten de interés para un análisis forense. Así, tomando como base la tipificación propuesta por (Banday, 2011) para el análisis forense de un correo electrónico resulta necesario identificar tres componentes: los actores participantes en la transmisión, la arquitectura lógica y la organización interna de un correo electrónico.

Si bien la comunicación de un correo electrónico requiere de un emisor y un receptor del mensaje, no son éstos los únicos partícipes de la transmisión. Existen procesos responsables de sostener el servicio –denominados ***actores***- que operan internamente durante la transmisión, como ser: procesos de notificación de fallas en el envío/recepción de los mensajes, procesos de gestión del servicio de mensajes (retransmisores y Gateway), procesos de gestión de políticas de autorización de dominios (comunicación con los servicios ISP, servicios de web mail), entre otros.

El proceso de transmisión de datos involucrado en un correo electrónico requiere de un conjunto de servicios y protocolos de comunicación integrados en una plataforma de software y hardware que conforman la arquitectura de procesamiento.

Un correo electrónico es manejado por un mínimo de cuatro equipos distintos: el equipo emisor, el servidor de correo del remitente, el servidor de correo del receptor y el equipo receptor. En todos ellos, el proceso de transmisión *deja una huella* del correo emitido, que se encuentra en la *cabecera del correo.* Los MTA (Message Transfer Agents), agentes de transferencia del mensaje, responsables de registrar en la cabecera del mensaje los datos referidos al servidor en donde se almacenó el correo durante el camino de transmisión, añaden una etiqueta de identificación en la cabecera del correo cada vez que el mail ingresa a un servidor.

El proceso de transmisión también requiere de un servidor de DNS (Domain Name System), responsable de "traducir" el nombre de dominio escrito por el usuario en las correspondientes direcciones IP asociadas con los equipos conectados a la red, esto con el propósito de poder localizar y direccionar estos equipos en todo el contexto de la WWW. Además de los dispositivos de emisión/recepción del correo utilizados por el usuario que emite/recibe el correo, sean éstos una computadora, un celular o cualquier otro equipo utilizado para el acceso al sistema de correo electrónico, se requiere de entidades no visibles al usuario que actúan en el proceso de creación, envío, transmisión, entrega y lectura del e-mail, como por ejemplo: componentes de la capa inferior de TCP/IP (routers, bridges) y de la capa

de aplicación de TCP/IP (nodos de e-mail) así como los diversos protocolos utilizados (SMTP, HTTP, etc.).

Según se muestra en la Figura 2-3, básicamente el envío del correo requiere de 7 pasos:

1. Desde la cuenta del emisor (*pedro@dominio1.com*) se compone el correo y mediante el protocolo SMTP (Simple Mail Transfer Protocol) se realiza el envío al servidor de correo de la cuenta de envío.
2. El servidor de correo de la cuenta del emisor (*smtp.dominio1.com*) envía la consulta al servidor DNS sobre la dirección IP (Internet Protocol) correspondiente a *dominio2.com.*
3. El servidor DNS responde informando la dirección IP correspondiente al servidor de correo *smtp.dominio2.com.*
4. Luego, mediante el protocolo SMTP, se realiza el envío del correo al servidor de correo de la cuenta del receptor (*smtp.dominio2.com).*
5. También mediante el protocolo SMTP, el servidor de correo de la cuenta del receptor recibe el correo en cuestión.
6. El correo recibido queda almacenado en el servidor *smtp.dominio2.com* hasta que el usuario accede al correo.
7. Mediante el protocolo POP3 (Post Office Protocol Ver.3), el usuario baja el correo desde la cuenta *jose@dominio2.com* .

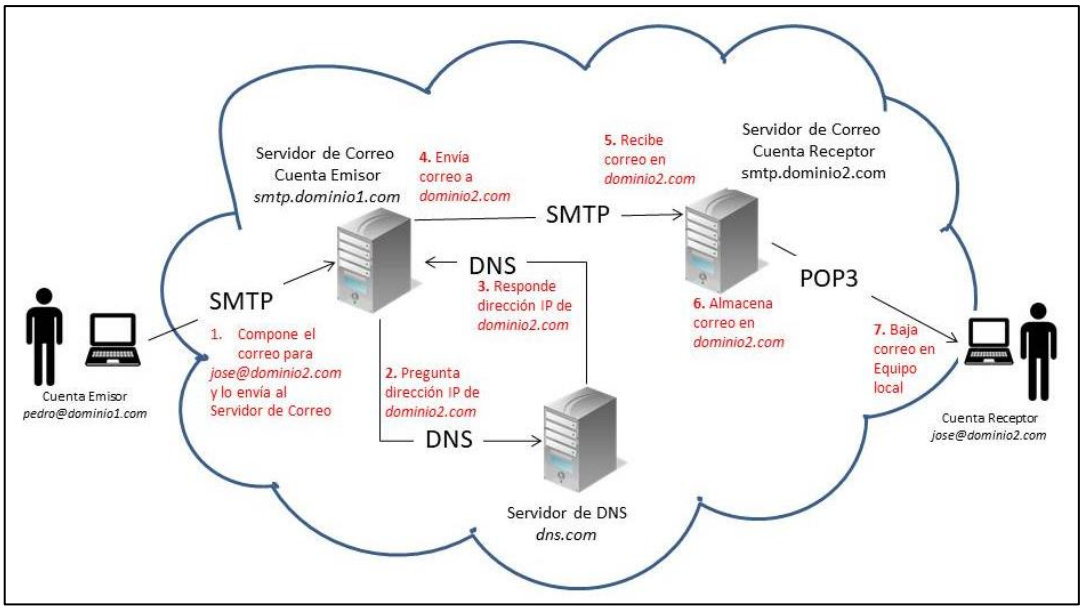


Figura 2-3: Proceso de transmisión de un correo electrónico

Se debe aclarar que la comunicación entre el servidor de correo de la cuenta del emisor y el servidor de correo de la cuenta del receptor no siempre es tan directa, la

mayoría de las veces se requiere de servidores intermedios que hacen la vinculación entre un extremo y otro de la transmisión. Así, en la cabecera se van registrando los datos de cada servidor en donde el correo se va almacenando durante el proceso de envío.

Por otra parte, a los efectos del análisis pericial, se considera que la función del servidor DNS no es relevante ya que sobre éste solo se realiza la consulta de identificación de la IP de los servidores, y el correo no se almacena en el mismo.

Se puede decir que un correo electrónico es un archivo digital con componentes que se visualizan de determinada manera, según sea el cliente de correo utilizado. La Figura 2-4 muestra el detalle de la cuenta del emisor, las cuentas que actúan como receptoras, asunto, cuerpo y archivo adjunto de un correo electrónico visto en el cliente de correo *Thunderbird*[24].

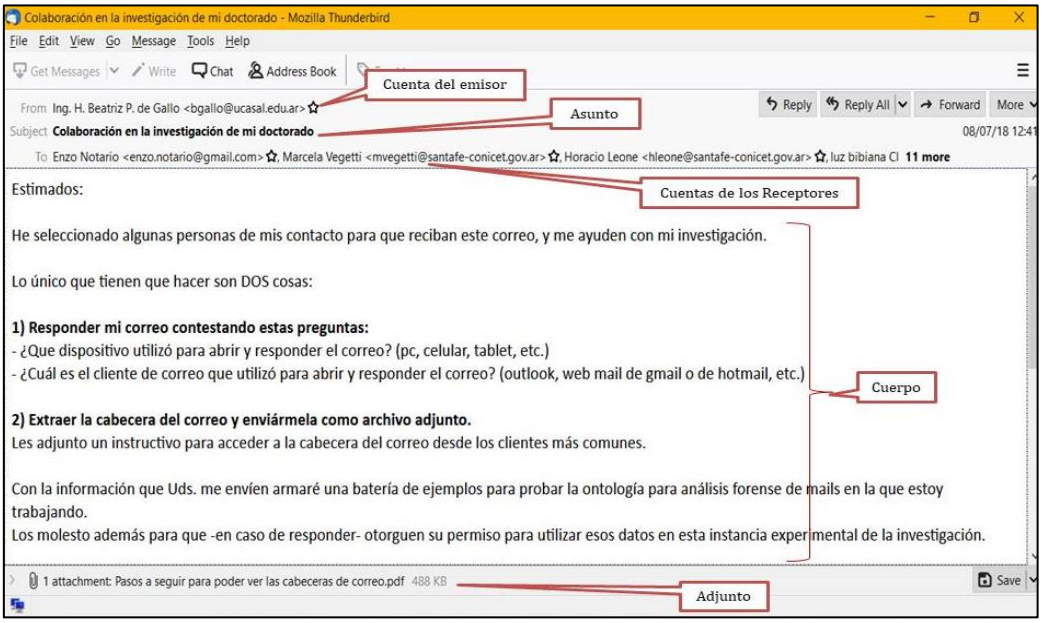


Figura 2-4: Partes de un Correo Electrónico

Por otra parte, la norma RTF 822 define al correo electrónico como un archivo digital de texto plano que contiene dos componentes muy bien identificados: un *encabezado o cabecera*, con la información pertinente al recorrido que realiza desde la cuenta del emisor a la cuenta del receptor, y un *cuerpo* que corresponde al contenido del envío. Obviamente, solo contendrá información sobre el envío aquella cabecera correspondiente a un correo *recibido*, ya que en ella se van almacenando los datos de cada servidor por donde pasa el correo durante la transmisión, hasta que llega al equipo receptor.

[24] https://www.thunderbird.net/es-ES/

La cabecera no es visible a simple vista, se debe acceder al menú del cliente de correo para llegar a ella, en el caso del ejemplo de la Figura 2-4 se toma uno de los correos recibidos como respuesta, y en la Figura 2-5 se muestra una vista parcial de la cabecera correspondiente. Allí se indica en letra cursiva y en color rojo las partes del correo: cuenta emisor, cuenta receptor, asunto y cuerpo del mensaje.

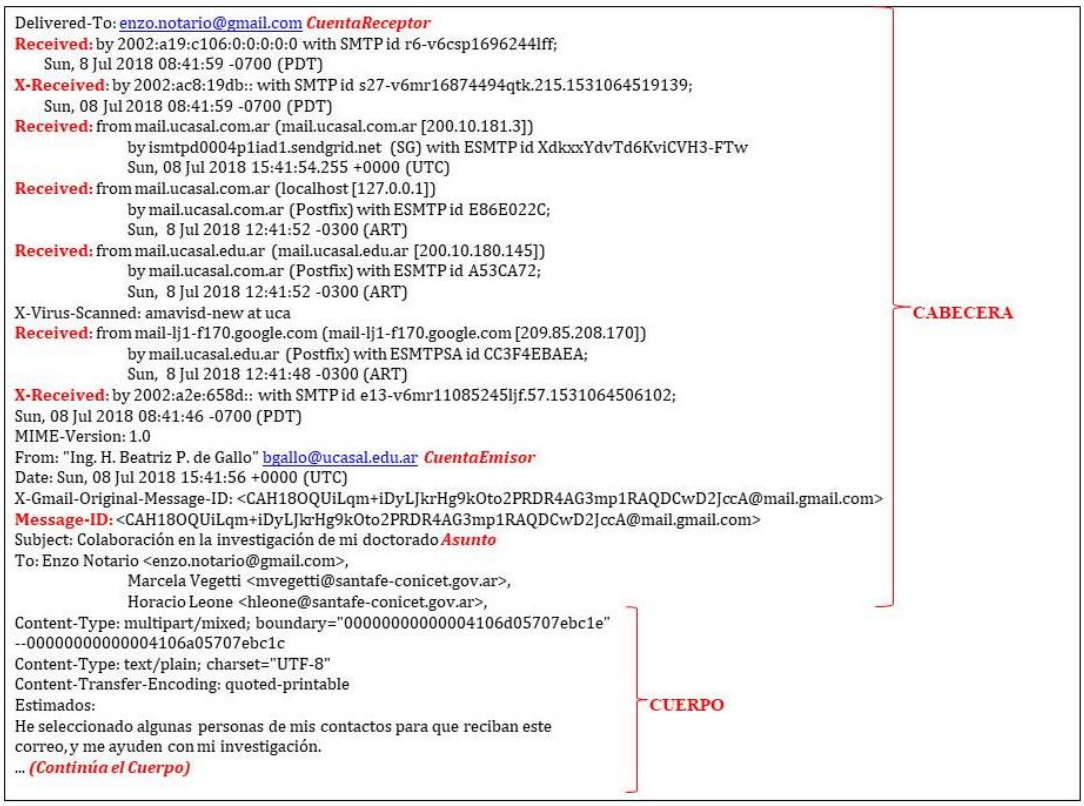


Figura 2-5: Cabecera de un Correo Electrónico Recibido

Es importante destacar algunas características de la cabecera de los mensajes electrónicos:

- Generalmente todos los campos del encabezado necesarios son generados por la interfaz de correo.
- Algunos son opcionales y pueden ser añadidos por el usuario, mediante las opciones de configuración de la cuenta y del cliente de correo que se utilice, como por ejemplo: el *alias* de usuario o el pie de firma del correo.
- El cuerpo se ubica al final del archivo de la cabecera, y se mantiene inalterado durante la transmisión, ya que los servicios de MTA incorporan los datos de identificación de cada servidor como primera línea de la cabecera. Es decir, por cada servidor de paso, se agrega una línea de identificación que se coloca al principio del archivo.

  Al realizar el análisis del encabezado del mensaje debe tenerse presente que:

- El orden de lectura se realiza de abajo hacia arriba, es decir, la primera dirección IP que se encuentra señala la dirección del último servidor visitado.
- Las líneas de la cabecera que comienzan con la palabra *Received* y *X-Received* brindan información para determinar la ruta del mensaje. En esta línea figura habitualmente la dirección IP del equipo/servidor por el que pasó el correo y puede ocurrir que allí figure además un *hostname* (Nombre de Dominio), ambos datos son de interés para la pericia.
- El identificador del Mensaje –*Message ID*- es asignado por el cliente de correo que creó el mensaje. Este identificador permite buscar un determinado mensaje entre los registros de uno o varios servidores.

2.3.2 Proceso de Transmisión y su Trazabilidad

En base a la información agregada a la cabecera durante la transmisión, es posible establecer la ***trazabilidad*** de un correo electrónico.

La norma ISO 9000:2015[25] define trazabilidad como la "*capacidad para seguir el (desarrollo) histórico, la aplicación o la localización de un objeto; al tratarse de un producto o servicio, la trazabilidad puede estar relacionada con el origen de los materiales y las partes, el histórico del proceso y la distribución y localización del producto o servicio después de la entrega*". La principal ventaja que reporta la trazabilidad (o logística inversa) es poder conocer a ciencia cierta la procedencia y la historia que atañe a un producto.

Así, mediante la trazabilidad se puede probar la existencia del correo electrónico recibido en una cuenta. La reconstrucción del camino de inverso de un correo electrónico permite avalar el carácter probatorio de este documento digital y refuerza la capacidad de no repudio del mismo.

Considerando el *camino* que realiza un correo electrónico desde su emisión hasta la recepción por parte del destinatario, ocurren diferentes procesos que se van desarrollando durante la transmisión. De esta secuencia de acciones, interesan en particular aquellas que pueden impactar en el *proceso de transmisión del envío.*

Así, se plantea un principio básico para representar la autenticidad del correo electrónico que dice:

[25] Se puede consultar en https://www.iso.org/obp/ui/es/#iso:std:iso:9000:ed-4:v1:es (vigente al 22/02/2018)

Un correo electrónico es auténtico cuando se identifican los datos del remitente (cuenta de correo y dirección IP), la trazabilidad del mismo (diferentes dispositivos que intervienen en la transmisión) y los datos del destinatario (cuenta de correo y dirección IP).

Este principio es el que toma para representar el proceso de transmisión y su trazabilidad mediante una ontología.

2.4 Forensia Digital

Las tecnologías de la información y de las comunicaciones (TIC) han invadido todas las áreas de la sociedad. El quehacer diario de las personas está vinculado de una forma u otra a las tecnologías informáticas, ya sea como usuario final mediante un celular, una PC o una Tablet, o como entidad que cumple un rol social, económico o comunitario, en el cual exige a sus integrantes una interacción virtual.

Esta inclusión de las tecnologías en la sociedad ha posibilitado importantes mejoras en las actividades en general, notándose su mayor impacto en el ámbito de las comunicaciones interpersonales mediante las redes sociales, la mensajería instantánea y el correo electrónico.

Pero de igual forma, así como ha favorecido la vida de las personas, también se utiliza para el desarrollo de actividades delictivas, en las cuales las TIC participan con idéntica fuerza que en el resto de los quehaceres sociales.

Ubicados en el contexto legal, a partir de 1990 surge la necesidad de convocar a *peritos informáticos* para que actúen como auxiliares de la justicia cuando se presenta una *prueba digital.* Con el tiempo, y la evolución de las tecnologías, esta primera acción del profesional informático que solo hacía un aporte técnico pasó a convertirse en una rama de la disciplina informática con entidad propia.

Proveniente de la Informática aplicada, el desarrollo de la *Informática Jurídica* tuvo una variante distintiva cuando se abordaron las pericias informáticas. Así, surge primeramente la *Informática Forense* y se transforma en lo que hoy se conoce como *Forensia Digital.*

A partir del año 2000, comienzan a surgir los ataques a la seguridad informática, lo que produce un crecimiento en las normas y procesos necesarios para atender la problemática de hacking e intrusión sobre los sistemas informáticos. Ya en el 2005, con la incorporación de aplicaciones web, se hace más crítica la cuestión de la

seguridad y resguardo de los datos, al punto de tener que generar esquemas de seguimiento y búsqueda de vulnerabilidades. Aparecen nuevas formas de la seguridad informática (hacking ético, por ejemplo) y allí se formaliza la *Forensia Digital*, para dar una respuesta al análisis de los incidentes de seguridad informática considerando no solamente las aplicaciones informáticas y los datos que se procesan, sino además los artefactos forenses de carácter digital (se denomina así a los distintos componentes en los que reside o constituye la evidencia digital, sea éste un dispositivo de hardware o software).

Por su parte, la *Informática Forense* toma para sí las herramientas y métodos de la *Forensia Digital* y le agrega algunos de los procedimientos propios de la criminalística como la *cadena de custodia*.

En el marco de esta tesis, se toma como definición de *Informática Forense* la propuesta por el Grupo de Investigación en Sistemas Operativos e Informática Forense de la UFASTA (Di Iorio et al., 2017) que dice: *La Informática Forense es considerada una rama de las ciencias forenses que se encarga de adquirir, analizar, preservar y presentar datos que han sido procesados electrónicamente, y almacenados en un medio digital. Es el uso de las Tecnologías de la Información para recuperar evidencia digital.*

Asimismo, se adhiere a la definición de *Forensia Digital* propuesta por (Zuccardi et al., 2006) que dice: *Forma de aplicar los conceptos, estrategias y procedimientos de la criminalística tradicional a los medios informáticos especializados, con el fin de apoyar a la administración de justicia en su lucha contra los posibles delincuentes o como una disciplina especializada que procura el esclarecimiento de los hechos (¿quién?, ¿cómo?, ¿dónde?, ¿cuándo?, ¿por qué?) de eventos que podrían catalogarse como incidentes, fraudes o usos indebidos bien sea en el contexto de la justicia especializada o como apoyo a las acciones internas de las organizaciones en el contexto de la administración de la inseguridad informática.*

La Forensia Digital se aplica principalmente en dos áreas: en el ámbito de la justicia mediante las pericias informáticas con la inclusión de las evidencias digitales, y en el ámbito empresarial/institucional cuando se analizan fallas de seguridad y acciones de intrusión indebidas. A los fines del presente trabajo, solo se considerará la Forensia Digital desde el punto de vista del rol de auxilio de la justicia en el proceso de validez de la prueba digital.

2.4.1 Análisis Forense Digital

En sus inicios, las pericias informáticas se realizaban recurriendo a las normas y protocolos que la Informática había desarrollado para el ámbito de la Seguridad de los Sistemas de Información. Las primeras normas de este tipo aparecen en 1995, con la BS 7799 de BSI (British Standards Institution) que establece un conjunto de buenas prácticas para la gestión de la seguridad de los sistemas de información empresariales. Esta norma fue evolucionando con el aporte de otras instituciones internacionales, hasta conformar lo que hoy se conoce como Norma ISO 27000:2013[26], referida a un Sistema de Gestión de Seguridad de la Información (SGSI).

Es importante resaltar que OntoFoCE se desarrolló considerando el marco legal de la justicia argentina, y si bien son varios los autores que abordan la definición del proceso de análisis forense digital, conjugando cuestiones propias de la criminalística con protocolos y normas de la ingeniería, conviene considerar aportes de investigadores argentinos pues referencian el proceso pericial que se sigue en dicho país.

En particular, se toma la propuesta del Grupo de Investigación sobre Forensia Digital de la Universidad FASTA (Di Iorio et al., 2017), quienes incorporan componentes de la Ingeniería de Software y proponen el Proceso Unificado de Recuperación de Información (PURI). Se selecciona esta metodología pues conjuga aspectos informáticos y criminalísticos que ordenan el procedimiento pericial, y en ese marco, se entiende mejor la utilización de una herramienta como la propuesta en esta tesis, en cuanto a dotar a la misma de las características necesarias para cumplir el condicionamiento de "principios científicos y técnicos" requeridos por el derecho procesal argentino[27] a las herramientas forenses.

La Figura 2-6 esquematiza el proceso PURI, el cual consta de cinco fases que se describen a continuación:

La ***Fase de Relevamiento*** abarca la investigación para conocer el caso e identificar los posibles objetos de interés, para considerar la documentación legal y técnica y la infraestructura de IT con que se va a trabajar.

---

[26] https://www.isotools.org

[27] Art. 477 del Código Procesal Civil y Comercial de la Nación (CPCyC), Ley Nacional 17454, 1967.

La ***Fase de Recolección*** abarca las acciones y medidas necesarias para obtener los equipos físicos, y/o las posibles fuentes de datos, sobre los cuales se deberá trabajar posteriormente.

La ***Fase de Adquisición*** abarca todas las actividades en las que se obtiene la imagen forense[28] del contenido que se analizará.

La ***Fase de Preparación*** involucra las actividades técnicas en las que se prepara el ambiente de trabajo del informático forense, la restauración de las imágenes forenses y volcados de datos, junto con su correspondiente validación, y la selección de las herramientas y técnicas apropiadas para trabajar en la extracción y el análisis, de acuerdo al objeto origen, y a las necesidades del caso.

La ***Fase de Extracción y Análisis*** comprende las tareas forenses de extracción de la información de las imágenes forenses, la selección de la potencial evidencia digital, y su análisis en relación al caso y a los puntos periciales o requerimientos de servicio forense.

Finalmente, la ***Fase de Presentación*** comprende el armado de los informes necesarios y la presentación del caso en un juicio o a los solicitantes.

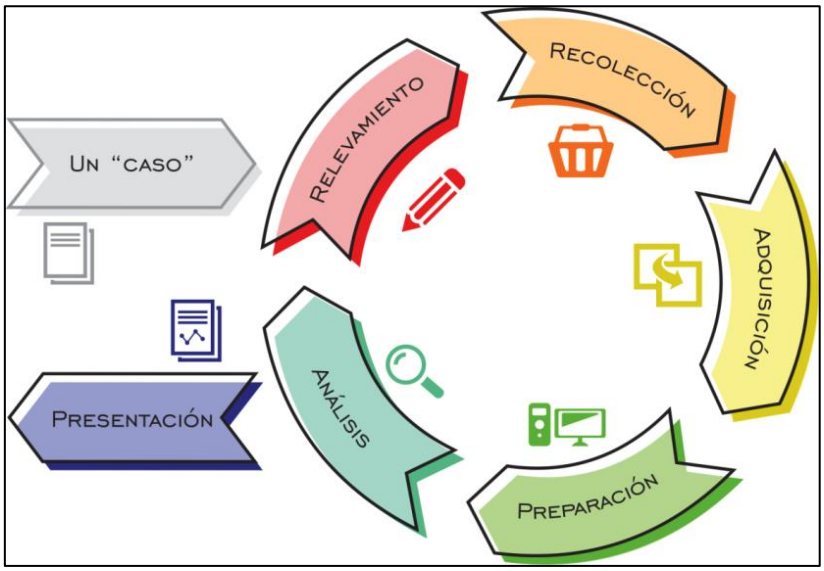


Figura 2-6: Fases del Proceso Unificado de Recolección de Información (PURI).

Fte: (Di Iorio et al., 2017)

Más adelante, en este capítulo, se describe el escenario del procedimiento pericial sobre correos electrónicos, señalando las fases de PURI en las que OntoFoCE puede dar soporte como herramienta para el análisis forense.

[28] Una ***imagen forense*** es una copia exacta, sector por sector, bit a bit, de un medio de almacenamiento. De esta manera, es posible trabajar con la imagen de la misma manera que si se hiciera sobre el original.

Por otra parte, es importante destacar la necesidad de contar con una *metodología científica* para el desarrollo del análisis forense, toda vez que la justicia está demandando la participación de peritos informáticos en la obtención y tratamiento científico de evidencias digitales que se presentan como *prueba* de un hecho, dado que los rastros digitales son cada vez más numerosos en los procesos investigativos.

De igual modo, la justicia demanda a los peritos informáticos un entrenamiento en la materia judicial y criminalística, particularmente respecto de los principios de mantenimiento de la cadena de custodia, no contaminación de la prueba y el uso de criterios de actuación compatibles con el derecho procesal de los distintos ámbitos judiciales.

2.4.2 Los Puntos de Pericia

La legislación argentina establece los procedimientos a seguir para la obtención de evidencia digital. Estos procedimientos varían –en amplitud y profundidad- según se trate del ámbito del derecho que se aborda (Civil, Penal, Laboral, etc.).

Para (Cafferata Nores & García, 2003) la pericia *es un medio probatorio con el cual se intenta obtener, para el proceso, un dictamen fundado en especiales conocimientos científicos, técnicos o artísticos, útiles para el descubrimiento o la valoración de un elemento de prueba*. Su regulación se encuentra definida de manera general en los códigos procesales de forma meramente enunciativa.

Se puede tomar la definición de prueba pericial de (Gilardi & Unzaga Domínguez, 2007): *La prueba pericial consiste en el informe brindado por una persona ajena al proceso, con especiales conocimientos técnicos, y/o científicos sobre la materia en litigio, que a través de un proceso deductivo (de lo general a lo particular), partiendo de sus conocimientos específicos, los aplica al caso concreto y elabora su opinión fundada con los elementos ciertos que surgen de la causa en análisis.*

De una pericia interesa particularmente el *objeto de la pericia* o *los puntos de pericia*, mediante los cuales el Juez define el alcance de la actividad pericial. Estos elementos usualmente se expresan en términos de acción "*verificar… constatar… informar… explicar…"*, en la cual el perito recurre al conocimiento científico de su área, para responder a la solicitud del Juez, atendiendo a las normas y buenas

prácticas de cada disciplina. Valgan como ejemplos los siguientes puntos de pericia sobre correos electrónicos:

- Acceder al equipo y extraer toda información de la que pudieran surgir los correos electrónicos en los que la casilla "xxxxxxx@xxxx.com.ar" figure como remitente o destinatario.
- Informe los correos intercambiados entre las partes en el período indicado.

Los puntos de pericia definen el alcance de la actividad pericial y tienen carácter restrictivo, es decir, el informe pericial debe contener solo las respuestas que el Juez solicita y no pueden realizarse interpretaciones subjetivas o juicios de valor respecto de la evidencia digital. La tarea del perito es solo eso, responder expresamente a lo solicitado en los puntos de pericia y sin exceso. Ahora bien, queda en libertad del perito la selección de las herramientas, técnicas o métodos que utilice para analizar la evidencia, y al respecto, solo se exige la utilización de *métodos científicos* que garanticen el análisis forense desde un procedimiento formal.

2.5 Procedimiento de Realización de una Pericia sobre Correos Electrónicos

La actividad pericial sobre un Correo Electrónico comienza cuando el Juez cita al perito para realizar el acto de posesión en la causa, supuesto que no hay impedimentos que lo inhabiliten para actuar en la misma[29].

El procedimiento de realización de la pericia incluye una serie de pasos a realizar para lograr resultados exitosos que se enuncian siguiendo la metodología PURI.

2.5.1 Fase de Relevamiento

El análisis pericial de correos electrónicos debe realizarse siempre en el correo *recibido,* ya que el correo emitido en sí mismo no garantiza que haya sido recibido por el destinatario. Si bien usualmente en el expediente judicial se incorpora un impreso del correo electrónico a peritar, es necesario acceder al cliente de correo en el que se encuentra éste y –de ser posible- al propio equipo en el cual se recibió el correo, pues la investigación debe realizarse sobre el correo original. Cualquier acción de reenvío del correo bajo análisis impacta en la cabecera del correo,

[29] Relación familiar o comercial con alguno de los imputados/procesados en la causa, incompetencia técnica para realizar la pericia, etc.

agregando los datos de la transacción en el propio correo, y alterando con ello la base de análisis forense de este tipo de documento digital.

Se debe considerar, además, que tanto el proveedor de Internet como el del correo electrónico, pueden ser consultados para rastrear un mensaje. Dado el volumen de tráfico de mensajes que manejan los proveedores de Internet, no suelen almacenar los registros por mucho tiempo, por lo tanto, cuando se efectúa el seguimiento de un correo electrónico se debe consultar al proveedor lo más pronto posible. Desde el punto de vista legal, no existe obligatoriedad de guardar estos registros.

Atendiendo a la condición de *documentación epistolar* que se le asigna legalmente al correo electrónico, es importante salvaguardar la privacía del mismo. (Bender, 2017) dice que para preservar la inviolabilidad de la correspondencia epistolar es importante que la búsqueda de información en la casilla de correo se realice con la supervisión de la afectada y que se individualice con la mayor precisión posible los documentos que deben buscarse, evitando acceder a otros que no se encuentren directamente vinculados con el objeto de la litis.

2.5.2 Fase de Recolección

El procedimiento para extracción y preservación del material o documento digital a peritar, depende de varios escenarios vinculados al área del derecho en el que se desarrolle la pericia. Si es el Fuero Penal, los procedimientos son sumamente rigurosos puesto que está en juego la libertad y/o la vida de las personas. En el Fuero Comercial o Laboral es distinto, usualmente la pericia de la documentación digital consiste en certificar y validar la prueba digital presentada. Es en este último foro –el laboral- en donde mayormente se solicitan pericias de correos electrónicos, de modo que se abordará el procedimiento bajo el supuesto de la solicitud de pericia en dicho contexto.

Para poder realizar el análisis forense del correo electrónico, es necesario obtener la *cabecera* del correo electrónico. El conjunto de datos que permiten que un correo electrónico tenga validez como evidencia digital se encuentra en su encabezado. Allí figura la información relativa al emisor del correo, fecha de envío, camino de recorrido por los servidores intermedios hasta llegar a destino y los datos referidos al destinatario del correo electrónico.

Se debe garantizar la posibilidad de la otra parte del juicio de poder inspeccionar el procedimiento seguido para la obtención de la prueba. Esto se hace a través de un procedimiento adecuado de extracción y conservación denominado *cadena de custodia*. Lo ideal es contratar los servicios de un escribano (o contar con la presencia de un oficial de justicia) para que dé fe del procedimiento seguido para su extracción, además que la actividad sea realizada por un experto informático conocedor de la importancia y características que debe cumplir un archivo digital para que luego se presente como evidencia digital.

Por otra parte, para garantizar que la prueba no ha sido alterada desde su extracción y depósito hasta la entrega en el juzgado, se recurre a herramientas de cifrado que proporcionan una función *HASH*[30], de tal forma que, a través de la secuencia de caracteres resultante, se puede comprobar que el fichero extraído por el perito y el depositado en el juzgado son idénticos. Así, una vez finalizado el copiado forense, el perito debe aplicar la función hash de dicha copia y entregarla al juzgado para su resguardo. Usualmente ese archivo se entrega con copia para las partes, y el perito se apropia de una de ellas para realizar la pericia.

2.5.3 Fase de Adquisición

Esta fase incluye todas las actividades dirigidas a la obtención de la *imagen forense* de la evidencia. Y esto cobra especial interés cuando la prueba digital incluye logs de transacciones, imágenes, audios, bases de datos o datos de estructuras de archivos poco comunes. Para el caso de los correos electrónicos, no es necesario obtener una imagen forense y a partir de ella realizar el análisis, sino que, se debe acceder a la cuenta en la que reside el correo electrónico que se aporta como prueba, y desde allí realizar la pericia.

La particularidad del análisis forense de correos electrónicos radica en que éste se realiza sobre la cabecera del mensaje, de allí que la fase de *adquisición* toma otra forma.

[30] Una *función criptográfica hash* -usualmente conocida como "hash"- es un algoritmo matemático que transforma cualquier bloque arbitrario de datos en una secuencia de caracteres única. Si se cambia un solo caracter del texto original, la secuencia resultante será diferente. Las funciones hash permiten identificar unívocamente a un conjunto de datos digitales, sea éste un documento, un video, etc.

Para seguir el rastro de los mensajes de correos electrónicos se deben analizar las cabeceras del mensaje. (Darahuge & Arellano González, 2016) indican cuales son las funciones que cumple el encabezado del correo electrónico:

- Indican a los servidores de correo donde entrega el mensaje.
- Indican a las aplicaciones lectoras de correo electrónico como procesar el contenido de los mensajes de correo.
- Ofrece un registro de la ruta seguida por el mensaje desde su origen a su destino.

La cabecera del correo electrónico figura como metadato[31] en el archivo digital que lo contiene, es decir, no está visible directamente, sino que debe obtenerse mediante al acceso a las propiedades del archivo digital del correo, que cambian según sea el cliente de correo utilizado para la gestión del mismo. Así, hay dos modos generales de obtener los datos de la cabecera del correo electrónico:

- Seleccionar "Código Fuente, o Mostrar Original" por medio de opciones o preferencias en la barra de herramientas de la página Web del cliente de correo (muy común en el caso de los servicios de correo con dominio gratuito).
- Si se trata de un gestor de correo como ser Outlook[32] o Thunderbird, se debe abrir el mensaje, seleccionar opciones y luego Encabezados de Internet o Fuente de mensaje, dependiendo de la versión que se trate.

2.5.4 Fase de Extracción y Análisis

Sobre la base de lo definido en (Rivetti, Araoz Fleming, Parra de Gallo, & Leone, 2016) en este apartado se formulan los pasos a seguir durante el procedimiento de obtención de la cabecera de un correo electrónico, resguardando todos los requerimientos señalados en el apartado anterior, a fin de que ese documento digital tenga validez como evidencia en una litis. La Figura 2-7 muestra el diagrama de actividades del procedimiento correspondiente, el cual es posteriormente explicado. A continuación se detalla cada paso del procedimiento indicado en la Figura 2.7.

1) Identificar la cuenta de correo a analizar, y determinar si el proveedor de la cuenta es un servicio de dominio gratuito o se trata de una cuenta corporativa

[31] Los ***metadatos*** (metadata) son campos de texto incrustados en los archivos con información adicional sobre el mismo (fecha de creación, resolución, tamaño, fecha de modificación, autor, etc.). Usualmente son transparentes al usuario y no están visibles a simple vista.

[32] https://products.office.com

con dominio propio. En este último caso, se abre la posibilidad de buscar una copia del correo en el servidor de correos corporativos de manera más fácil que en el caso de un servicio de dominio gratuito.

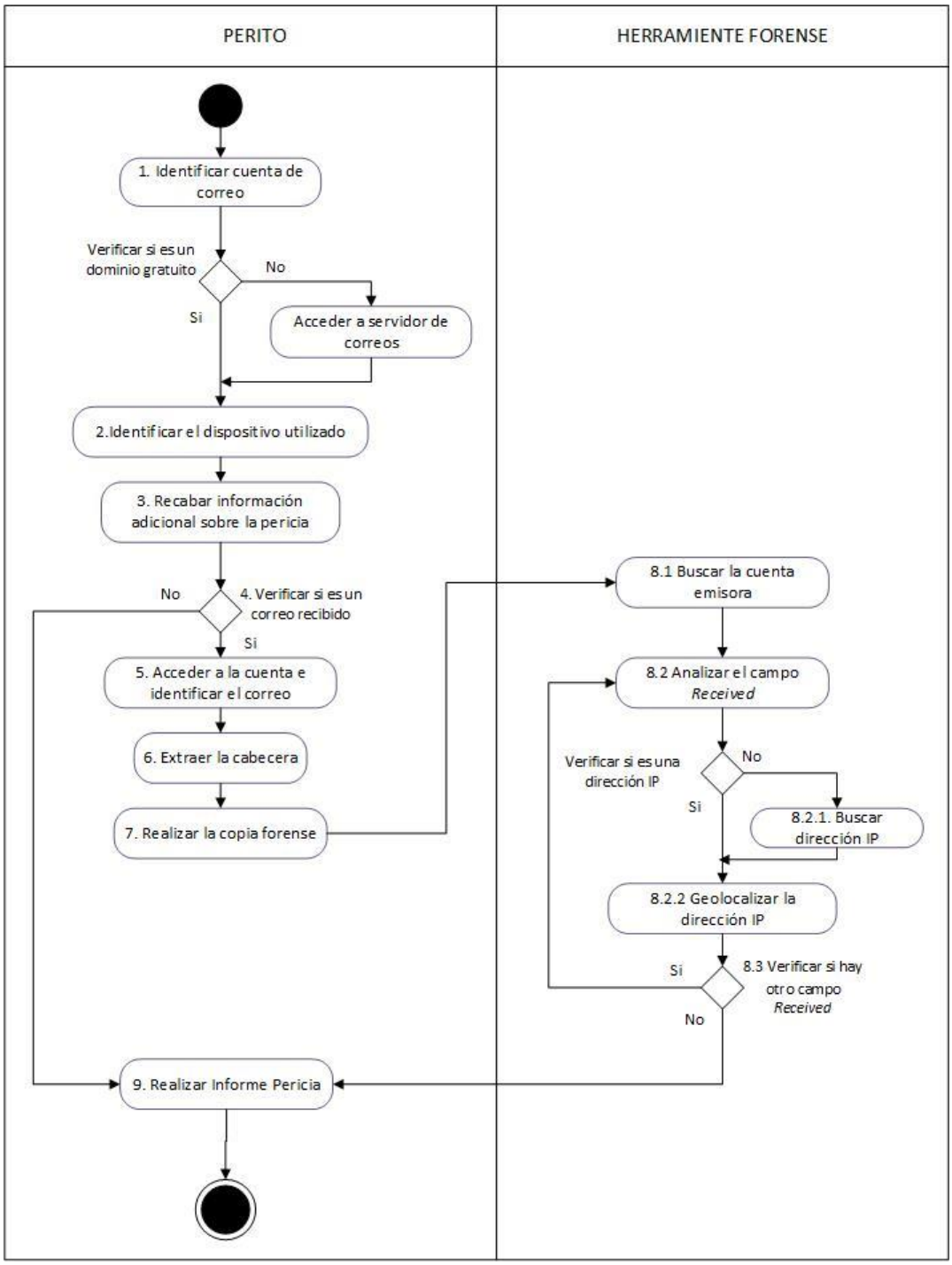


Figura 2-7: Procedimiento de Obtención y Análisis de la cabecera de un correo electrónico

2) Identificar el dispositivo utilizado (PC, Celular, Tablet, servidor de correo, etc.) en el cual reside el correo electrónico aportado como prueba.

3) Recabar información adicional sobre la pericia, es en este momento donde se obtiene la información auxiliar para el informe pericial: dirección física del

equipo, cliente de correo utilizado para gestionar la cuenta y otra información relevante para el informe pericial (datos del usuario, lugar donde se realiza la pericia, personas presentes en el acto, entre otras).

4) Identificar si el correo en análisis corresponde a un correo emitido por el usuario o a un correo recibido por el usuario. En el primer caso, el encabezado solo dará certeza de que el correo salió de la cuenta de correo en análisis, por lo tanto no se puede verificar el envío del mismo, indicar esta situación en el Informe de Pericia y finalizar la pericia. Si se trata de un correo recibido, el encabezado permite trabajar con la trazabilidad del correo hasta su origen, y en ese caso se debe continuar con el paso 5.
5) Acceder a la cuenta en la cual se encuentra residente el correo, sea éste un cliente remoto o local. Aunque al cliente remoto se puede acceder desde cualquier computadora, siempre es preferible hacerlo desde el propio dispositivo del aportante del correo. En cambio, si se utilizó un cliente local solo se puede acceder al correo electrónico que se pretende analizar, desde la computadora o dispositivo del aportante. Se debe destacar que, para no acceder a correos no vinculados a lo indicado por el Juez en los puntos de pericia e incurrir en el delito de acceso no autorizado a datos privados, el perito debe acceder a la cuenta teniendo cuidado de identificar y separar solo los correos pertinentes.
6) Una vez identificado el correo electrónico se debe extraer la cabecera completa, accediendo a la misma a través de los metadatos correspondientes.
7) Se debe realizar una copia forense de la evidencia con su correspondiente valor hash. Accediendo a esta copia forense, se debe utilizar la herramienta forense adecuada para analizar la cabecera.
8) Se debe analizar la cabecera del correo. Cabe mencionar que lo indicado a continuación en este paso, es lo que usualmente realiza el perito cuando analiza un único correo electrónico, pero cuando se trata de un conjunto de correos, es necesario utilizar una herramienta forense que le permita realizar el análisis con cierto grado de confianza, y evitar errores producto de una revisión manual.

   Supuesto que fuera una única cabecera a analizar, se deben realizar las siguientes verificaciones:

   8.1 Buscar la cuenta del emisor del correo en análisis, que puede estar en el campo *X-Originating-IP* o en el campo *From.*

8.2 Analizar el parámetro *Received*. Para cada parámetro encontrado (se debe leer el encabezado de abajo hacia arriba) realizar los siguientes pasos:

8.2.1 Identificar los datos de la dirección IP encontrada en el paso anterior. Puede ocurrir que dicha dirección se encuentra en formato IPV4 ó IPV6, o bien, figure como un hostname en cuyo caso debe identificarse la IP correspondiente si fuera posible.

8.2.2 Con la información del encabezado se verifica el origen del mensaje enviado, buscando con el número IP registrado el dominio de donde se originó el mensaje. Para ello se puede utilizar una interfaz de identificación de direcciones IP, que ayudan a identificar el servicio utilizado, ubicar la dirección geográfica de los servidores y los puntos de contacto, y geolocalizar (cuando es posible) la instalación donde se encuentra un equipo[33].

8.3 Realizar el paso 8.2 con las sucesivas direcciones IP halladas en el encabezado.

9) Con la información obtenida elaborar el Informe de Pericia, de acuerdo a las normas de estilo de presentación de documentos judiciales.

Todo el procedimiento debe ser efectuado en presencia de un escribano u oficial de justicia el cual validará las tareas realizadas.

Obsérvese que el procedimiento es sencillo cuando se debe peritar un único correo electrónico, pero no lo es cuando se trata de múltiples correos. Principalmente porque al tratarse de un solo correo, el análisis del encabezado puede realizarse manualmente y si el perito es ordenado en su tarea, es probable que no incurra en errores u omisiones producto de realizar una tarea rigurosa un número importante de veces.

Necesariamente el perito debe recurrir a alguna herramienta que automatice la tarea de analizar los correos cuando la cantidad lo amerita en pos de ganar tiempo y realizar una tarea eficiente y sin errores técnicos.

[33] Existen varias páginas web que ofrecen el servicio gratuito de identificación de una IP, entre ellas: http://network-tools.com/, http://whois.domaintools.com. Estas páginas web brindan información de la dirección IP, propietario del registro en ARIN (American Registry for Internet Numbers). LACNIC (Latin American and Caribbean Internet Addresses Registry y DNS(Sistema de Nombres de Dominios), reverso de la dirección IP del hostname (lookup), nombre asociado con la dirección IP, Nombre del contacto del proveedor y Nombre del Responsable de tramitar el dominio.

2.6 Herramientas para el análisis forense de Correos Electrónicos

La Forensia Digital ha entrado en una crisis producto del impacto de dos elementos que marcan la época actual de la tecnología informática: la masividad de los datos y la multiplicidad de plataformas tecnológicas. (Garfinkel, 2010) presenta varios desafíos, involucrando no solo los modelos de "visibilidad y búsqueda" que proponen las herramientas forenses de uso actual sino también la falta de integración de las estrategias (como la ingeniería reversa) con dichas herramientas para reducir tiempos y costos. Cita este autor como próximos desafíos a resolver:

- **Diseño de las herramientas orientadas a la evidencia**: usualmente las herramientas actuales se orientan a la búsqueda de elementos digitales (evidencia) pero no a la presentación, resumen o análisis de correlaciones entre los datos encontrados.
- **Modelo de visibilidad, filtro e informe**: las herramientas utilizan interfaces de comunicación con el experto forense que habitualmente no permiten establecer vínculos o relaciones de prioridad entre los datos encontrados. Incluso algunas herramientas se basan en algoritmos computacionales costosos en tiempo y pueden faltarle características de usabilidad para el usuario final. La automatización o generación de scripts para búsqueda y filtro no siempre resultan. Y se complica aún más ante el avance continuo de las tecnologías (procesamiento paralelo, virtualización, deep web, etc.).
- **Problemas estructurales en las herramientas forenses**: en muchos casos se recurre a software desarrollado para el contexto de negocios o para sistemas transaccionales y no responden exactamente a las necesidades puntuales de la búsqueda de evidencia digital. Ocurre lo mismo con tecnologías integradas, tales como las aplicaciones monolíticas.
- **Abstracción y modularización**: debido al volumen de datos que se procesan en la búsqueda de la evidencia digital, se requiere fijar estándares para la identificación, transmisión e intercambio de los datos; igualmente es importante generar arquitecturas de procesamiento que superen los conflictos del software abierto y propietario.

- **Enfoque en la identidad del individuo**: tomando como atributos todos aquellos datos que puedan generar una "imagen" de la persona (datos de identificación, datos bancarios, correos, vínculos de las redes sociales, etc.).

Existen muchas y diversas herramientas disponibles para analizar un correo electrónico que fueron analizadas en (Rivetti & Parra de Gallo, 2017). A los allí citados se agrega el trabajo de (Devendran, Shahriar, & Clincy, 2015) acerca de un estudio comparativo de varios software open source para el análisis de correos electrónicos.

La elección de la técnica y herramientas más adecuadas se deduce de la estrategia de investigación que siga el perito, la cual dependerá de ciertos factores: dispositivo a analizar (PC, celular, servidor, etc.); cliente de correo (residente en el dispositivo o web mail); cantidad de correos (se debe analizar toda la cuenta o solo un correo determinado) y facilidad de acceso a la prueba (acceso al email enviado y al recibido, solo a uno de ellos, al servidor de correo, etc.),

Para realizar el análisis del encabezado de un correo electrónico es necesario utilizar herramientas forenses las cuales nos proporcionan información que extrae del encabezado y nos brinda la posibilidad de generar distintos tipos de reportes que pueden integrarse el informe parcial. Sin agotarse, el siguiente listado enumera las herramientas forenses más usuales para el análisis de correos electrónicos:

Aid4Mail[34] soporta más de 40 formatos de correo electrónico y programas de cliente de correo, así como muchos servicios populares de correo web y cuentas remotas a través de IMAP (Internet Message Access Protocol). Como resultado, los desarrolladores de este producto indican que es posible procesar prácticamente cualquier tipo de buzón que llegue a su destino. Las carpetas y archivos de correo local se pueden procesar fácilmente cuando se desconectan de su cliente de correo electrónico, incluidos los almacenados en discos duros externos y medios como DVD y dispositivos USB. *Aid4Mail* puede leer archivos *mbox*[35] de sistemas Mac y Linux sin conversión previa.

EmailTrackerPro[36] no sólo ofrece la capacidad de rastrear un correo electrónico usando el encabezado de correo electrónico, sino que también viene con un filtro de

[34] http://www.aid4mail.com

[35] mbox es un conjunto de formatos de archivo de texto plano utilizado para almacenar mensajes de correo electrónico en un único archivo como texto ASCII de 7 bits y los archivos adjuntos se almacenan en formato codificado.

[36] http://www.emailtrackerpro.com

spam (edición avanzada), que escanea cada correo electrónico a medida que llega y advierte al usuario si se sospecha de spam La característica más valiosa de *EmailTrackerPro* es la capacidad de rastrear más de una dirección IP o nombre de dominio a la vez. Se puede trazar tantas direcciones IP y nombres de dominio como sea necesario y se envían los resultados a una nueva pestaña o un archivo Excel / HTML.

MailNavigator[37] fue creado a partir de dos herramientas para la lectura de email y canales de noticias: FILTER, que es un poderoso sistema para la búsqueda de correos en ficheros de los distintos programas de e-mail; y NAVIGATOR, que es un lector de correos y noticias con funciones avanzadas.

OSForensics[38] permite extraer pruebas forenses de computadoras rápidamente con búsquedas e indexación de archivos de alto rendimiento. Puede identificar archivos sospechosos y actividad con coincidencia hash, comparaciones de firmas de unidad, correos electrónicos, memoria y datos binarios. También permite administrar una investigación digital y crear informes de datos forenses recopilados.

La empresa creadora de *E-mail Examiner*[39] ha sido una de las pioneras en soluciones para dispositivos móviles, teléfonos inteligentes y correo electrónico, y su enfoque de trabajo en la movilidad le permitió avanzar en muchas otras áreas de la innovación incluyendo la investigación y el desarrollo en el Internet de Cosas (IoT) con el *Forensics of Everything TM* (FoE). *E-mail Examiner* permite analizar los encabezados, los cuerpos y los archivos adjuntos de los correos electrónicos. Analiza el mensaje de principio a fin, incluyendo la clasificación y análisis detallado de archivos adjuntos. Soporta los principales tipos de correo electrónico que se almacenan en equipos locales para análisis, generación de informes y exportación/conversión de datos.

Systools Software, creador de *MailXaminer*[40], se dedica a proporcionar herramientas de alta tecnología con interfaz de usuario amigable. Ha contribuido con la recuperación de datos, soluciones de copia de seguridad, así como herramientas forenses de investigación y análisis de correo electrónico. El primer lanzamiento importante realizado en el campo de las aplicaciones de *eDiscovery* fue

[37] http://www.mailnavigator.com
[38] http://www.osforensics.com
[39] https://www.paraben.com
[40] https://www.mailxaminer.com/

MailXaminer, es un conjunto completo de herramientas para la documentación, análisis, examen y notificación de evidencias de correo electrónico.

Incluso existen frameworks integrados para el análisis forense de correos electrónicos, tales como el *Integrated E-mail Análisis Forense Framework (IEFAF)*, propuesto por (Hadjidj et al., 2009), que consta de 5 módulos: Navegador Interbase de datos, Explorador de estadísticas, Explorador de minería de datos, submódulo Weka y Explorador de E-mail. Consta además de una interface gráfica e intuitiva con 5 visores: Visor de Edición de Detalles del e-mail, Visor de mapas para la ubicación geográfica de las IP encontradas en el e-mail, Visor de Estadísticas acerca de la frecuencia de uso de palabras, eventos u objetos repetitivos; Visor de red social que muestra la vinculación entre todos los IP y nombres de clientes de mail; y Visor de minería de datos que opera con Weka.

EnCase Forensic[41] es una poderosa plataforma de investigación que recolecta datos digitales, realiza análisis, informa sobre descubrimientos y los preserva en un formato válido a efectos legales, en el caso de correos electrónicos cuenta con una amplia compatibilidad con los distintos formatos de archivos de correos y permite obtener un archivo imagen del mismo con el objeto de preservar la prueba original libre de manipulación. Aquí también es distintiva la funcionalidad de la herramienta para el análisis forense de correos electrónicos, y los informes de resultados que muestran los datos desde varias ópticas, pero dejando a consideración del perito la selección de los resultados que le permitan responder a los puntos de pericia.

Para el caso particular de los dispositivos móviles que se encuentran en el mercado, se pueden utilizar las siguientes herramientas de análisis forense: *TULP2G*[42], y *MOBILedit FORENSIC*[43]. Incluso existen herramientas diseñadas específicamente para el análisis forense de determinada tecnología celular, tal como el *Elcomsoft Phone Breaker*[44] diseñada específicamente para realizar el análisis forense mejorado en los dispositivos con iOS[45].

En el Capítulo 4, que describe ObE Forensic, la herramienta para el análisis forense de correos electrónicos basada en OntoFoCE, se discute acerca de las ventajas que presenta esta herramienta respecto de las aquí nombradas.

---

[41] https://www.guidancesoftware.com
[42] http://tulp2g.sourceforge.net/
[43] https://www.mobiledit.com/mobiledit-forensic
[44] https://www.elcomsoft.es/eppb.html
[45] Sistema operativo móvil desarrollado por Apple para los teléfonos inteligentes de esa marca

2.7 Estado del Arte

Se realizó un estudio sistemático del estado del arte en las áreas de interés para el desarrollo de esta tesis: ontologías y forensia digital, y de la aplicación del concepto de trazabilidad en ambas temáticas.

Respecto del tipo de estudio, se efectuó una investigación exploratoria para identificar el estado del arte sobre la aplicación de las tecnologías semánticas a la Forensia Digital, particularmente en la forensia de correos electrónicos, realizando un estudio crítico y ajustando el alcance del mismo a los objetivos propuestos. Éstos últimos se mencionan a continuación:

- Identificar y estudiar los aportes investigativos más actualizados sobre Ontologías y Forensia Digital.
- Establecer las áreas de vacancia sobre la aplicación de ontologías a la Forensia Digital
- Relacionar los trabajos desde atributos de cercanía (o distancia) con la aplicación de las ontologías al análisis forense de correos electrónicos.

A fin de realizar una revisión bibliográfica ordenada y ajustada a norma, se analizaron diversas metodologías de revisiones de trabajos científicos, identificando los criterios más útiles para la temática en estudio. Este análisis metodológico se detalla en el ANEXO I: INVESTIGACIÓN BIBLIOGRÁFICA.

A partir de lo dicho sobre el tipo de estudio, los objetivos que se persiguen y las metodologías analizadas, se optó por tomar como guía la metodología propuesta por (Medina Lopez, Marin Garcia, & Alfalla Luque, 2010) y lo dicho por (Portugal, Alencar, & Cowan, 2018) para realizar una revisión del estado del arte de la aplicación de las ontologías en problemáticas de forensia digital de correos electrónicos. Así, se propuso un método de revisión sistemática para el estudio que se introduce en el presente trabajo basado en las siguientes fases: A) Definición del Marco de Estudio y Alcance de la Revisión y B) Procesos de Búsqueda y Selección. En el ANEXO I se describe con detalle cada una de estas etapas y los resultados obtenidos de la revisión bibliográfica realizada.

En el caso que nos ocupa, el marco teórico de estudio surgió de la conjunción de dos temáticas principales: las tecnologías semánticas y la forensia digital. Ambas áreas, de amplísimo desarrollo por sí solas, se fusionan en trabajos de investigación particulares que cuentan cada uno de ellos con sus propios objetivos, en los que –en

la generalidad- se observa la aplicación de ontologías en la resolución de problemas de la forensia digital.

El estudio del estado del arte sobre Forensia Digital permitió identificar las temáticas más investigadas, tales como Tráfico de Redes particularmente en lo referido a identificación de direcciones IP, Minería de Datos con énfasis en la minería de textos y técnicas de extracción de datos y en la Seguridad Informática que se enfoca principalmente en la atención de ataques cibernéticos. Y fue posible identificar las áreas de vacancia que pueden ser de interés para los investigadores del área: Big Data, tecnologías semánticas y procesamiento de lenguaje natural. Respecto de la aplicación de las tecnologías semánticas, los trabajos relacionados al desarrollo de ontologías son de especial interés.

La herramienta CyBox propuesta en el estudio de (Casey, Back, & Barnum, 2015) basada en la ontología DFAX permite representar e intercambiar información forense digital, incorporando aspectos procesales de la forensia, como ser la cadena de custodia, manejo de casos y procesamiento forense. (Karie & Venter, 2014) proponen una ontología para categorizar las disciplinas forenses digitales, así como algunas metodologías que puedan ofrecer orientación en diferentes áreas de la forensia digital. Las investigaciones realizadas por (Amato, Cozzolino, & Mazzocca, 2017) y (Amato, Cozzolino, Mazzeo, & Moscato, 2018) describen un marco ontológico que recupera y muestra modelos de evidencias obtenidos a través de diferentes herramientas forenses, presentando un sistema capaz de añadir afirmación semántica a los datos generados por las herramientas de análisis forense durante los procesos de extracción.

Además de estas publicaciones que describen el uso de ontologías para resolver diferentes problemáticas de la forensia digital, se encontró el estudio de (Mehta, 2017), referido a la aplicación de tecnologías semánticas para el direccionamiento de correos electrónicos entre un grupo de usuarios especificados semánticamente, pero se debe destacar que no contempla la cabecera del correo electrónico en la ontología desarrollada.

Particularmente, en lo referente a las investigaciones sobre forensia de correos electrónicos, los trabajos de interés son varios. Por una parte, la herramienta RDAP presentada por (Nikkel, 2017), permite obtener información acerca de una dirección IP identificada, y a partir del trabajo de investigación en la que se presentó la herramienta, se consideró incluirla en ObE Forensic, como herramienta de búsqueda

de datos de geolocalización de la dirección IP de los servidores que intervienen en el proceso de transmisión del correo.

Respecto de las herramientas para forensia de correos electrónicos, se estudió con atención los trabajos de (Z. Chen et al., 2017), (L. Chen & Mao, 2017), (Koven, Bertini, Dubois, & Memon, 2016), (Jayan & Dija, 2015), (Msongaleli, 2018), (Romaios, Nikolaos, George, & Andreas, 2016), (Umar, Riadi, & Muthohirin, 2018) y (Devendran et al., 2015) considerando la amplitud y profundidad de la información que brindan estas herramientas y el grado de colaboración con el perito al momento de realizar el análisis pericial. Debe considerarse además, que todas estas herramientas se basan en el análisis de la cabecera del correo electrónico, y dan por sentado que al considerar las direcciones IP que figuran en ella, se analiza el proceso de transmisión, pero no lo enfocan de manera directa y concreta, con excepción de lo investigado por (Msongaleli, 2018). Se observó que si bien hay investigaciones que consideran el proceso de transmisión del correo electrónico al abordar el encabezado del mismo, en ninguno de ellos se aplica el concepto de trazabilidad como elemento vinculante de los distintos equipos o servidores actuantes.

Como síntesis de la búsqueda de investigaciones que tratan sobre la aplicación de tecnologías semánticas a la Forensia Digital, se puede decir que es escaso el uso de las tecnologías semánticas en la formulación de herramientas forenses.

Por otra parte, son varios los trabajos que abordan el estudio de la cabecera del correo electrónico para el análisis forense de los mismos, pero solo en uno se abordó detalladamente el proceso de transmisión.

La literatura relacionada a componentes de la Forensia Digital que no contemplen el uso de ontologías es variada y abundante. Se destaca el trabajo de (Chhabra & Bajwa, 2012) que revisa integralmente el tema, desde el funcionamiento y la arquitectura del sistema de correo electrónico y los protocolos de seguridad generales, hasta las técnicas de análisis forense y las herramientas más utilizadas.

Hay varios trabajos comparativos de *herramientas forenses para el análisis de correos electrónicos*, la investigación de (Devendran et al., 2015) presenta un estudio sobre las herramientas de análisis forense de email habitualmente más utilizadas, discutiendo acerca del enfoque no colaborativo que presentan y las dificultades de integración entre ellas. (L. Chen & Mao, 2017) estudia las particularidades de realizar forensia de emails en las oficinas móviles actuales, es decir, sobre notebook, celulares, tablets, etc., considerando el análisis forense del

contenido de la memoria volátil de un teléfono celular, a través de herramientas propuestas para tal fin.

Se han detectado trabajos que desarrollan casos puntuales de análisis forense descripto a través de ontologías. (Youn, 2014) definió SPONGY una ontología para modelar el filtrado de correo SPAM en una cuenta de correo electrónico. Otro ejemplo de casos de uso de interés para la presente investigación, es el trabajo de (Carvalho, Goldsmith, & Creese, 2016) sobre un caso de fraude bancario on line, en donde se propone una ontología OBMO (Online Banking Malware Ontology) para modelar las organizaciones criminales intervinientes e identificar a los desarrolladores de malware.

Respecto de la *trazabilidad y su aplicación en la ingeniería ontológica* la investigación realizada permitió encontrar trabajos de interés para esta tesis. (Souali, Rahmaoui, & Ouzzif, 2017) aborda el estado del arte de la trazabilidad y su aplicación. (Vale et al., 2017) realizan un estudio profundo de la rastreabilidad en las líneas de productos de software, las familias de productos y los procesos de desarrollo subyacentes. (Selamat, Shahrin, Hafeizah, Yusof, & Abdollah, 2013) aborda la investigación de aspectos de trazabilidad en el proceso de investigación forense digital, particularmente cuando se trata de volumen complejo y enorme de evidencia y es necesario conectar relaciones significativas entre ellos. Hay dos trabajos de un mismo autor (Morgan, 2017a) y (Morgan, 2017b), en el que se presenta un modelo conceptual (FoRTE) que describe la naturaleza holística de la evidencia en el "esfuerzo" de la reconstrucción forense, indicando la importancia del rastreo de los componentes. Es decir, se coincide en el concepto de logística inversa (del resultado al origen) para darle validez a la prueba forense. Por su parte (Ali, 2016) avanza sobre la rastreabilidad basada en procedencias, para capturar y consultar eventos que ocurrieron en el pasado y así comprender cómo y por qué se llevaron a cabo, y aplica este concepto a los servicios en la red.

2.7.1 Impacto de los trabajos revisados en la presente investigación

A partir de los textos citados es posible sintetizar el estado del arte en los tres temas analizados en forma integrada: ontologías, forensia digital y trazabilidad. A continuación, se resumen los resultados más destacados.

Existen desarrollos ontológicos que se utilizaron para describir las entidades del correo electrónico y sus relaciones. Estos estudios definen el marco conceptual y el alcance de la investigación que se está realizando. (Amato et al., 2017) señalan que *"... las tecnologías de la Web Semántica pueden mejorar el proceso de las investigaciones digitales debido a sus capacidades relacionadas con la automatización y la integración de datos...*".

El estudio de CyBox, la herramienta propuesta por (Casey et al., 2015), permitió incorporar en OntoFoCE aquellos aspectos necesarios para mantener la cadena de custodia, tales como la aplicación de funciones Hash para la encriptación del cuerpo del mensaje que va circulando durante el proceso de envío del correo.

Las metodologías que describe (Karie & Venter, 2014) se estudiaron para orientar la selección de la metodología utilizada para el análisis forense digital, y la definición del aporte de la herramienta propuesta, ObE Forensic, en el procedimiento pericial.

De los trabajos de (Amato et al., 2017) y (Amato et al., 2018) se valoró el modo en que las tecnologías semánticas se aplican a la representación de la evidencia digital, y permitió confirmar que el conjunto de conceptos considerados para el análisis forense de correos electrónicos –y que luego se representan en OntoFoCE- es el adecuado.

Aquellas investigaciones sobre ontologías que utilizaron Methontology como metodología de desarrollo, (Ellison, Venter, & Adeyemi, 2016) y (Alzaabi, Martin, Taha, & Jones, 2017), se revisaron a fin de considerar sugerencias que ayudaron durante las sucesivas fases iterativas realizadas para construir OntoFoCE utilizando dicha metodología.

Además, los trabajos que describen la utilización de herramientas y lenguajes para la construcción de ontologías, que también son utilizados para la construcción de OntoFoCE, sirvieron para solucionar problemas de sintaxis, uso y buenas prácticas de dichas herramientas. Así, se consideraron los trabajos de (Mavroeidis, 2018), (Amato et al., 2018), (Carvalho et al., 2016), (Wimmer, Chen, Narock, & Chen, 2018) y (Kalemi & Yildirim-Yayilgan, 2016).

Los trabajos que abordan los métodos de la forensia digital enmarcan la investigación forense a partir de los criterios ya tomados por estos investigadores y permiten definir el criterio de no repudio del correo electrónico como prueba digital. Por ejemplo: heterogeneidad de los datos, procedimiento pericial, cadena de

custodia, entre otros. En particular, de la publicación de (Chabot, Bertaux, Nicolle, & Kechadi, 2015) interesa destacar el análisis que realizan los autores acerca de la necesidad de garantizar la trazabilidad de la información en el proceso forense, para cumplir con los criterios de credibilidad, integridad y reproducibilidad de la evidencia digital. Cumplir con estos criterios aumenta significativamente el peso de la prueba digital, y recurrir a las tecnologías semánticas para modelar la trazabilidad de la prueba, permite al tribunal judicial entender plenamente como se llega a una conclusión válida partiendo de la evidencia que es un elemento técnico incompresible en sí mismo para ellos.

Al analizar los trabajos sobre Trazabilidad y Ontologías, se destaca la definición propuesta por (Souali et al., 2017) que define la trazabilidad como "*la capacidad para mantener un historial detallado de todas las actividades y cambios que un objeto en particular puede experimentar a lo largo de todo su ciclo de vida, teniendo en cuenta las diferentes relaciones que pueden aparecer. Este objeto en particular puede ser un material, un producto, un modelo o incluso una clase en una plataforma de desarrollo de software*". Este autor elabora el estudio considerando dos contextos: la industria alimentaria y la tecnología de la información. A partir de ellos, describe las principales técnicas utilizadas en trazabilidad y propone modelos de definición para los dos sectores. Si bien el trabajo no incluye una relación directa con las ontologías, sienta las bases para estructurar el marco conceptual de la trazabilidad y agregarlo a la investigación en curso.

Si consideramos el objeto de estudio en sí –el correo electrónico- no se encontraron trabajos que resuelvan el análisis forense de este componente utilizando la ingeniería ontológica. Puede decirse que cada investigación analizada contribuye con un aporte de interés para OntoFoCE, pero no se encontró ninguna que reúna todos los requerimientos como un todo.

Así, se puede observar trabajos que abordan la forensia de correos electrónicos desde enfoques complementarios al propuesto en la presente investigación. En el estudio denominado "A Forensic Analysis Solution of the Email Network Based on Email Contents" de (Xie, Liu, & Chen, 2016), se aborda el análisis forense conformando una red que permite marcar la correlación de correos en una cuenta; por su parte el estudio de (Y. Zhang, Liu, & Chen, 2015) trata el análisis forense de email anónimos mediante patrones estructurales de escritura; el trabajo "Email Forensic Analysis Based on k-means clustering" de (Nampoothiri, 2015) emplea

minería de datos para recopilar y analizar correos electrónicos de usuarios sospechosos y establecer una posible asociación criminal; la publicación de (L. Chen & Mao, 2017), bajo el nombre de "Forensic analysis of Email on Android volatile Memory" propone una técnica para realizar la forensia de emails desde el contenido de la memoria volátil de un teléfono celular. Otro trabajo de interés para la presente tesis es "Identification and Analysis of Email and Contacts Artefacts on iOS and OS X" de (Ovens & Morison, 2016) en el que se aborda la complejidad de establecer el origen de correos electrónicos en un sistema de múltiples dispositivos que comparten una sola cuenta, y también el de (Stadlinger & Dewald, 2017) que propone un método estadístico para la visualización rápida de patrones de comunicación entre los correos de una cuenta. Particularmente se estudió la investigación realizada por (Chhabra & Bajwa, 2012) que abordan la forensia de correos electrónicos pero solo desde las herramientas que se pueden utilizar, así como la de (Armknecht & Dewald, 2015) sobre "Privacy-preserving email forensics" que ayudan al perito en el acceso a datos no autorizados, mediante la encriptación previa de porciones de la cuenta, permitiendo el acceso solo a determinados correos. El trabajo denominado "Email Visualization Correlation Analysis Forensics Research" de (Z. Chen et al., 2017) en el que se describe una herramienta visual propia para el cliente de correo Foxmail, permite extraer, asociar y graficar información referida a contactos, datos del cuerpo del correo y archivos adjuntos, y establecer correlaciones entre el remitente y el receptor.

Se puede considerar los siguientes trabajos como los más cercanos a la presente investigación: la publicación denominada "Electronic Mail Forensic Algorithm for Crime Investigation and Dispute Settlement" de (Msongaleli, 2018), que aborda el análisis forense a partir del análisis del encabezado, y de los registros de servidores y de equipos de emisión/recepción para la identificación de correos falsos únicamente, sin considerar la ingeniería ontológica como estructura base de la propuesta ni tampoco responde de manera directa a los requerimientos solicitados por el Juez en los puntos periciales. También se analizó el estudio de (Kota, 2012) denominado "An Ontological Approach for Digital Evidence Search" en el que propone una ontología de correos electrónicos con instanciación dinámica a partir de los conceptos, propiedades y relaciones entre los elementos que resulten relevantes para la consulta del forense, sin abordar la verificación de la veracidad del proceso de transmisión ni la respuesta a los puntos de pericia.

Se debe destacar que en ninguno de los trabajos precitados que tratan sobre el correo electrónico y sus componentes internos, se recurre a la ingeniería ontológica como marco contextual, ni se aborda el proceso de transmisión para fijar criterios que permitan establecer la veracidad del mismo.

La revisión bibliográfica realizada muestra que la Forensia Digital es un tema de interés en la comunidad científica, así como la necesidad de formalizar y aplicar metodologías, técnicas y herramientas propias para abordar el procesamiento de la evidencia digital. Por otra parte, las ontologías se han transformado en una opción válida para representar modelos pluridisciplinares, como es el caso de estudio de esta tesis. Y obviamente, el análisis de correos electrónicos encarado desde diferentes enfoques ha permitido identificar el estado del arte del contexto científico en el que se propone esta tesis.

2.8 Conclusiones del Capítulo

En este capítulo se abordó el estudio de dos áreas: las tecnologías semánticas, más específicamente el uso de ontologías, y la Forensia Digital, con foco en el análisis forense de correos electrónicos a partir de la cabecera del mismo. Un punto en común de este estudio fue la identificación de la trazabilidad, en cuanto a cómo se representa mediante ontologías, y en cuanto a su aplicación a diferentes procesos.

Como resumen del marco teórico definido y el análisis realizado sobre las investigaciones recabadas hasta la fecha, se pueden indicar los aspectos destacados encontrados:

- En la comunidad científica existe una marcada preocupación por desarrollar una respuesta adecuada al avance del cibercrimen o ciberdelito.
- Las tecnologías semánticas están siendo utilizadas por los investigadores del tema para formular modelos que soporten científicamente la actividad pericial.
- Existen numerosos estudios sobre diversos aspectos del análisis forense, entre los que se destacan: estudios comparativos de herramientas, formulación de técnicas o métodos forenses, estudio de casos de uso, herramientas para el análisis forense de dispositivos varios, entre otros.
- No se encontraron estudios que aborden directamente la verificación o validación del proceso de transmisión del correo electrónico, siendo que en la generalidad

de las veces, ésta es la actividad técnica central del perito, y sobre este proceso se sustenta la no repudiabilidad de la evidencia digital.

- Las herramientas de análisis forense se limitan a generar los resultados en términos técnicos, dejando a cargo del perito la interpretación de los mismos a la luz de los puntos periciales. Respecto de esto, ObE Forensics es una herramienta que concluye el análisis forense con la emisión de un Informe Técnico de Pericia, cuyo tenor y contenido es suficiente para presentar directamente al Juez las conclusiones arribadas en la pericia.
- En algunos casos, como con la herramienta *MailXaminer* se ha podido comprobar que no identifican la totalidad de direcciones IP y/o nombres de dominio que figuran en la cabecera.

A partir de los resultados de la investigación bibliográfica realizada, si se consideran las herramientas para el análisis forense de correos electrónicos, es importante el aporte que propone esta tesis, en cuanto al modelo ontológico que contempla el proceso de transmisión y su trazabilidad, y plantea resultados en términos de los puntos periciales que se deben responder a partir de las preguntas de competencia de la ontología, ajustando el análisis forense a la condición de no repudiabilidad de la prueba digital.

Además de OntoFoCE, el aporte de la tesis incluye una aplicación web, ObE Forensics basada en dicha ontología, que actúa como herramienta para el análisis forense de correos electrónicos. Considerando el proceso PURI de realización de pericias señalado en la Figura 2-6, ObE Forensic es una aplicación que acompaña al perito en la ***Fase de Extracción y Análisis*** de la evidencia digital, permitiendo representar la trazabilidad del proceso de transmisión del correo, y señalando los datos de identificación de las cuentas y equipos utilizados en la transmisión. Por otra parte, los informes técnicos con las respuestas a las preguntas de competencia que se emiten desde OntoFoCE pueden utilizarse para respaldar la ***Fase de Presentación*** del análisis forense realizado.

Hasta aquí, la descripción del marco teórico de la presente investigación. Los siguientes capítulos abordan la construcción e implementación de OntoFoCE, la ontología para el análisis forense de los correos electrónicos, así como ObE Forensic, la aplicación web que la utiliza. En el capítulo 3 se describe ampliamente el proceso de desarrollo de OntoFoCE, así como la metodología y las herramientas

utilizadas para la construcción de la ontología; y en el capítulo 4 se detallan las características técnicas y funcionales de OBE Forensic, la aplicación informática propuesta para el análisis forense de correos electrónicos basado en OntoFoCE.

CAPÍTULO 3. OntoFoCE: UNA ONTOLOGÍA PARA EL ANÁLISIS FORENSE DE CORREOS ELECTRÓNICOS

3.1 Introducción

La ontología para el Análisis Forense de Correos Electrónicos (OntoFoCE) es uno de los aportes del presente trabajo de tesis. La ontología propuesta tiene por objetivo permitir la representación del correo electrónico y su proceso de transmisión a fin de *comprobar la autenticidad del correo electrónico como prueba digital y en consecuencia la condición de no repudiabilidad de la prueba*, recurriendo a la ingeniería ontológica para garantizar la obtención de resultados científica y técnicamente aceptados como válidos según las normas que la justicia exige a los resultados periciales.

En este capítulo se presentan los conceptos principales de OntoFoCE, así como su proceso de desarrollo. Se inicia la descripción con la sección 3.2 en la que se explica cómo se aplicó la metodología Methontology para la construcción de OntoFoCE, detallando los aspectos generales de la metodología. Luego, se introducen los resultados más importantes de dos de estas etapas: la especificación de requerimientos y la conceptualización.

En la sección 3.3 se describe los resultados obtenidos en la fase de especificación de requerimientos. Estos resultados especifican el alcance de la ontología propuesta.

A continuación, en la sección 3.4, se especifica en detalle el Modelo Conceptual de OntoFoCE organizando la descripción según los distintos conjuntos de conceptos que se utilizaron para desarrollar la ontología: los conceptos asociados al Correo Electrónico en la sección 3.4.1, luego los conceptos asociados al proceso de transmisión en la sección 3.4.2 expresando los tres momentos de ese proceso: emisión, recepción y transmisión interna, y por último los conceptos complementarios requeridos para documentar la pericia y que están presentes en el modelo ontológico (sección 3.4.3).

El capítulo continúa luego con la sección 3.5 en la se describen ejemplos de aplicación, en el cual la ontología propuesta es utilizada para representar el proceso de transmisión de correos electrónicos en tres escenarios diferentes.

Por último, la sección 3.6 presenta las conclusiones de lo trabajado en este capítulo.

3.2 Aspectos Generales de la Aplicación de Methontology para el Desarrollo de OntoFoCE

Como se mencionó en el Capítulo 2, la metodología Methontology propone guías de actividades para la especificación, conceptualización, formalización, implementación y mantenimiento de la ontología a construir, bajo un esquema de procesos iterativos que ayudan en el ajuste del modelo a construir. A continuación, se describen los aspectos generales de la aplicación de esta metodología para la construcción de OntoFoCE.

En la fase de *especificación* se identificó el dominio de OntoFoCE a partir de un conjunto de definiciones sobre el análisis forense de correos electrónicos que permitieron delimitar el alcance del mismo, como por ejemplo: objetivos propuestos para OntoFoCE, utilización de esta ontología en los sucesivos pasos del procedimiento del análisis forense, y rol de los peritos como usuarios finales de la misma. La definición de los requerimientos se formula en términos de *preguntas de competencia,* que son un componente ampliamente utilizado en el área de ingeniería ontológica para esta actividad.

Durante la fase de *conceptualización* se estudió toda la información recabada sobre el análisis forense de correos electrónicos y se organizó en estructuras de representación intermedia, tipo diagramas UML y tablas, obteniendo como resultado la definición de los conceptos básicos de la ontología: clases, relaciones, propiedades e instancias de ejemplo. Asimismo, a fin de limitar la interpretación de los diagramas UML, se definieron restricciones y reglas utilizando la lógica descriptiva (DL) y el lenguaje SWRL (Semantic Web Rule Language). En particular, para las expresiones en DL se utilizó la sintaxis OWL Manchester[46], la cual se deriva de la sintaxis OWL original y es más fácil de leer (Horridge et al., 2006).

En la fase de *formalización e implementación* se utilizó la herramienta Protégé, en su versión 5.5.0 en el que se utilizó el lenguaje OWL para la definición de las clases y propiedades de OntoFoCE, SWRL para la formulación de las reglas de

[46] https://www.w3.org/2007/OWL/draft/ED-owl2-manchester-syntax-20081128/

inferencia y SPARQL para la definición de consultas que formalizan las preguntas de competencia.

Por último, se realizó la *validación* de la ontología, mediante una metodología integrada que contempla cuatro criterios que permiten enfocar la validación desde varios puntos: a) validación del uso correcto del lenguaje, b) exactitud de la estructura taxonómica, c) validez del vocabulario y d) adecuación a requerimientos. Este proceso de validación se explica en detalle en el Capítulo 5.

Por otra parte, los procesos iterativos seguidos incluyeron la realización de las actividades centrales de la construcción de la ontología en cada iteración, con distinto peso y grado de avance para cada actividad, según el momento de la construcción. Esto se grafica en la Figura 3-1.

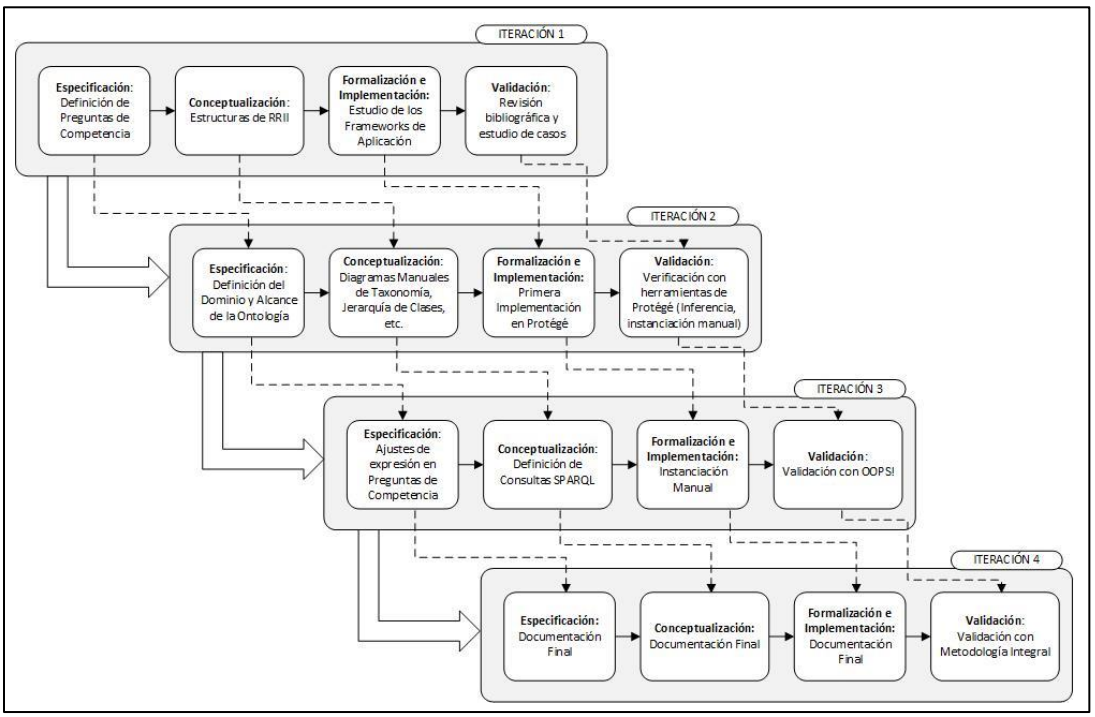


Figura 3-1: Procesos Iterativos de Construcción de OntoFoCE

Como muestra la Figura 3-1, la primera iteración se inició con la definición de las preguntas de competencia de la fase de especificación, luego se avanzó en la construcción de las estructuras de representación intermedias (señaladas como Estructuras RRII en la Figura 3-1), elaborando las tablas y diagramas UML necesarios para el grado de representación de la conceptualización que hasta ese momento se había logrado. Respecto de la fase de Formalización e Implementación, en esta primera iteración se realizó el estudio de los frameworks de trabajo más comunes para la construcción de una ontología y se seleccionó Protégé Versión 5.5.0. Esta primera iteración finalizó con una investigación sobre las diferentes

técnicas y metodologías de validación de ontologías, a modo de aproximación al tema.

En la segunda iteración ya se pudo avanzar con la definición más acotada del Dominio de la Ontología, y se ajustó el alcance de la misma, discutiendo acerca de los conceptos que debía (o no) incluirse en el modelo. Realizado ese ajuste del dominio, se procedió a elaborar en detalle la tabla de conceptos, el diccionario de conceptos, la taxonomía en su primera versión y otros diagramas intermedios referidos a la ontología que se estaba construyendo. En esta iteración se realizó la primera implementación del modelo lógico en Protégé, avanzando así en la fase de Formalización e Implementación. Al tener una primera versión implementada en Protégé se pudo realizar la primera validación de consistencia del modelo, mediante las herramientas y el razonador provisto por este framework.

Durante la tercera iteración, fueron escasos los ajustes realizados en la fase de conceptualización, mínimamente se ajustaron algunas palabras de las preguntas de competencia, expresándolas con más riqueza y certeza según los conceptos representados en OntoFoCE. En esta iteración el modelo implementado en Protégé ya estuvo suficientemente ajustado y maduro como para avanzar con las consultas SPARQL de las preguntas de competencia, esta actividad fue la que se agregó en la fase de Conceptualización. Llegado a este punto, la fase de Formalización e Implementación consistió en realizar las primeras instanciaciones con datos de prueba, vislumbrando ya los distintos escenarios que debían plantearse para el análisis forense de correos electrónicos. Durante la fase de Validación de esta iteración se recurrió a las herramientas automáticas (OOPS! OntOlogy Pitfall Scanner[47]) para validar el código OWL.

La última iteración, incluyó la documentación final de las fases de Conceptualización, Especificación, Formalización e Implementación, y se realizó la validación exhaustiva del modelo logrado.

Por supuesto que, además de mantener una relación transversal de todas las fases en cada iteración, se mantuvo también una línea de trabajo dentro de cada una de estas fases en la que se marcaba la complejidad creciente de una iteración a otra, manteniendo un feedback adecuado de cada fase en las restantes. Por ejemplo: en la fase de conceptualización, lo realizado en la iteración 1 se tomaba como insumo para

[47] http://oops.linkeddata.es/

trabajar esta misma fase en la iteración 2, y se revisaban los resultados de la iteración 1 de las otras fases para verificar el impacto que podría tener en ellas el ajuste propuesto para la conceptuación en la iteración 2.

3.3 Delimitación del Alcance de la Ontología

En esta sección se describe el alcance de OntoFoCE, el cual surge como resultado de la etapa de Especificación, allí se realizaron las siguientes acciones:

- Identificar el contexto o ámbito en el cual se utilizará la ontología
- Considerar los puntos de vista de los usuarios a la hora de modelizar los conceptos de la realidad, es decir, como conciben ellos cada concepto definiéndolos desde el propio ámbito de aplicación de la ontología.
- Definir las preguntas de competencia que representan los requerimientos funcionales de la ontología.

A fin de determinar debidamente el alcance de OntoFoCE, se define en más detalle el dominio de aplicación así como los requerimientos planteados para la ontología.

En el caso de OntoFoCE, el dominio está definido por el contexto en el que se realiza el análisis forense de correos electrónicos, particularmente interesa representar tres elementos: el objeto de estudio (correo electrónico), el procedimiento de análisis forense (a partir de los datos de la cabecera del correo electrónico y los datos complementarios de la pericia) y los puntos de pericia que actúan como fin último de realización de la pericia. Estos elementos se describen en el apartado 2.5 del Capítulo 2 de esta tesis, pero se extiende la definición de los requerimientos periciales en el presente capítulo, ya que a partir de ellos se definieron las preguntas de competencia de OntoFoCE.

En particular, es necesario explayarse acerca del objetivo propuesto para OntoFoCE respecto de sostener la autenticidad del correo para evitar el repudio de esta prueba digital.

La condición de no repudiabilidad se basa en la autenticidad del correo, que a su vez se logra cuando se verifica la premisa ya señalada en la sección 2.3 del Capítulo 2, acerca de que un correo electrónico es auténtico cuando se identifican los datos del remitente (cuenta de correo y dirección IP), la trazabilidad del mismo (diferentes

dispositivos que intervienen en la transmisión) y los datos del destinatario (cuenta de correo y dirección IP).

En esta tesis, se propone una herramienta que permita cumplir con la condición de no repudiabilidad de la evidencia digital, recurriendo a la ingeniería ontológica para garantizar la obtención de resultados según el rigor de cientificidad que la justicia exige a los peritos para realizar su tarea.

Los puntos de pericia son particulares para cada caso judicial, usualmente no se repiten en su enunciado ya que deben explicitar la tarea del perito con toda claridad y según los detalles de la causa. Aun así, es posible considerar que existen datos comunes que habitualmente se toman como base para responder a estos requerimientos de la pericia.

Para encontrar un conjunto de conceptos comunes se realizó un relevamiento sobre pericias de correos electrónicos entre los profesionales informáticos que actúan como peritos habilitados, recurriendo a la Red Universitaria de Informáticas Forense (REDUNIF) la cual integra la Universidad Católica de Salta.

Con la colaboración de aquellos peritos que respondieron a la encuesta, se obtuvieron 86 puntos de pericia, los cuales se listan en el ANEXO II: PUNTOS DE PERICIA. Luego se realizó un Focus Group con la participación de los investigadores del Grupo de Forensia Digital de la UCASAL, que además actúan como peritos, a fin de reducir los 86 puntos a un número adecuado considerando que muchos de los resultados inicialmente obtenidos de las encuestas expresaban lo mismo pero con diferentes palabras.

Como resultado de esta segunda revisión, se obtuvieron 46 puntos de pericia diferentes, y además se observó que en varios casos, solicitan datos que se pueden responder desde una misma consigna, por esa razón, los 46 requerimientos periciales identificados se responden a partir de 21 preguntas de competencia. La relación entre ambos conjuntos se muestra claramente en la Tabla II-3: Matriz de Relación Puntos de Pericia y Preguntas de Competencia, del ANEXO II: PUNTOS DE PERICIA.

Valga el siguiente ejemplo para explicar cómo se relacionan los puntos de pericia con las preguntas de competencia. El Juez puede solicitar uno de los requerimientos señalado a continuación**:**

> *Acceder al equipo y extraer toda información de la que pudieran surgir los correos electrónicos en los que la casilla "xxxxxxx@xxxx.com.ar" figure como remitente o destinatario.*

Esto se responde a partir de la información recabada en el equipo analizado, pero debe identificarse particularmente. Es decir, cuando se pide *"... en los que la casilla…figure como remitente o destinatario*" no se puede indicar simplemente cuales son los correos encontrados, sino que cada correo que responda al requerimiento pericial debe identificarse debidamente a través de la fecha, hora, dirección IP, cuenta del emisor, cuenta del receptor, equipo emisor y equipo receptor. Así, el ejemplo señalado en el párrafo anterior se puede responder a partir de las preguntas de competencia siguientes:

- PC01: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?
- PC02: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?
- PC04: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del emisor?
- PC05: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del Receptor?
- PC07: Dado un correo CE ¿Cuál fue el equipo desde el cual se emitió el correo?
- PC08: Dado un correo CE ¿Cuál fue el equipo en el que se recibió el correo?

A su vez, la pregunta de competencia "PC08: Dado un correo CE ¿Cuál fue el equipo en el que se recibió el correo?" se utiliza para responder los puntos de pericia 10 y 18, además del indicado en el ejemplo anterior, que dicen:

> *10) Determinar si los correos que se adjuntan como prueba se encuentran en los equipos informáticos de la demandada*.
>
> *18) Identifique cuales fueron los equipos de origen y de destino del mensaje. Además, deberá recabar datos de utilidad que permitan determinar el contenido del e-mail en cuestión. Los servidores o proveedores de Internet – que operan a modo de enlace entre el remitente y el destinatario del e-mail- cuentan con información inherente a la fecha en que se produjo el envío o reenvío del e-mail, duración de conexión, archivos adjuntos, número de identificación de los equipos de origen y destino de las comunicaciones – datos de tráfico.*

Por otra parte, se observó que hay puntos de pericia que no se pueden responder exactamente según lo solicita el Juez, o bien, requieren de alguna interpretación de lo solicitado considerando que se trata de una pericia informática, como ocurre con

los requerimientos periciales identificados como N° 1, N° 5, N° 43 y N°45 en la Tabla II-2: Puntos de Pericia Resultantes del ANEXO II: PUNTOS DE PERICIA.

Cabe mencionar que las causas de esta falta de respuesta no es una característica de la herramienta o del proceso pericial o del propio perito, sino más bien, surgen de la falta de acuerdo entre el Juez y el perito respecto de nombres, conceptos y procesos informáticos, es decir, muchas veces, el desconocimiento tecnológico de quien propone los puntos de pericia no le permite formular los requerimientos adecuadamente, y por su parte, el perito está obligado a responder tal y como se solicita en el punto de pericia. Es por ello que normalmente, existe una instancia de asesoramiento del perito al Juez a fin de comprender con más claridad lo que está solicitando. Para entender esta situación conviene detallar cada caso.

El punto de pericia N° 1 dice:

> *¿Quién es el titular de la casilla sss@yahoo.com.ar , ¿Cuándo se habilitó la misma?, señale si en el período de la causa fue el sistema de contacto del actor con la demandada o sus casillas de email; persona1@dominio.com.ar, usuario@yahoo.com.ar, persona2@dominio.com.ar, persona3@dominio. com.ar y persona4@dominio.com.ar e imprimir o transcribir los mails recepcionados y emitidos desde esa bandeja de servicios entre los siguientes sujetos y fecha.*

El ejemplo contiene varios interrogantes. El concerniente a "*¿Cuándo se habilitó la misma?* ", en referencia a la cuenta de correo en análisis, no se puede responder desde la cabecera de un correo electrónico porque dicha información no figura allí, pero sí es posible asesorar al Juez indicando que tal información se puede solicitar al proveedor de la cuenta de correo.

El punto N° 5 solicita:

> *Determinar el lugar físico de origen de los correos extraídos.*

En este caso se entiende que el Juez se refiere al equipo desde el cual se emitieron los correos electrónicos y no a un lugar geográfico físico en donde está ubicada la computadora. El perito hace esta aclaración en su informe, ya que si el equipo de emisión es un celular o una notebook podría ser de interés para la causa expresar puntualmente esta consideración que toma el perito en su análisis.

En el caso del punto de pericia N° 43 que indica:

> *"Teniendo a la vista los correos electrónicos determine la autenticidad de los mismos, puerto de entrada y salida y todo otro dato de interés",*

El perito debe realizar ante el Juez una consulta solicitando que le indique a que se refiere con puerto de entrada y salida, es altamente posible que haya querido indicar las cuentas de emisión y recepción, pero el perito no puede tomarse la atribución de interpretarlo así sin hacer la consulta previa.

El requerimiento del Juez indicado como N° 45:

> *"Verifique los mails que fueron remitidos entre PADRE (padre@dominio.com), persona (persona@dominio.com) e HIJO (hijo@dominio.com)".*

Cuando el perito elabora el informe, debe aclarar que no es posible constatar *quien fue la persona que emitió el correo,* solo se puede constatar los datos referidos al alias de usuario y tal vez firma del usuario de la cuenta, pero no se puede precisar nada acerca de la persona que utilizó la cuenta y realizó el envío de los correos.

El mismo caso ocurre con los puntos de pericia N°14, N°16, N°34, N°36 y N°40, en donde no es posible certificar fehacientemente la *pertenencia* de una cuenta de correo a una persona en particular, en cuanto a que este último haya generado la cuenta, la haya utilizado y sea el único que actúa como usuario de la cuenta.

El modelo de la encuesta, así como los 86 requerimientos periciales relevados y los 46 puntos resultantes de la fase de reducción de las preguntas por similitud de contenidos se detallan en el ANEXO II: PUNTOS DE PERICIA.

En la Tabla 3-1, se introduce el Documento de especificación de Requerimientos de la Ontología (OSRD), propuesta por (Suárez-figueroa et al., 2009), que resume el resultado de la actividad de especificación de requerimientos del proceso de desarrollo de OntoFoCE.

Tabla 3-1: Documento OSRD

Propósito	
	Modelar el proceso de análisis forense de correos electrónicos.
Objetivo General	
	Comprobar la autenticidad del correo electrónico como prueba digital y en consecuencia la condición de no repudiabilidad de la prueba.
Objetivos Específicos	
	• Representar el proceso de transmisión del correo y derivar de ella la trazabilidad del envío. • Identificar la información correspondiente a las cuentas y equipos que intervienen en la transmisión.
Alcance	
	La ontología comprende la representación de la trazabilidad del correo electrónico a partir del contenido de la cabecera del mismo, con el agregado de datos externos a la cabecera y que son necesarios para dar respuesta a las preguntas de competencia
Lenguaje de Implementación	
	El lenguaje formal utilizado en la implementación de OntoFoCE será el lenguaje estándar OWL.
Usuarios Finales	
	Profesionales informáticos que actúan como Peritos Judiciales y que – por el volumen de datos o complejidad de la causa- requieren de procesos automatizados para realizarla.
Usos Previstos	
	• Verificación de la validez de la cabecera de un correo electrónico para que pueda considerarse como prueba digital periciable.

	• Realización del análisis forense de correos electrónicos a partir de la cabecera de los mismos. • Obtención de información relevante a la pericia a partir de los datos resultantes de las preguntas de competencia.
Requerimientos de la Ontología	
	Requerimientos Funcionales:
	PC01: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico? PC02: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico? PC03: Dado un correo CE ¿A qué cuentas se remitió el correo? PC04: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del emisor? PC05: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del Receptor? PC06: Dado un correo CE ¿Cuál fue el cliente de correo utilizado por cada usuario? PC07: Dado un correo CE ¿Cuál fue el equipo desde el cual se emitió el correo? PC08: Dado un correo CE ¿Cuál fue el equipo en el que se recibió el correo? PC09: Dado un correo, un emisor y un receptor ¿cuál es la secuencia de equipos por los que pasó ese correo? PC10: Dado una cuenta C ¿cuáles son los correos que emitió? PC11: Dado una cuenta C ¿cuáles son los correos que recibió? PC12: Dado una cuenta C1 ¿se ha emitido un correo hacia la cuenta C2? PC13: Dado una cuenta C1 ¿se ha recibido un correo desde la cuenta C2? PC14: Dada una dirección IP ¿cuál sería la localización geográfica del mismo? PC15: ¿Cuáles son los correos que han pasado por el dispositivo que posee una IP dada? PC16: ¿Cuáles son los mails enviados desde una determinada cuenta en una fecha dada? PC17: ¿Cuáles son los mails recibidos por una determinada cuenta en una fecha dada? PC18: Dada una palabra clave ¿Figura en el asunto de un correo? PC19: Dada una palabra clave ¿Figura en el cuerpo de un correo? PC20: Dada una palabra clave ¿Figura en el adjunto de un correo? PC21: ¿Cuáles son los correos intercambiados entre las cuentas C1 y C2 en un rango de fechas dado?
	Pre-Glosario de Términos de las preguntas de competencia (nombre y frecuencia de aparición):
	• Correo electrónico: 17 • Cuentas de correo electrónico: 10 • Fecha y hora (de paso del correo en un equipo o servidor): 5 • Dirección IP (del equipo o servidor): 5 • Equipo: 3 • Palabra Clave: 3 • Alias de usuario: 2 • Cliente de correo: 1 • Localización geográfica: 1 • Asunto: 1 • Cuerpo: 1 • Adjunto: 1

3.4 Modelo Conceptual de OntoFoCE

Como resultado de la fase de conceptualización se obtiene un modelo conceptual que define los conceptos y relaciones que conforman la ontología. Siguiendo las guías de Methontology el modelo conceptual de OntoFoCE se define utilizando representaciones intermedias no formales. En particular, se utilizaron diagramas de clases UML, que permiten mostrar gráficamente los conceptos, así como sus atributos y relaciones. Estos diagramas se complementaron con tablas que representan el Diccionario de Conceptos y el de Relaciones. La primera contiene la definición de cada concepto, sus atributos de instancia y relaciones en la que participa. La segunda tabla que se menciona especifica las relaciones definidas en la ontología y sus inversas, indicando para cada una de ellas el dominio y rango

correspondiente. Estas representaciones intermedias se muestran detalladamente en el ANEXO III: REPRESENTACIONES INTERMEDIAS DE LA CONCEPTUALIZACIÓN DE OntoFoCE.

A continuación, se describe el modelo conceptual de OntoFoCE. Para una mejor organización de la presentación, este modelo se introduce mostrando diferentes vistas, a saber:

- Representación del Concepto de Correo Electrónico.
- Representación del Proceso de Transmisión
- Representación de Conceptos Complementarios

En las siguientes secciones se utilizan diagramas de clases UML para ilustrar cada una de estas vistas. Como se mencionó al inicio de la sección 3.2, la interpretación de los modelos se restringe por medio de la definición de axiomas y reglas en DL (utilizando la sintaxis Manchester de OWL) y SWRL. Estos axiomas y reglas permiten también inferir conocimiento implícito a partir de los conceptos y relaciones explícitas de la ontología.

En los diagramas de clases se representan los conceptos y relaciones cuyas instancias se infieren mediante axiomas y/o reglas, utilizando clases derivadas y asociaciones derivadas. Este tipo de elementos se identifican en un diagrama de clases porque tienen al inicio de su nombre el símbolo "\".

Asimismo, cabe aclarar que en OWL las relaciones son unidireccionales y a los fines de mantener la navegabilidad de una clase a la siguiente y viceversa se crean dos propiedades de una misma relación, una definida explícitamente y la otra derivada como relación inversa e inferida oportunamente por el razonador. Debido a que en las reglas, axiomas y en las consultas SPARQL que se muestran en este capítulo se usan indistintamente las propiedades explícitas y las inferidas, se decidió representar las relaciones entre clases mediante dos asociaciones unidireccionales (que son las que se implementaron luego en OWL) en lugar de una única bidireccional (definida originalmente en la etapa de conceptualización). En el siguiente sitio: https://digilab.ucasal.edu.ar/owl/ se encuentra publicado el código OWL de OnfoFoCE, según se implementó en Protégé Ver. 5.5.0.

Por otra parte, en referencia a los nombres de los conceptos que se utilizan en los axiomas y reglas, si no se indica un espacio de nombre, se trata de conceptos definidos en la ontología. Para los conceptos definidos en otras ontologías, se indican los espacios de nombres correspondientes.

3.4.1 Representación del Concepto de Correo Electrónico

En el apartado sobre objeto de estudio de esta tesis (sección 2.3 del Capítulo 2) se indicó que a los fines del análisis forense se define *correo electrónico* como *documento digital que consta de dos partes: a) una cabecera que contiene información sobre el proceso de transmisión que se desarrolla con identificación de las cuentas intervinientes y los distintos servidores en que el correo se fue almacenando durante la transmisión; y b) un cuerpo que contiene el mensaje que se transmite más los archivos adjuntos que opcionalmente integran el mensaje.*

Si bien el usuario sólo ve algunos de los datos de un correo (remitente, destinatario, asunto, cuerpo del mensaje y archivo adjunto), el análisis pericial requiere de información técnica asociadas a esos datos, la cual se encuentra en un archivo de texto plano comúnmente denominado *cabecera o encabezado (*ver sección 2.3.1 del Capítulo 2 sobre la Estructura del Correo Electrónico), que además de contar con toda la información visible al usuario desde la pantalla de la aplicación cliente en la que se envía/recibe el correo, incluye los datos referidos al proceso de transmisión del correo.

Con el fin de poder representar todos los componentes de la cabecera del correo, el modelo propuesto en OntoFoCE incorpora conceptos que representan las diversas partes del correo electrónico.

De este modo, el correo electrónico se define mediante un conjunto de conceptos diferenciados: **CABECERA, CUENTA DE CORREO, CUERPO, ASUNTO, ADJUNTO** y **OCURRENCIAS**, cuyas representaciones se van a describir en los párrafos que siguen. Pero además de estos componentes, es preciso representar el concepto vinculante de todos ellos, así, se considera el concepto de **CORREO** entendiendo por tal la entidad abstracta que contiene todas las partes desagregadas del correo electrónico que está siendo analizado.

En la Figura 3-2 se presenta el diagrama de clases UML que muestran las clases que representan los conceptos propuestos en OntoFoCE para definir un correo electrónico. Se describen a continuación cada una de las clases incluidas en el gráfico, así como las relaciones que las vinculan.

El documento digital sobre el cual se realiza el análisis forense se representa en OntoFoCE mediante la clase ***Correo*** y su atributo es *idCorreo* que actúa como identificador único de dicho documento.

Durante el proceso de transmisión, el correo electrónico se va almacenando en cada equipo o servidor por el que pasa. Estas copias, en una trasmisión sin adulteración expresa del correo deberían ser iguales, en cuanto al mensaje que transmiten en el cuerpo y los archivos adjuntos, pero difieren en la cabecera, ya que en cada servidor, se agregan a la misma los datos de identificación del mismo.

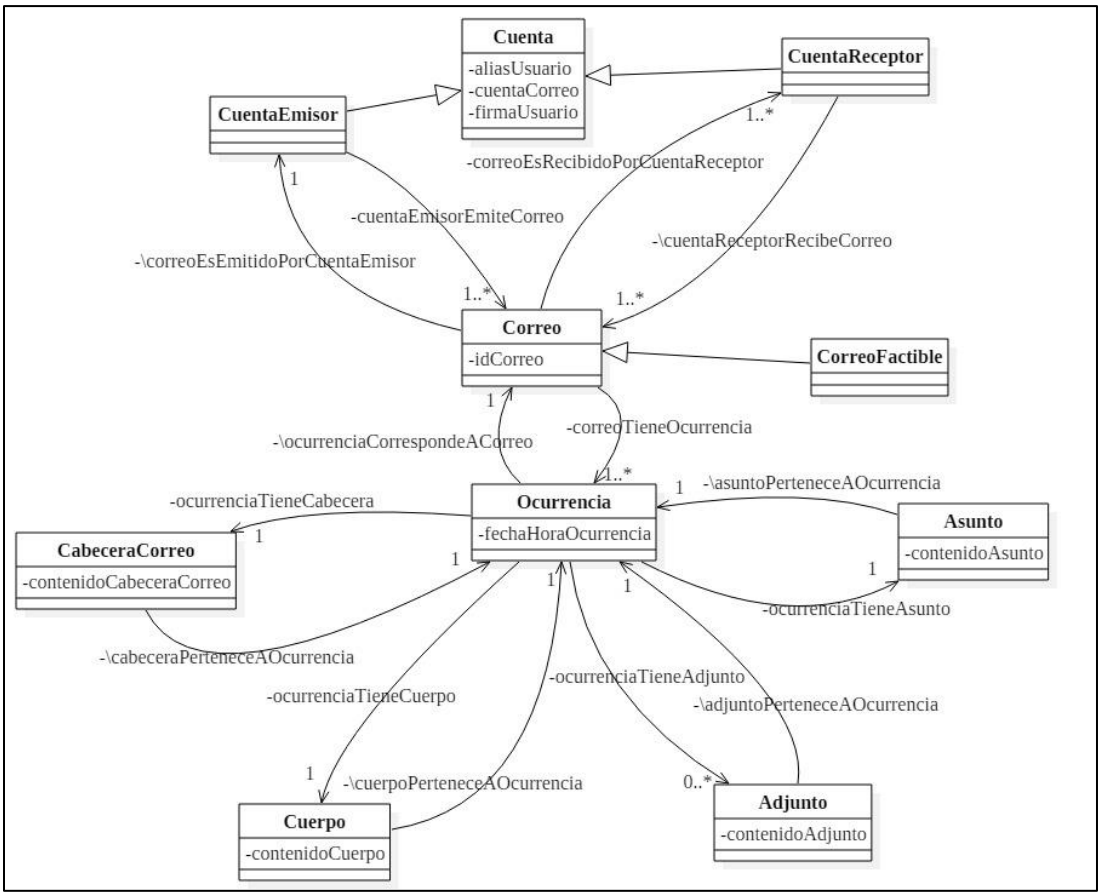


Figura 3-2: Vista de Representación del Concepto de Correo

Es decir: cada vez que el correo llega a un servidor, el proceso agrega al inicio de la cabecera los datos de dirección IP, fecha y hora en que arribó al servidor. A fin de representar las sucesivas copias de un correo, se incorpora a OntoFoCE el concepto de **OCURRENCIA,** que se define como "Copia del correo electrónico que se almacena en cada dispositivo que participa en el proceso de transmisión". Este concepto es representado en la figura mediante la clase ***Ocurrencia,*** para la cual se define el atributo *fechaHoraOcurrencia* que indica la fecha y hora en la que la copia del correo es almacenada en el equipo/servidor.

Ambas clases, ***Correo*** y ***Ocurrencia***, se vinculan mediante una relación derivada denominada *correoTieneOcurrencia*. Esta relación permite vincular todas las ocurrencias que corresponden a un mismo correo por ello su cardinalidad señala que

un correo puede vincularse a varias ocurrencias y la relación inversa se define como *ocurrenciaCorrespondeACorreo* e indica que cada ocurrencia se vincula a un único correo.

La relación derivada entre ***Correo*** y ***Ocurrencia*** se expresa en términos de SWRL de la siguiente manera:

Correo(?x),Secuencia(?s),Hilo(?h),Ocurrencia(?o),
correoTieneSecuencia(?x, ?s),secuenciaTieneHilo(?s, ?h), hiloTieneOcurrencia(?h, ?o),
-> correoTieneOcurrencia (?x, ?o)

El correo electrónico es un mensaje que se transmite desde un emisor hacia uno o varios receptores. Ese proceso requiere de dos conceptos primarios: el correo y las cuentas de correo electrónico de origen y destino. Se define **CUENTA** como "servicio online que provee un espacio para la recepción, envío y almacenamiento de mensajes de correo electrónico", se asocia a un único usuario que actúa como emisor/receptor del correo. Este concepto se representa en OntoFoCE mediante la clase ***Cuenta***, cuyos atributos son:

- *aliasUsuario* representa el nombre, apodo o sinónimo elegido por el usuario para identificar su cuenta,
- *cuentaCorreo* identifica el nombre de la cuenta bajo la estructura de *usuario@dominio* en el que *usuario* es el identificador único de la cuenta dentro del *dominio* correspondiente
- *firmaUsuario* corresponde a la sigla o párrafo con el cual el usuario se describe a sí mismo como autor del correo, no siempre está presente.

En la práctica, el usuario utiliza su cuenta para realizar las dos acciones del servicio, o sea, para enviar correos electrónicos y para recibir correos electrónicos de otros usuarios. Una misma cuenta puede utilizarse para emitir y/o para recibir correos, por ello es necesario identificar cual es el rol que cumple la cuenta en cada correo que se instancie, de allí la necesidad de identificar ambas situaciones. Para ello se introduce en la ontología como subclases de Cuenta, las subclases ***CuentaEmisor*** y ***CuentaReceptor***, y la clase ***Correo*** ya mencionada.

Considerando estos tres conceptos: **CORREO, CUENTA_EMISOR** y **CUENTA_RECEPTOR,** el vínculo entre ellos se establece en función de la acción de emisión/recepción que en cada instancia señala como actúa la cuenta. Esto se refleja en OntoFoCE mediante las relaciones *cuentaEmisorEmiteCorreo* y *cuentaReceptorRecibeCorreo* que a continuación se describen.

La relación *cuentaEmisorEmiteCorreo* vincula el correo y la cuenta que actúa como emisora, es decir, vincula la clase ***CuentaEmisor*** con la clase ***Correo***. Esta relación permite identificar la cuenta que da origen al correo electrónico y es exigencia que el correo tenga una cuenta que actúa como emisora y ésta sea única, pero desde una cuenta pueden emitirse más de un correo. Asimismo se definió la relación inversa entre ***Correo*** y ***CuentaEmisor***, denominada *correoEsEmitidoPorCuentaEmisor* que vincula cada correo con su correspondiente cuenta emisor, y puede ocurrir que una ***CuentaEmisor*** tenga varias instancias de ***Correo***.

En igual sentido, la relación *cuentaReceptorRecibeCorreo* representa el vínculo entre la clase ***CuentaReceptor*** y la clase ***Correo***. Esta relación permite identificar las cuentas a las cuales se remite el correo, su cardinalidad es múltiple en ambos extremos de la relación, pues si el usuario utiliza una lista de distribución o la opción ”con copia a” o “con copia oculta a”, es posible enviar el correo a más de una cuenta, y éstas a su vez pueden recibir más de un correo. La relación inversa a esta relación es *correoEsRecibidoPorCuentaReceptor.*

Las relaciones introducidas en los párrafos previos permiten definir las clases ***CuentaEmisor*** y ***CuentaReceptor***, las cuales representan los roles de emisión/recepción de una Cuenta. Los siguientes axiomas definen estos conceptos:

Class: CuentaEmisor
EquivalentTo: Cuenta and (cuentaEmisorEmiteCorreo some Correo) and
(cuentaEmisorEmiteCorreo only Correo)

Este axioma indica que toda instancia de CuentaEmisor es todo individuo que es una instancia de la clase Cuenta que se vincula con al menos una instancia de Correo mediante la relación *cuentaEmisorEmiteCorreo* y que, para toda instanciación de dicha relación se vincula solamente con instancias de la clase Correo.

De manera análoga, el siguiente axioma define el concepto de *CuentaEmisor*:

Class: CuentaReceptor
EquivalentTo: Cuenta and (cuentaReceptorRecibeCorreo some Correo)
and (cuentaReceptorRecibeCorreo only Correo)

Con el objetivo de restringir las Clases con las cuales puede relacionarse la clase ***Correo*** mediante las relaciones *cuentaReceptorRecibeCorreo y cuentaEmisorEmiteCorreo* se define el siguiente axioma de clase:

Class: Correo
SubClassof: (correoEsRecibidoPorCuentaReceptor only CuentaReceptor) and
(correoEsRecibidoPorCuentaReceptor some CuentaReceptor) and
(correoEsEmitidoPorCuentaEmisor only CuentaEmisor) and
(correoEsEmitidoPorCuentaEmisor some CuentaEmisor)

la cual expresa que toda instancia de correo debe estar relacionada, mediante la relación *correoEsRecibidoPorCuentaReceptor* con al menos una instancia de la clase ***CuentaReceptor*** y, además para toda instancia de la relación mencionada que dicha instancia de Correo tenga, debe estar vinculada solamente con instancias de ***CuentaReceptor***. La misma restricción se impone sobre las instancias de la relación *correoEsEmitidoPorCuentaEmisor* que vinculan una instancia de Correo con al menos una instancia de ***CuentaEmisor*** y sólo con instancias de ***CuentaEmisor***.

A esta restricción se suma la que indica la unicidad de la cuenta que actúa como emisora para un correo: "*Todo Correo tiene una única Cuenta que actúa como emisora*". Dicha restricción se expresa en términos de SWRL de la siguiente manera:

Correo(?x), CuentaEmisor(?e1), CuentaEmisor(?e2),
cuentaEmisorEmiteCorreo(?e1, ?x), cuentaEmisorEmiteCorreo(?e2, ?x)
-> SameAs (?e1, ?e2)

Esta expresión indica que si la variable ?x representa una instancia de clase ***Correo*** y *?e1, ?e2* representan sendas instancias de ***CuentaEmisor*** que se vinculan con el correo ?x mediante las correspondientes instancias de la relación *cuentaEmisorEmiteCorreo* (*cuentaEmisorEmiteCorreo(?e1,?x)* y de *cuentaEmisorEmiteCorreo(?e2,?x))* implica que ?e1 y ?e2 representan el mismo individuo. El predicado SameAs(?e1,?e2) está predefinido en SWRL y evalúa verdadero si ?e1 y ?e2 son el mismo individuo. Si este predicado falla, implica que el correo ?x está vinculado con dos cuentas emisoras (?e1 y ?e2), situación que no es válida en el dominio.

Hasta aquí se describió la parte superior de la Vista de Definición de Correo (Figura 3-2), corresponde describir la representatividad de las ocurrencias de un correo y los componentes del correo vinculados a ellas: Asunto, Adjunto, Cuerpo y Cabecera.

Si bien hasta el momento se ha considerado el correo electrónico como un todo, conviene aclarar que a los fines de representarlo convenientemente en OntoFoCE, resultó necesario particionar su representación utilizando distintos conceptos: **CABECERA_DE_CORREO**, **ASUNTO**, **CUERPO** y **ADJUNTO**, esto es así ya que cada uno de estos conceptos cumple una función propia en el análisis forense del correo electrónico: de la cabecera del correo se obtiene la información para la trazabilidad, sobre el contenido del asunto, cuerpo y adjunto se realizan búsqueda por palabras claves de interés para la causa judicial.

Como se mencionara antes, cada una de estas partes, se mantienen en cada copia del correo, por lo cual las clases ***Cabecera, Asunto***, ***Cuerpo*** y ***Adjunto*** se vinculan con la clase ***Ocurrencia,*** para representar las diferentes partes de cada copia del correo.

El concepto **CABECERA_CORREO** se define como "Bloque de texto plano que contiene información relativa al correo y al proceso de transmisión realizado". En OntoFoCE se encuentra representado por la clase ***CabeceraCorreo*** y en el único atributo que los describe *contenidoCabeceraDeCorreo*. Las clases ***Ocurrencia*** y ***CabeceraCorreo*** se vinculan mediante la relación *ocurrenciaTieneCabecera* cuya cardinalidad indica que cada ocurrencia debe mantener una única cabecera. La relación inversa es *cabeceraPerteneceAOcurrencia*.

El componente del correo denominado **ASUNTO** es un concepto que se define como "Texto que expresa el tema del que trata el correo electrónico", está representado por la clase de nombre homónimo y contiene un único atributo denominado *contenidoAsunto*. La relación *asuntoPerteneceAOcurrencia* vincula la clase ***Asunto*** con la clase ***Ocurrencia*** con una cardinalidad que expresa que cada ***Asunto*** se vincula a su correspondiente ***Ocurrencia***. La relación inversa es *ocurrenciaTieneAsunto*.

El mensaje contenido en el correo electrónico, comúnmente denominado **CUERPO** se representa en OntoFoCE mediante la clase del mismo nombre y contiene un único atributo denominado *contenidoCuerpo* que almacena el cuerpo del mensaje del correo electrónico. La relación *cuerpoPerteneceAOcurrencia* vincula la clase ***Cuerpo*** con ***Ocurrencia*** expresando mediante su cardinalidad que cada instancia de ***Cuerpo*** se relaciona con su correspondiente ***Ocurrencia***. La relación inversa es *ocurrenciaTieneCuerpo*.

Otro elemento complementario para el análisis pericial es el concepto de **ADJUNTO,** definido como "Archivo asociado al correo electrónico con información complementaria al contenido del correo", cabe mencionar que este elemento es opcional en el correo electrónico, y que de estar presente, puede estar conformado por uno o varios archivos. Este concepto se representa en OntoFoCE mediante la clase ***Adjunto*** y contiene un único atributo denominado *contenidoAdjunto*. La relación *adjuntoPerteneceAOcurrencia* vincula la clase ***Adjunto*** con la clase ***Ocurrencia***, manteniendo una cardinalidad que permite representar que cada instancia de ***Adjunto*** corresponde a una única instancia de

Ocurrencia, mientras que la relación inversa *ocurrenciaTieneAdjunto* vincula *Ocurrencia* con *Adjunto* y su cardinalidad representa los distintos archivos adjuntos (ninguno o varios) que corresponden a cada ***Ocurrencia***. Es importante aclarar que en la presente propuesta no se incluye el análisis *del contenido del archivo Adjunto*, y se deja para investigaciones futuras el problema de resolver como instanciar cada archivo de modo que luego pueda analizarse su contenido, considerando que el archivo adjunto puede contener una gran diversidad de formatos de texto enriquecido, imagen, video o formatos propietarios de aplicaciones informáticas (DWG, ZIP, WAW, etc.).

Por otra parte, para realizar el análisis forense, es necesario que el correo cumpla con determinados requisitos (debe tener una cabecera, y en ella deben figurar las direcciones IP de envío y recepción del correo). Esta condición se establece en OntoFoCE mediante la subclase ***CorreoFactible***, que se explica en detalle más adelante, cuando se describe la representación de las ocurrencias en la sección 3.4.2.3.

Por otra parte, cada instancia de ocurrencia se asocia a una única instancia de cada uno de los conceptos **ASUNTO, CUERPO y CABECERA** correspondientes, ya que así se representan las partes de cada copia del correo que está siendo almacenado en los equipos y servidores de paso.

En el caso de las clases ***Cabecera, Asunto*** *y* ***Cuerpo***, la única restricción que se establece para estas clases es que sean únicas para cada instancia de la clase ***Ocurrencia***. Esto se define mediante las siguientes reglas: *Toda Ocurrencia tiene una única Cabecera*

Ocurrencia (?o), CabeceraCorreo(?ca1), CabeceraCorreo(?ca2),
cabeceraPerteneceAOcurrencia (?ca1, ?o), cabeceraPerteneceAOcurrencia (?ca2, ?o)
-> SameAs (?ca1, ?ca2)

Con esta regla se establece la unicidad de ***Cabecera*** respecto de ***Ocurrencia,*** indicando que para toda ocurrencia *?o*, si existen dos cabeceras denominadas *?ca1* y *?ca2* que se relacionan a *?o* mediante la relación *cabeceraPerteneceAOcurrencia*, entonces *?ca1* y *?ca2* deben representar la misma cabecera.

Por otra parte se debe establecer la unicidad de ***Asunto*** respecto de ***Ocurrencia,*** que se puede expresar mediante la siguiente regla: *Toda Ocurrencia tiene un único Asunto.*

Ocurrencia (?o), Asunto(?a1), Asunto(?a2),
asuntoPerteneceAOcurrencia(?a1, ?o),asuntoPerteneceAOcurrencia(?a2, ?o)
-> SameAs (?a1, ?a2)

Esta regla expresa la unicidad de ***Asunto*** respecto de ***Ocurrencia,*** indicando que para toda ocurrencia *?o*, si existen dos asuntos *?a1* y *?a2* que se relacionan a *?o* mediante la relación *asuntoPerteneceAOcurrencia*, entonces, *?a1* y *?a2* representan el mismo asunto.

De igual forma, se expresa la regla sobre unicidad del ***Cuerpo*** respecto de la ***Ocurrencia,*** diciendo: *Toda Ocurrencia tiene un único Cuerpo*

Ocurrencia(?o), Cuerpo(?cu1), Cuerpo(?cu2),
cuerpoPerteneceAOcurrencia(?cu1, ?o), cuerpoPerteneceAOcurrencia(?cu2, ?o)
-> SameAs (?cu1, ?cu2)

Esta regla expresa la unicidad de ***Cuerpo*** respecto de ***Ocurrencia,*** y señala que para toda ocurrencia *?o*, si se relacionan a ella dos instancias de la clase *Cuerpo* denominados *?cu1* y *?cu2* mediante la relación *cuerpoPerteneceAOcurrencia,* entonces, *?cu1* y *?cu2* deben ser el mismo individuo.

3.4.2 Representación del Proceso de Transmisión

Definida la representación del correo electrónico, corresponde ahora describir la representación del proceso de transmisión en OntoFoCE. Es importante considerar aquí que dicho proceso tiene tres momentos (emisión, transmisión interna y recepción), los cuales debe modelarse para derivar luego la *trazabilidad* de los envíos.

La Figura 3-3 muestra la vista parcial, correspondiente al proceso de transmisión del correo electrónico.

El proceso de emisión/recepción de un correo electrónico requiere de dos componentes básicos: los equipos (o hardware) utilizados para la emisión/recepción y los gestores de correo electrónico (o software) denominadas comúnmente Clientes de Correo. Pero además, es de interés especial para el análisis forense, la información que permite señalar e identificar a los usuarios actuantes a partir de los equipos utilizados. Así, se describen a continuación los conceptos más relevantes del proceso de transmisión: **EQUIPO, IDENTIFICACIONEQUIPO** y **CLIENTECORREO**.

El concepto **EQUIPO** se define como “Componente de hardware que almacena el correo electrónico” y representa el dispositivo (PC, Celular, Notebook, servidor de correo, etc.) utilizado durante el proceso de transmisión del correo electrónico.

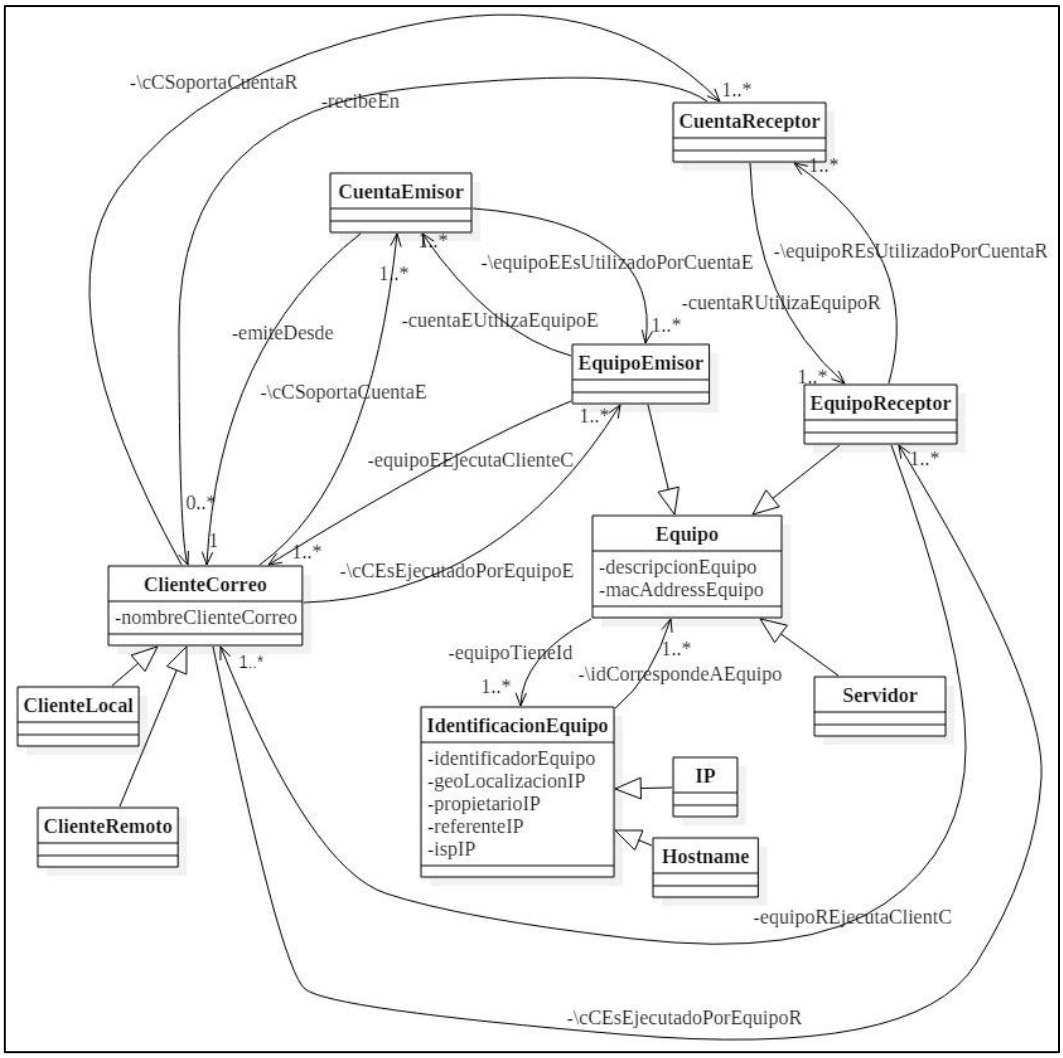


Figura 3-3: Vista de Representación del Proceso de Transmisión

En la ontología este concepto está representado por la clase ***Equipo*** y sus atributos son:

- *descripcionEquipo* que contiene un breve detalle de las características del equipo obtenidas por el perito al momento de realizar la pericia, y
- *macAddressEquipo* que es un número que identifica a cada equipo dentro de una red, y que se obtiene del registro inicial del dispositivo que realiza el perito.

El concepto **IDENTIFICACION_EQUIPO** se define como "Identificación única del hardware conectado a internet", y comprende el conjunto de datos referidos a la ubicación del equipo en el contexto de la red utilizada durante la transmisión del correo electrónico. Este concepto se identifica en OntoFoCE mediante la clase del mismo nombre y sus atributos son:

- *identificadorEquipo,* nombre de dominio o dirección IP asignada al equipo que está contenido en la cabecera del correo electrónico,

- *geolocalizacionIP*, que contiene los datos de ubicación geográfica de la dirección IP, en valores de longitud y latitud geográfica
- *propietarioIP*, nombre de empresa o institución responsable de esa dirección IP según los registros de internet[48],
- *referenteIP*, nombre de la persona indicada como referente de la empresa o institución responsable de la dirección IP,
- *ispIP*, nombre del proveedor del servicio de internet

Las direcciones IP de los equipos que intervienen en el proceso de transmisión figuran en la cabecera del correo electrónico. Según las características y funcionalidades de los distintos protocolos y aplicaciones utilizadas para la transmisión (que dependen además de las condiciones del servicio y características de los proveedores) la dirección IP puede figurar en el encabezado ya sea de manera expresa, bajo el formato IP4 o IP6, o bien se indica en términos del dominio de procedencia.

Para representar estas dos formas de identificación, OntoFoCE especializa la clase ***IdentificacionEquipo*** en las subclases ***IP*** *y* ***Hostname***. Es necesario establecer esta especialización ya si en la cabecera figura un nombre de dominio en vez de una dirección IP, es probable que dicho dominio tenga una dirección IP dinámica, es decir, se obtenga diferentes direcciones IP según la fecha o momento en que se realice la consulta para identificarla. Y esta situación es importante reflejarla en la ontología ya que, las direcciones IP encontradas en la pericia se toman luego como referencia para la ubicación geográfica de los usuarios, de modo que, si no se tiene certeza de una dirección IP, se debe expresar tal situación en el informe pericial. Es decir, cuando se trata de una hostname, igualmente se pueden obtener los valores de los atributos *geolocalizaciónIP*, *propietarioIP*, etc., pero no se tiene certeza que esos valores sean correctos, ya que es posible que sean los que tenga esa IP asignada en la actualidad al momento en que se realiza la consulta en el directorio público de IPs, y no los que tenía al momento en que el correo pasó por el servidor cuyo hostname se está analizando.

Considerando los conceptos **EQUIPO** e **IDENTIFICACION_EQUIPO**, las relaciones entre ambos se establecen a partir de la necesidad de identificar los datos

[48] Son varias las instituciones que registran los datos sobre los dominios y direcciones IP, entre ellas: LANIC (Registro de Direcciones de Internet de América Latina y Caribe), en la que a la fecha se encuentran registradas más de 800 empresas e instituciones argentinas. (https://www.lacnic.net/971/1/lacnic/nuestros-asociados consultado el 15/02/2019).

referidos a la dirección IP y ubicación geográfica de cada equipo. Dado que esos datos pueden variar debido a la asignación aleatoria de las direcciones IP que puede realizarse desde el proveedor del servicio, es posible que un mismo equipo tenga asignadas más de una identificación, de allí la necesidad de particionar los datos de identificación del equipo en una clase separada.

La relación *equipoTieneId* permite asociar el equipo a un conjunto de datos de identificación, es decir, vincula la clase ***Equipo*** con la clase ***IdentificacionEquipo***, señalando que un equipo puede contener varias identificaciones, y una identificación puede corresponder a más de un equipo. Asimismo se definió la relación inversa entre ***IdentificacionEquipo*** y ***Equipo***, denominada *idCorrespondeAEquipo*.

Ahora bien, durante el proceso de transmisión, se identifican tres tipos de equipos según sea la función que cumplen:

- equipo emisor: utilizado por el usuario para escribir y enviar el correo
- servidores: utilizados por el servicio de correo electrónico durante la transmisión del mismo; y
- equipo receptor: utilizado por el usuario para recibir y leer el correo.

Esta tipificación se expresa en la ontología mediante las subclases ***EquipoEmisor***, ***Servidor*** y ***EquipoReceptor*** que son especializaciones de la clase ***Equipo***, y se explican en las secciones siguientes según los tres momentos del proceso de envío del correo electrónico: emisión, transmisión interna y recepción.

Para explicar el proceso de envío y recepción de un correo falta definir el concepto **CLIENTE_CORREO,** que se define como “Aplicación informática que gestiona una cuenta de correo electrónico”, y se refiere al software utilizado por el usuario para acceder a la cuenta de correo.

En OntoFoCE se representa este concepto mediante la clase ***ClienteCorreo*** y cuenta con un único atributo llamado *nombreClienteCorreo,* que señala la denominación de esa aplicación informática.

Es de interés para el análisis pericial, la identificación del cliente remoto o cliente local utilizado para la emisión/recepción del mensaje, ya que de tratarse de un cliente local, se puede encontrar una copia del correo electrónico en el disco o medio de almacenamiento del equipo utilizado, mientras que en el caso de un cliente remoto, el correo electrónico se almacena en los servidores de correo. El acceso a una copia del correo electrónico enviado/recibido es más rápido si se trata de un cliente de correo local, ya que con la orden del Juez es suficiente para acceder a esos

datos, mientras que si el correo electrónico se encuentra en los servidores de correo, se debe realizar un procedimiento judicial engorroso (a veces involucra a la justicia internacional) para que los propietarios de esos servidores entreguen una copia del correo electrónico. De allí que la clase ***ClienteCorreo*** se especializa en dos subclases: ***ClienteLocal*** y ***ClienteRemoto***.

Consideremos a continuación la descripción detallada del proceso de transmisión pero identificando los tres momentos que la componen: la emisión, la transmisión en sí y la recepción del correo electrónico.

3.4.2.1 Proceso de Emisión de un Correo Electrónico

Cuando se ***envía un correo*** el usuario accede a la cuenta (que actúa como cuenta del emisor) mediante la aplicación informática (que actúa como cliente de correo) para escribir el correo. Esto se representa mediante la relación *emiteDesde* que vincula la ***CuentaEmisor*** con el ***ClienteCorreo*** utilizado, establecida de manera de representar la cardinalidad múltiple de esta relación, es decir, una cuenta emisor puede utilizar distintos clientes de correo, y a su vez, cada cliente de correo puede gestionar varias cuentas. La relación inversa denominada *cCSoportaCuentaE* vincula ***ClienteCorreo*** con ***CuentaEmisor***, muestra que un mismo cliente de correo puede ser usado a la vez por una o más cuentas.

Por otra parte, la ejecución de la aplicación informática que actúa como cliente de correo, requiere de un dispositivo en el cual esté instalada. Esta relación, que vincula la clase ***Equipo*** con la clase ***ClienteCorreo*** se describe como *equipoEEjecutaClienteC,* y su cardinalidad expresa que un equipo puede contener a varios clientes de correo, y que un mismo cliente de correo puede estar instalado en varios equipos. La relación inversa se denomina *cCEsEjecutadoPorEquipoE* y vincula ***ClienteCorreo*** con ***EquipoEmisor***.

Cuando se efectúa el envío, el acceso al cliente de correo se realiza mediante un equipo, que actúa como equipo emisor. De modo que la relación *cuentaEUtilizaEquipoE* permite identificar cual es el equipo utilizado por el usuario para acceder a la cuenta del emisor y enviar el correo, es decir, vincula la clase ***CuentaEmisor*** con la clase ***EquipoEmisor***, cuya cardinalidad indica que una cuenta puede utilizarse desde varios equipos y un equipo puede ser utilizado por más de una

cuenta que actúa como emisora. Asimismo se definió la relación inversa entre ***EquipoEmisor*** y ***CuentaEmisor***, denominada *equipoEEsUtilizadoPorCuentaE*. Para las clases que se vinculan mediante la relación *emiteDesde* durante el envío del correo se definen la regla de inferencia que dice "*Toda Cuenta que actúa como emisora utiliza un Equipo Emisor que ejecuta un Cliente de Correo"* expresadas en términos de SWRL de la siguiente manera:

CuentaEmisor(?e), EquipoEmisor(?q), ClienteCorreo(?cc),
cuentaEUtilizaEquipoE(?e, ?q), equipoEEjecutaClienteC(?q, ?cc),
-> emiteDesde(?e, ?cc)

Esta regla expresa que para toda cuenta que interviene cómo emisora *?e*, si ésta se relaciona con un equipo emisor *?q* mediante la relación *cuentaEUtilizaEquipoE,* y el equipo *?q* se relaciona con un cliente de correo *?cc* mediante la relación *equipoEEjecutaClienteC,* entonces, es posible inferir que la cuenta del emisor *?e* se vincula con el cliente de correo *?cc* mediante la asociación *emiteDesde.*

3.4.2.2 Proceso de Recepción de un Correo Electrónico

De igual modo, se puede considerar la relación entre las subclases ***EquipoReceptor***, ***CuentaReceptor*** y ***ClienteCorreo*** para explicar el proceso de **recepción de un correo.**

Para recibir un correo el usuario accede a la cuenta (que actúa como cuenta del receptor esta vez) mediante la aplicación informática (que actúa como cliente de correo). Esto se representa mediante la relación *recibeEn* que vincula una ***CuentaReceptor*** con un ***ClienteCorreo*** utilizado. La relación inversa, señalada como *cCSoportaCuentaR* vincula ***ClienteCorreo*** con ***CuentaReceptor*** manteniendo una cardinalidad 1-N, ya que un mismo cliente de correo puede utilizarse simultáneamente por las cuentas que se están analizando.

Por otra parte, la ejecución de la aplicación informática que actúa como cliente de correo y el dispositivo en el cual está instalada se describe como *equipoREjecutaClienteC,* manteniendo una cardinalidad que permite relacionar un equipo receptor a varios clientes de correo y viceversa, esto último se expresa mediante la relación inversa denominada *cCEsEjecutadoPorEquipoR* que vincula ***ClienteCorreo*** con ***EquipoReceptor***.

Además, al momento de la recepción, el acceso al cliente de correo se realiza mediante un equipo, que actúa como equipo receptor. De modo que la relación

cuentaRUtilizaEquipoR permite identificar cual es el equipo utilizado por el usuario para acceder a la cuenta que actúa como receptora y recibir el correo, es decir, vincula la clase ***CuentaReceptor*** con la clase ***EquipoReceptor***. Esta relación (representada con las asociaciones unidireccionales *cuentaRUtilizaEquipoR* y *equipoREsUtilizadoPorCuentaR* en el diagrama de la Figura 3-3) mantiene una cardinalidad múltiple entre ambos extremos, ya que una cuenta puede utilizarse desde varios equipos y un equipo puede ser utilizado por más de una cuenta en carácter de receptora, esto último expresado en la relación inversa entre ***EquipoReceptor*** y ***CuentaReceptor***, denominada *equipoREsUtilizadoPorCuentaR*. Utilizando las relaciones presentadas en los párrafos previos es posible establecer una regla que permita inferir la relación *recibeEn* vinculando una cuenta receptora con el cliente utilizado. Esta regla podría expresarse como "*Toda Cuenta que actúa como receptora utiliza un Equipo Receptor que ejecuta un Cliente de Correo"* y se expresada en términos de SWRL:

CuentaReceptor(?r), EquipoReceptor(?q), ClienteCorreo(?cc),
cuentaRUtilizaEquipoR(?r, ?q), equipoREjecutaClienteC(?q, ?cc)
-> recibeEn(?r, ?cc)

La regla anterior especifica que para toda cuenta que actúa como receptora *?r*, si ésta se relaciona con un equipo receptor *?q* mediante la relación *cuentaRUtilizaEquipoR*, y el equipo *?q* se relaciona con un cliente de correo *?cc* mediante la relación *equipoREjecutaClienteC*, entonces, es posible inferir que *?re* se vincula con *?cc* mediante la asociación *recibeEn.*

Respecto de la tercera subclase de ***Equipo*** denominada ***Servidor***, no se vincula a ninguna de las cuentas de correo, sino que identifica el equipo utilizado para almacenar las copias del correo electrónico que se van almacenando en los servidores intermedios del proceso de transmisión, es decir aquellas que no se almacenan en el equipo emisor ni en el equipo receptor, y cuya representación se introduce en la siguiente sección.

3.4.2.3 Representación de Las Ocurrencias

Como se explicó en el Capítulo 2, la condición de no repudio de un correo electrónico que se presenta como evidencia digital, se sustenta en que es posible establecer el *camino recorrido* por el mismo desde la cuenta emisora hasta la cuenta receptora. Siempre se realiza el análisis forense sobre un correo electrónico *recibido,*

ya que éste contiene en su cabecera toda la información correspondiente a los sucesivos servidores de paso utilizados durante la transmisión, haciendo posible establecer la *trazabilidad* de la comunicación mediante el recorrido inverso de la transmisión (desde la recepción del correo hasta la emisión del mismo).

Durante la transmisión, el correo *viaja* pasando de servidor a servidor hasta llegar a destino, en cada uno de estos servidores el correo queda almacenado hasta que es enviado al siguiente servidor, de modo que en realidad, cada servidor contiene una *copia* del correo electrónico que transmite. En esta propuesta se denomina *ocurrencia* a cada una de esas copias. Se incluyeron en OntoFoCE los conceptos de **OCURRENCIA, HILO** y **SECUENCIA**, que permiten representar el proceso de transmisión e inferir la trazabilidad del correo electrónico. En la Figura 3-4 se ilustran las clases que representan dichos conceptos, así como las relaciones que los vinculan.

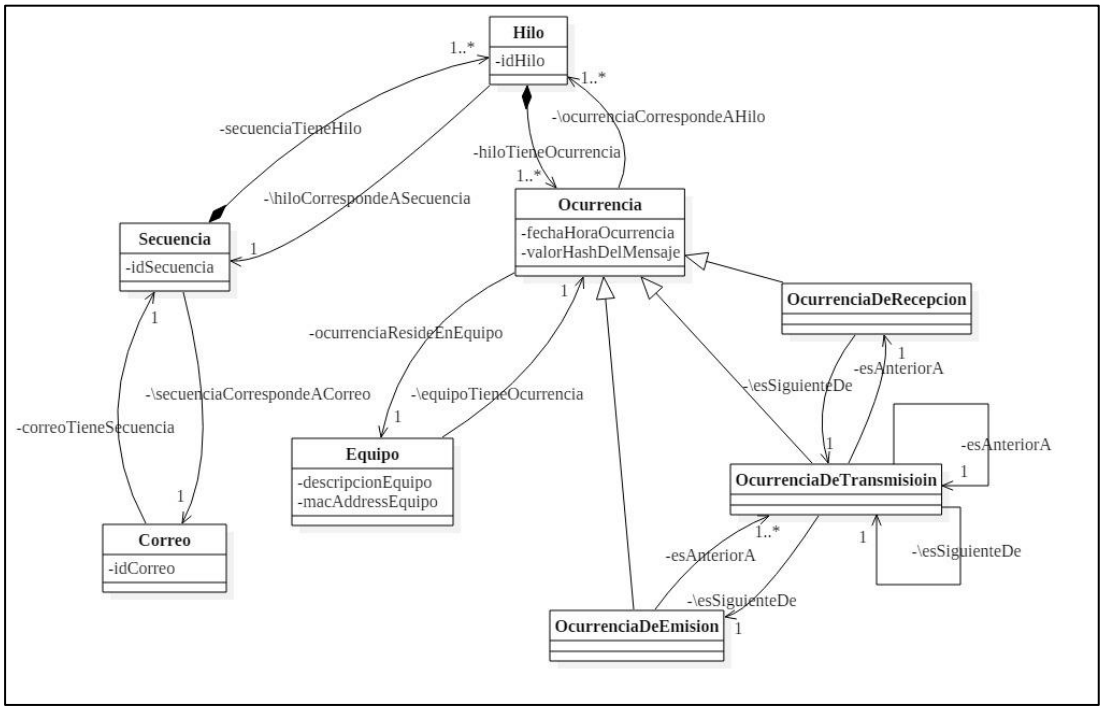


Figura 3-4: Vista Representación de las Ocurrencias

Como se observa en la figura, se identifican tres tipos de ocurrencias según sea el orden o prelación que mantienen durante la transmisión:

- Ocurrencia de Emisión: es la copia que se almacena en el equipo emisor,
- Ocurrencias de Transmisión: son las sucesivas copias del correo almacenadas en los servidores intermedios que participan de la transmisión, y
- Ocurrencia de Recepción: copia residente en el equipo receptor.

Esta tipificación se expresa en la ontología mediante las subclases ***OcurrenciaDeEmision, OcurrenciaDeTransmision*** y ***OcurrenciaDeRecepcion*** que especializan la clase ***Ocurrencia***.

Las distintas ocurrencias se agrupan mediante un **HILO,** definido como "Agrupación de ocurrencias relacionadas a una cuenta receptora". En OntoFoCE este concepto se representa según la clase de igual nombre y contiene un único atributo denominado *idHilo* que actúa como identificador de cada hilo respecto de otros que pudieran existir (habrá tantos hilos como cuentas receptoras tenga el correo).

La clase ***Hilo*** se vincula con la clase ***Ocurrencia*** mediante la relación *hiloTieneOcurrencia*, manteniendo una cardinalidad que permite que cada hilo asocie varias ocurrencias, aquellas que representan el proceso de transmisión desde la cuenta emisora a una única cuenta receptora. La relación *ocurrenciaCorrespondeAHilo* es la inversa, según una cardinalidad que permite que una ocurrencia corresponda a varios hilos, como sucede con las ocurrencias de emisión que son comunes a todos los hilos del correo.

Teniendo en cuenta que para un determinado correo existen tantos hilos como destinatarios tenga el mismo y a fin de agrupar todos los hilos que corresponden a un correo se propone el concepto de **SECUENCIA**, el cual se define como "Serie de hilos de ocurrencias de correo electrónico asociadas a un mismo correo electrónico". Este concepto agrupa todos los hilos (conjuntos de ocurrencias) que pertenecen a un correo. Por lo tanto, habrá tantas secuencias como correos electrónicos distintos. Es decir, el concepto **SECUENCIA** tiene sentido cuando en OntoFoCE se instanció más de un correo electrónico emitido y es necesario identificar los hilos que correspondan a cada correo particular. Como se observa en la Figura 3-4, este concepto se representa en OntoFoCE mediante la clase ***Secuencia*** y tiene un único atributo denominado *idSecuencia* que actúa como identificador de cada correo electrónico instanciado para la misma cuenta emisora.

La vinculación entre ***Correo*** y ***Secuencia*** se representa mediante la relación *correoTieneSecuencia,* manteniendo una cardinalidad que permite que cada correo tenga una única secuencia en la que se asocian tantos hilos como cuentas receptoras tenga dicho correo. La relación inversa *secuenciaCorrespondeACorreo* vincula las clases ***Secuencia*** y ***Correo***. La restricción de que *todo Correo tiene una única Secuencia de Ocurrencias* se expresa mediante la siguiente regla:

Correo(?x), Secuencia(?s1), correoTieneSecuencia(?x, ?s1),
Secuencia(?s2), correoTieneSecuencia(?x, ?s2)
-> SameAs (?s1, ?s2)

Esta regla expresa la unicidad de ***Secuencia*** respecto de ***Correo,*** indicando que para todo correo *?x*, si existen dos secuencias *?s1* y *?s2* que se relacionan a *?x* mediante la relación *correoTieneSecuencia*, entonces, *?s1* y *?s2* representan el mismo individuo.

Entre las clases ***Secuencia*** e ***Hilo*** se establece la relación *secuenciaTieneHilo* manteniendo una cardinalidad para representar que puede haber tantos hilos como cuentas receptoras tenga el correo. La relación inversa es *hiloCorrespondeASecuencia* y se establece entre ***Hilo*** y ***Secuencia***.

Ahora bien ¿de qué manera se establece el orden de cada ocurrencia en el hilo correspondiente? Se sabe que, en cada hilo, hay una única ***OcurrenciaDeEmision*** y una única ***OcurrenciaDeRecepcion***, que son la primera y la última, respectivamente de dicho ***Hilo***. Resta establecer el orden de las ***OcurrenciaDeTransmision*** dentro del ***Hilo***. En OntoFoCE, este orden de prelación de las ocurrencias se establece mediante la relación *esAnteriorA* que vincula una ***Ocurrencia*** con la que le precede, y mediante la relación *esSiguienteDe,* que es la relación inversa de la mencionada antes. Gráficamente, esto se expresa en la Figura 3-5:

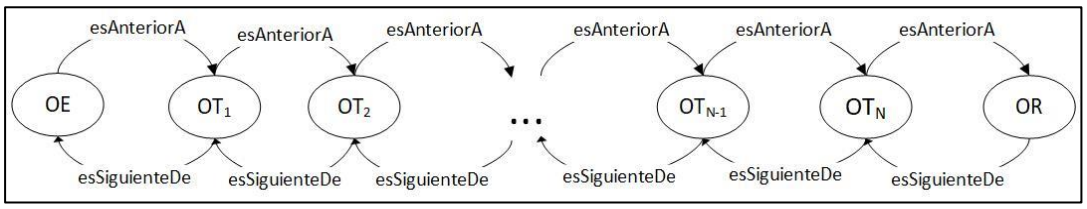


Figura 3-5: Relaciones de precedencia de las Ocurrencias de un Hilo

Ambas relaciones *esAnteriorA* y *esSiguienteDe* permiten definir las clases ***OcurrenciaDeEmision***, ***OcurrenciaDeTransmision*** *y* ***OcurrenciaDeRecepcion***, que especializan Ocurrencia. Los tres tipos de ocurrencias señaladas se definen mediante los axiomas de clases que se explican a continuación para cada tipo de ocurrencia.

La clase ***OcurrenciaDeEmision*** se define como "*Una Ocurrencia de Emisión es la primera Ocurrencia del Hilo"* y se especifica mediante el siguiente axioma de clase:

Class: OcurrenciaDeEmision
EquivalentTo: Ocurrencia and (esAnteriorA only OcurrenciaDeTransmision) and (esAnteriorA some OcurrenciaDeTransmision) and (esSiguienteDe max 0 Ocurrencia)

Con esto se quiere indicar que una instancia de ***OcurrenciaDeEmision*** es una instancia de ***Ocurrencia*** que está vinculada mediante la relación *esAnteriorA* con al

menos una ocurrencia de transmisión, si existe más de una instancia de la relación mencionada éstas la vinculan solamente con instancias de la clase ***OcurrenciaDeTransmision*** y, además, no sigue a ninguna otra ocurrencia (esSiguienteDe max 0 Ocurrencia)

Por su parte, una ***OcurrenciaDeTransmision*** se define como "*Una Ocurrencia de Transmisión es aquella que es anterior a otra Ocurrencia de Transmisión o a una Ocurrencia de Recepción o es siguiente de una Ocurrencia de Emisión"*. Esto se expresa con el siguiente axioma de clase:

Class: OcurrenciaDeTransmision
EquivalentTo: Ocurrencia and
((esAnteriorA exactly 1 OcurrenciaDeTransmision) or (esAnteriorA exactly 1 OcurrenciaDeRecepcion)) and (esAnteriorA only (OcurrenciaDeRecepcion or OcurrenciaDeTransmision)) and
((esSiguienteDe exactly 1 OcurrenciaDeTransmision) or (esSiguienteDe exactly 1 OcurrenciaDeEmision))and (esSiguienteDe only (OcurrenciaDeEmision or OcurrenciaDeTransmision))

Este axioma señala que una instancia de ***OcurrenciaDeTransmision*** es una instancia de ***Ocurrencia,*** que cumple con los siguientes criterios:

- Se asocia mediante la relación *esAnteriorA* con una única instancia de ***OcurrenciaDeTransmision*** o con una única instancia de ***OcurrenciaDeRecepcion***.
- Se vinculará mediante la relación *esAnteriorA*, solamente con instancias de las clases ***OcurrenciaDeRecepcion*** u ***OcurrenciaDeTransmision,*** es decir, no puede vincularse con instancias de otras clases mediante esta relación.
- Se relaciona mediante la relación *esSiguienteDe* con una única instancia de la clase ***OcurrenciaDeTransmision*** o con una única instancia de ***OcurrenciaDeEmision.***
- Se asocia mediante la relación *esSiguienteDe* solamente con instancias de ***OcurrenciaDeTransmision*** o de ***OcurrenciaDeRecepcion***

Finalmente, una ***OcurrenciaDeRecepcion*** se define como "*Una Ocurrencia de Recepción es la última Ocurrencia del Hilo"*. Esto se expresa mediante el siguiente axioma:

Class: OcurrenciaDeRecepcion
EquivalentTo: Ocurrencia and (esAnteriorA max 0 Ocurrencia)
and (esSiguienteDe only OcurrenciaDeTransmision)
and (esSiguienteDe exactly 1 OcurrenciaDeTransmision)

Este axioma define que una instancia de ***OcurrenciaDeRecepcion*** es una instancia de ***Ocurrencia*** que no está asociado a ninguna otra ***Ocurrencia*** por la

relación *esAnteriorA*, sólo se vincula con una única instancia de ***OcurrenciaDeTransmision***, y mediante la asociación *esSiguienteDe* sólo se puede vincular con instancias de la clase ***OcurrenciaDeTransmision***.

Por otra parte, es importante aclarar que por cada cuenta receptora de un correo, se define un hilo de ocurrencias. La unicidad de ***Hilo*** respecto de ***CuentaReceptor*** se expresa mediante la siguiente regla *Toda Cuenta que actúa como receptora tiene un único Hilo de Ocurrencia:*

CuentaReceptor(?r), Correo(?x), cuentaReceptorRecibeCorreo(?r, ?x),
Secuencia(?s), Hilo(?h1), Hilo(?h2), correoTieneSecuencia(?x, ?s),
secuenciaTieneHilo(?s, ?h1), secuenciaTieneHilo(?s, ?h2),
OcurrenciaDeRecepcion(?or),
hiloTieneOcurrencia(?h1, ?or), hiloTieneOcurrencia(?h2, ?or)
-> SameAs (?h1, ?h2)

Esta regla señala que para todo correo *?x* que es recibido por una cuenta receptor *?r* (expresando esta vinculación mediante la relación *cuentaReceptorRecibeCorreo),* se cumple que si el correo *?x* tiene una secuencia *?s* (indicada mediante la relación *correoTieneSecuencia),* y existen dos hilos identificados cómo *?h1* y *?h2*, que son hilos de esa secuencia *?s* (según lo indica la relación *secuenciaTieneHilo* establecida entre secuencia y cada hilo), y además existe una ocurrencia de recepción *?or* a la cual se vinculan los hilos *?h1* y *?h2* mediante la relación *hiloTieneOcurrencia,* entonces, *?h1* e *?h2* representan el mismo hilo.

Además, se debe considerar que la ***OcurrenciaDeEmision*** es la única ocurrencia que es compartida por todos los hilos que representan la transmisión a los distintos receptores, entonces ocurre que *Todos los Hilos de un Correo comparten la misma Ocurrencia de Emisión:*

Correo(?x), Secuencia(?s), correoTieneSecuencia(?x, ?s),
Hilo(?h1), Hilo(?h2),
secuenciaTieneHilo(?s, ?h1), secuenciaTieneHilo(?s, ?h2),
OcurrenciaDeEmision(?oe1), OcurrenciaDeEmision(?oe2),
hiloTieneOcurrencia(?h1, ?oe1), hiloTieneOcurrencia(?h2, ?oe2)
-> SameAs (?oe1, ?oe2)

Esta regla expresa que para todo correo *?c* que tiene una secuencia *?s* (expresando esta vinculación mediante la relación *correoTieneSecuencia)* y dos hilos identificados cómo *?h1* y *?h2*, que son hilos de esa secuencia *?s* (según lo indica la relación *secuenciaTieneHilo* establecida entre secuencia y cada hilo), y además existen dos ocurrencias de emisión identificadas cómo *?oe1* y *?oe2* las cuales se vinculan respectivamente con los hilos *?h1* y *?h2* mediante la relación

hiloTieneOcurrencia, entonces, ambas ocurrencias *?oe1* y *?oe2* representan la misma ocurrencia.

Ahora bien, no siempre es posible realizar la pericia sobre un correo electrónico, mayormente por dos razones: la cabecera del correo fue alterada ó el gestor del cliente de correo no responde mínimamente a los requerimientos de la RFC 822 y no es posible identificar los elementos internos de la cabecera. Por ello, el perito solo puede dar fe de su actuación cuando el correo es *factible* de ser analizado. Cabe mencionar que OntoFoCE permite instanciar el correo aun cuando esta condición de factibilidad no se cumpla, dando la posibilidad de implementar de la mejor manera posible los valores de las instancias sobre los conceptos representados en la ontología.

Ahora sí, teniendo presente la definición de todas las clases que representan el correo electrónico y su proceso de transmisión, se puede representar esta condición de factibilidad de análisis pericial, en la clase ***CorreoFactible***, mediante la regla que dice "*Un correo es factible de analizar cuando tiene Cabecera, y en ella figuran la IP del Equipo Emisor y la IP del Equipo Receptor".* Esta regla, expresada en términos de SWRL es:

Correo(?x), Secuencia(?s), correoTieneSecuencia(?x, ?s),
Hilo(?h), secuenciaTieneHilo(?s, ?h),
OcurrenciaDeEmision (?oe), hiloTieneOcurrencia(?h, ?oe),
OcurrenciaDeRecepcion (?or), hiloTieneOcurrencia(?h, ?or),
CabeceraCorreo (?cb), ocurrenciaTieneCabecera(?or, ?cb),
IP (?i1), IP (?i2), EquipoEmisor (?qe), EquipoReceptor (?qr),
equipoTieneId(?qe, ?i1), equipoTieneId(?qr, ?i2), ocurrenciaResideEnEquipo(?oe, ?qe),
ocurrenciaResideEnEquipo (?or, ?qr)
-> CorreoFactible (?x)

La regla indica que un correo *?x* es factible de analizar cuando:

- tiene una secuencia *?s* asociada mediante la relación *correoTieneSecuencia*, y existe un hilo *?h* que se relaciona con *?s* mediante *secuenciaTieneHilo,* y además existen al menos dos ocurrencias, de las cuales una es una ***OcurrenciaDeEmision*** **y la** otra es una ***OcurrenciaDeRecepcion*** identificadas cómo *?oe, y ?or,* que se asocian con *?h* mediante las respectivas relaciones *hiloTieneOcurrencia,*
- existe una ***CabeceraCorreo*** asociada a la ***OcurrenciaDeRecepcion*** mediante *ocurrenciaTieneCabecera,*
- existen dos direcciones IP identificadas cómo *?ip1 e ?ip2,* y dos equipos denominados *?qe, y ?qr* que son instancias de ***EquipoEmisor y EquipoReceptor***

respectivamente, y la relación *equipoTieneId* asocia las instancias de direcciones IP con las instancias de equipo señaladas, y finalmente, la relación *ocurrenciaResideEnEquipo* vincula cada ocurrencia con el respectivo equipo.

- Por último, las clases ***Ocurrencia*** y ***Equipo*** se enlazan a través de la relación *ocurrenciaResideEnEquipo* con una cardinalidad uno a uno, esto es así ya que, como se dijo anteriormente, al circular el correo en los diferentes servidores utilizados en la transmisión, queda una única copia del mismo (u ocurrencia) en cada servidor. La relación inversa, denominada *equipoTieneOcurrencia* se establece entre la clase ***Equipo*** y la clase ***Ocurrencia***.

3.4.3 Conceptos Complementarios

Si bien no están directamente vinculados al proceso de transmisión del correo electrónico, se considera necesario incluir un conjunto de conceptos requeridos para responder a los puntos de pericia que tratan el mensaje en sí transmitido en el correo electrónico. Así, resulta necesario identificar los conceptos **PALABRA_CLAVE y EXPEDIENTE.** La vista de la Figura 3-6 muestra las clases correspondientes, sus relaciones y atributos.

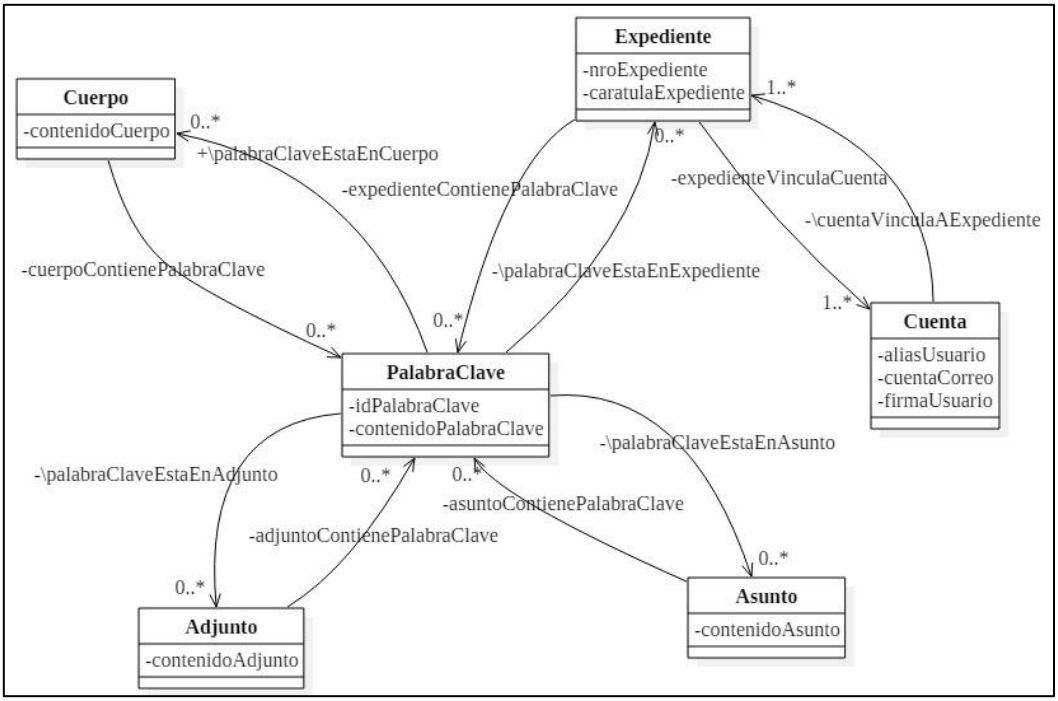


Figura 3-6: Vista de Representación de Conceptos Complementarios

A los fines de responder a los puntos de pericia P18, P19 y P20 relacionados con la búsqueda de textos, es necesario considerar el concepto **PALABRA_CLAVE,** el

cual se define como “Palabra utilizada para búsqueda de un tema de interés para la causa”. Este concepto se identifica en OntoFoCE con la clase del mismo nombre y sus atributos son:

- *idPalabraClave*, identificador único del tema de búsqueda,
- *contenidoPalabraClave,* texto específico que se utiliza para la búsqueda.

La búsqueda por palabra clave se realiza en determinados elementos del correo electrónico, ellos son: Asunto, Adjunto y Cuerpo. Con cada una de estas clases se establece la relación correspondiente:

Las clases ***PalabraClave*** y ***Asunto,*** se vinculan mediante la relación *asuntoContienePalabraClave* manteniendo una cardinalidad 0..N en el extremo final de la relación (PalabraClave), ya que es posible que una palabra clave esté en cero, uno o más asuntos. La relación inversa *palabraClaveEstaEnAsunto* tiene una cardinalidad que permite relacionar una, varias o ninguna instancia de ***Asunto*** a ***PalabraClave*** dependiendo de que el asunto se cero, una o más palabras claves.

De igual modo se relaciona la clase ***PalabraClave*** con la clase ***Cuerpo*** y con ***Adjunto,*** según las relaciones correspondientes *cuerpoContienePalabraClave* y *adjuntoContienePalabraClave*.

Por otra parte, el concepto **EXPEDIENTE** representa “el documento legal en donde consta la prueba digital de correo electrónico y los puntos de pericia” contiene el código único o número de identificación para referencia el Expediente judicial. OntoFoCE considera la clase ***Expediente*** que representa el concepto enunciado antes y contiene dos atributos:

- *nroExpediente*, identificador único del expediente, y término que se utiliza para referenciar todos los informes que elabora el perito a partir de OntoFoCE,
- *caratulaExpediente,* texto técnico que indica los datos de identificación de los involucrados en la causa y delito que se está juzgando.

El concepto **EXPEDIENTE** se relaciona primeramente con el concepto **CUENTA,** ya que allí se inicia el pedido de análisis forense del correo que se adjunta como evidencia digital (si bien el expediente incluye la prueba digital, los puntos de pericia indican una o más cuentas que figuran en esa evidencia digital y son las que deben analizarse según lo solicitado por el Juez). La vinculación se realiza mediante la relación *expedienteVinculaCuenta* entre *Expediente* y *Cuenta*. Esta relación puede vincular un expediente a una o más instancias de *Cuenta*, ya que

pueden ser varias las cuentas que figuran en el expediente. La relación inversa se denomina *cuentaVinculaAExpediente*.

Los conceptos **EXPEDIENTE** y **PALABRA_CLAVE** también se vinculan pues las palabras claves que se utilizarán para la búsqueda temática están enunciadas en el expediente. Esta relación denominada *expedienteContienePalabraClave* se establece entre la clase ***Expediente*** y la clase ***PalabraClave*** según una cardinalidad de 0-N (en el extremo final de la relación), dependiendo de si en el expediente se solicita la búsqueda de palabras claves. La relación inversa se denomina *palabraClaveEstaEnExpediente*.

Para los conceptos descriptos en la Vista de Representación de Conceptos Complementarios, se define la siguiente regla de inferencia: "*Una cuenta debe estar vinculada a un único Expediente"* y se expresa como axioma de la clase ***Cuenta***:

Class: Cuenta
SubClassOf: (cuentaVinculadaAExpediente only Expediente) and
(cuentaVinculadaAExpediente exactly 1 Expediente)

Esta regla indica que la cuenta en análisis debe estar vinculada a un único ***Expediente*** mediante la relación *cuentaVinculadaAExpediente*.

Hasta aquí la descripción del modelo ontológico según vistas parciales utilizadas para describir el modelo. A continuación, la Tabla 3-2 describe el Diccionario de Conceptos y la Figura 3-7 muestra el modelo completo de OntoFoCE, en el que –por cuestiones de legibilidad- no se representan las inversas de las relaciones entre los conceptos, no obstante, en la Tabla III-2 del ANEXO III: REPRESENTACIONES INTERMEDIAS DE LA CONCEPTUALIZACIÓN DE OntoFoCE, se pueden ver cuáles son las inversas de las asociaciones que se muestran en la figura.

Tabla 3-2: Diccionario de Conceptos

CLASE	DESCRIPCIÓN
Adjunto	Archivo asociado al correo electrónico con información complementaria al contenido del correo.
Asunto	Texto que expresa el tema del que trata el correo electrónico
CabeceraCorreo	Bloque de texto plano que contiene información relativa al correo y al proceso de transmisión realizado.
ClienteCorreo	Aplicación informática que gestiona una cuenta de correo electrónico
ClienteLocal	Cliente de Correo residente en un dispositivo (PC, Teléfono, etc) y que guarda una copia de los correos enviados/recibidos en dicho dispositivo
ClienteRemoto	Cliente de Correo al que se accede remotamente vía web
Correo	Identificación del Correo Electrónico que está siendo analizado
CorreoFactible	Correo electrónico que cumple con los requisitos mínimos para ser analizado
Cuenta	Servicio online que provee un espacio para la recepción, envío y almacenamiento de mensajes de correo electrónico.
CuentaEmisor	Cuenta de correo del usuario que emite un correo electrónico
CuentaReceptor	Cuenta de correo del usuario que recibe un correo electrónico

CLASE	DESCRIPCIÓN
Cuerpo	Contenido del mensaje del correo electrónico
Equipo	Componente de hardware que almacena el correo electrónico durante el proceso de transmisión
EquipoEmisor	Equipo utilizado para emitir un correo. Puede ser una PC, Notebook, teléfono, etc.
EquipoReceptor	Equipo utilizado para recibir el correo electrónico. Puede ser una PC, Notebook, Teléfono, etc.
Expediente	Documento legal en donde consta la prueba digital de correo electrónico y los puntos de pericia
Hilo	Agrupación de ocurrencias relacionadas a una cuenta receptora
HostName	Identificación del Equipo mediante un nombre de dominio
IdentificacionEquipo	Identificación única del hardware conectado a internet
IP	Identificación del Equipo mediante una dirección IP
Ocurrencia	Copia del correo electrónico que se almacena en cada dispositivo que participa en el proceso de transmisión
OcurrenciaDeEmision	Copia del correo electrónico residente en el equipo emisor
OcurrenciaDeRecepcion	Copia del correo electrónico residente en el equipo receptor
OcurrenciaDeTransmision	Copia del correo electrónico residente en el equipo servidor intermedio que participa en la transmisión del correo
PalabraClave	Palabra utilizada para búsqueda de un tema de interés para la causa
Secuencia	Serie de hilos de ocurrencias de correo electrónico asociadas a un mismo correo electrónico.
Servidor	Equipo utilizado para almacenar el correo electrónico, cada vez que se debe seleccionar un camino de distribución del correo durante la transmisión

A partir de estas representaciones semiformales que se implementaron en Protégé (Musen, 2015) Versión 5.5.0 se construyó el modelo lógico de OntoFoCE, utilizando OWL para la definición de los clases, propiedades y atributos de clases; SWRL para escribir las reglas y axiomas; y SPARQL para la elaboración de las consultas que dan respuesta a las preguntas de competencia.

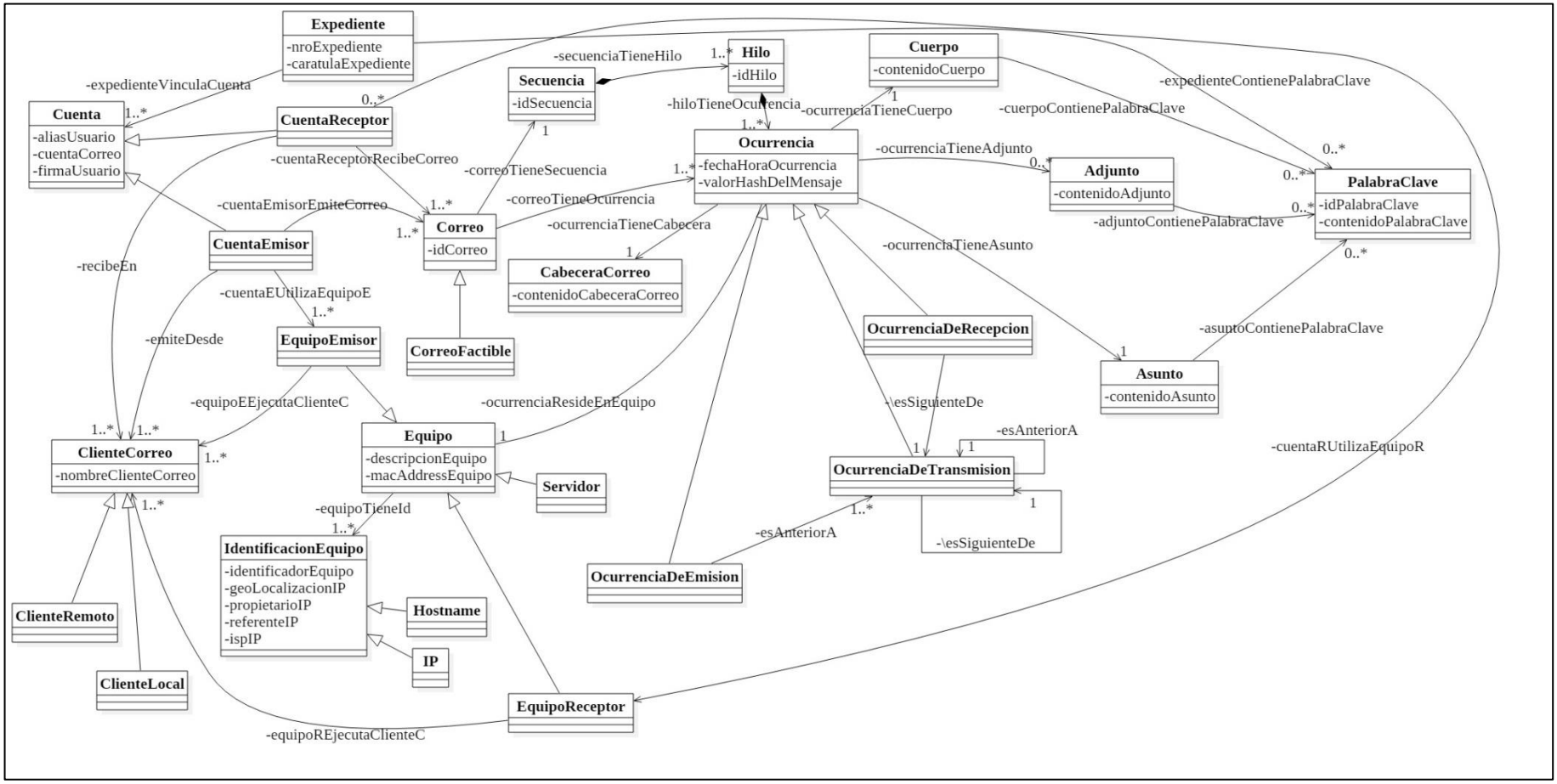


Figura 3-7: Modelo Conceptual de OntoFoCE

3.5 Aplicación de la Ontología Propuesta para el Análisis Forense de Correos Electrónicos

En esta sección se describe la aplicación de OntoFoCE a la forensia de correos electrónicos, considerando diversos escenarios. En general pueden presentarse 3 casos distintos en el análisis forense de correos:

- Cuando se pide el análisis forense de un único correo electrónico.
- Cuando se pide el análisis forense de un correo que fue enviado a más de un receptor
- Cuando se pide el análisis forense de un conjunto de correos.

La Figura 3-8 grafica los tres escenarios que luego se describen en los siguientes apartados.

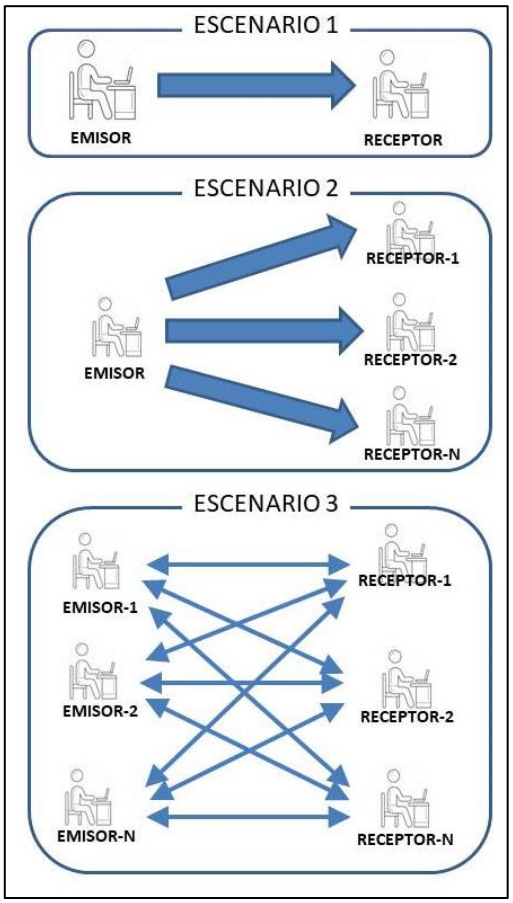


Figura 3-8: Escenarios de Pericias de Correos Electrónicos

Cabe mencionar que se abordan correos de ejemplo en el que figuran datos reales de usuarios y cuentas de correo. En esos casos, en el ANEXO IV – AUTORIZACIONES DE USO, se hacen constar la nota enviada por cada usuario en la que autoriza el uso de estos datos para la presente tesis.

3.5.1 Análisis Forense de un Correo Electrónico a un único receptor

Este caso trata sobre la pericia de un único correo, según el Escenario 1 señalado en la Figura 3-8. Para el análisis forense siempre se parte de un *correo recibido*¸ en consecuencia, se cuenta con la información necesaria en la cabecera para reconstruir el camino de transmisión desde la cuenta receptora a la cuenta emisora.

En la práctica, el perito realiza el análisis forense directamente en el dispositivo en que reside la cuenta de correo cuya cabecera se está analizando. Accediendo a ese dispositivo, obtiene la cabecera del correo recibido, así como también los restantes datos necesarios para el análisis forense: equipo receptor, equipo emisor, servidores, cliente de correo local/remoto. Todos estos datos, provenientes del análisis de la cabecera como los obtenidos de la inspección ocular al momento de la realización de la pericia, sirven para generar las instancias de las diferentes clases de la ontología y a partir de allí responder los puntos de pericia que el Juez solicite.

A continuación se ilustra con un ejemplo la forma en que los conceptos propuestos en OntoFoCE permiten representar el proceso de transmisión de un correo recibido por un usuario e inferir su trazabilidad. La instanciación que se muestra aquí se realiza de manera manual a fin de ejemplificar el uso de los conceptos.

Considere el correo que se muestra en la Figura 3-9 que ha sido recibido en la cuenta *carlos.neil@uai.edu.ar* y emitido desde *bgallo@ucasal.edu.ar*.

Colaboración en investigación
1 mensaje

Ing. H. Beatriz P. de Gallo <bgallo@ucasal.edu.ar> 8 de julio de 2018, 12:41
Para: Carlos Neil <carlos.neil@uai.edu.ar>

Estimados:

He seleccionado algunas personas de mis contactos para que reciban este correo, y me ayuden con mi investigación.
Lo único que tienen que hacer son DOS cosas:

1) Responder mi correo contestando estas preguntas:
- ¿Que dispositivo utilizó para abrir y responder el correo? (pc, celular, tablet, etc.)
- ¿Cuál es el cliente de correo que utilizó para abrir y responder el correo? (outlook, web mail de gmail o de hotmail, etc.)

2) Extraer la cabecera del correo y enviármela como archivo adjunto.

Con la información que Uds. me envíen armaré una batería de ejemplos para probar la ontología para análisis forense de mails en la que estoy trabajando.
Los molesto además para que -en caso de responder- otorguen su permiso para utilizar esos datos en esta instancia experimental de la investigación.

Muchísimas gracias por su ayuda!!

Sds
Beatriz

Figura3-9: Correo Ejemplo C1

En la Figura 3-10, se muestra el texto que corresponde a la cabecera del correo recibido (CR1) señalando los distintos datos que se extraen de la cabecera, y que luego se utilizan para instanciar OntoFoCE. Debe tenerse presente que el texto que

se muestra en la Figura 3-10 es lo que hemos denominado *archivo de texto plano*, que el Perito obtiene de la cabecera del correo recibido, y que contiene toda la información requerida para instanciar el ejemplo en OntoFoCE, con excepción de los datos complementarios (descripción y MAC Address de los equipos, datos del expediente, etc.) que el Perito obtiene por otra vía y que luego se instancian también en OntoFoCE.

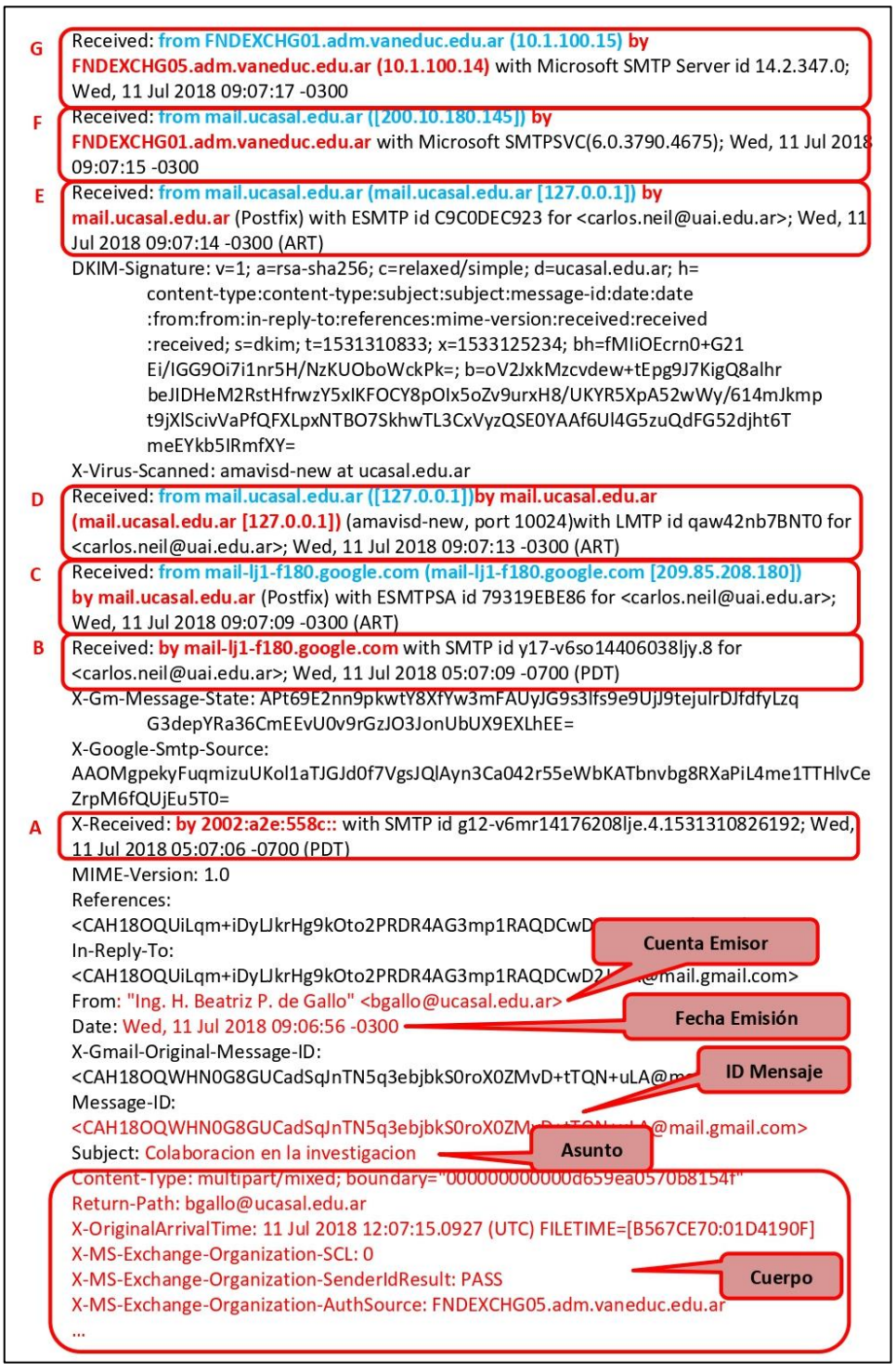


Figura 3-10: Cabecera correo CR1

Si se observa la Figura 3-10 de abajo hacia arriba, analizando línea por línea de la cabecera, se encuentran los siguientes datos:

- **CUERPO**: Identificado desde la línea que comienza con la palabra *Content-Type* hasta el final del archivo, en la Figura 3-10 se ha recortado para mostrar los datos relevantes a este estudio.
- **ASUNTO:** Se considera la línea que comienza con *Subject* y el contenido siguiente se identifica como el tema o asunto del correo.
- **MESSAGE-ID:** esta línea que comienza con *Message-ID* muestra el identificador único del correo electrónico.
- **FECHA DE EMISIÓN:** se obtiene de la línea que se inicia con *Date* y señala la fecha de emisión del correo.
- **CUENTA EMISOR:** en la línea que inicia con *from* se encontrarán los datos del alias y cuenta desde la cual se realizó el envío del correo.

La Figura 3-10 etiqueta un conjunto de líneas que comienzan con Received o X-Received. Cada una de estas líneas fue agregada por un equipo por el que pasó el correo durante el proceso de transmisión. Dicha figura identifica 7 líneas señaladas como A, B, C, D, E, F y G, a partir de las cuales se identificarán los equipos en los que residen las distintas copias del correo (ocurrencias) durante el proceso de transmisión.

La identificación de las distintas ocurrencias no es una tarea simple, debido a que no siempre se respeta la consigna definida por las normas RFC 822 y siguientes, respecto de que en cada *Received/X-Received* se encuentra en el *by* de cada paso.

La dirección IP que identifica en la cabecera cada equipo/servidor por el que pasa el correo, se obtiene de la línea de la que comienza con el parámetro *Received/X-Received.* Cada una de estas líneas, a su vez, puede contener sub-atributos, de los cuales interesan particularmente tres:

- *from:* cuyo valor indica la dirección IP/Hostname del equipo por el que había pasado el correo en el paso anterior.
- *by:* cuyo valor indica la dirección IP/Hostname del equipo actual. Entendiendo como tal al equipo que inserta la línea que se está analizando.
- *timeStamp:* cuyo valor identifica la fecha y hora en que el mensaje fue recibido por el equipo.

Los valores de los sub-atributos *from* y *by* interesan porque señalan cual es la IP/Hostname del equipo actual (*by)* y cual es la IP/Hostname del equipo anterior (*from)*. Si bien estos sub-atributos forman parte de la del parámetro *Received/X-Received*, puede ocurrir que a veces no estén presentes porque los procesos gestores de correo no respetan estrictamente las normas RFC sobre la estructuración interna del correo electrónico.

Así, cuando realiza el análisis forense, el perito identifica los equipos por los que pasa el correo y las copias que en ellos residen teniendo en cuenta las siguientes particularidades:

- Las líneas que comienzan con *Received/X-Received* indican que el correo estuvo almacenado en un equipo/servidor y el valor del sub-atributo *by* identifica a éste mediante su dirección IP/Hostname. Si el sub-atributo *by* contiene un hostname y una dirección IP entre corchetes, esta última se tomará como dirección de identificador del equipo/servidor.
- El valor IP/Hostname que se encuentra en el sub-atributo *from* de la primera línea que comienza con *Received/X-Received* identifica el equipo en la que reside la primera copia del correo (*OcurrenciaDeEmision*). Sin embargo, se observó que en los casos en que éste no estaba presente, la IP/Hostname del equipo emisor sí está presente en el sub-atributo *by* de dicha línea.
- Para el resto de las líneas, que identifican equipos en las que residen las ocurrencias de transmisión y de recepción, se toman las IP/Hostname presentes en el sub-atributo *by* y deben considerarse también la incluida en el sub-atributo *from* que debería ser igual -en su valor- al *by* de la línea *Received/X-Received* anterior. Aquí se establece una diferencia de la herramienta propuesta en esta tesis con otras herramientas existentes en el mercado, ya que de acuerdo a las pruebas realizadas, no todas consideran las direcciones IP/Hostname presentes en el sub-parámetro *from.* Esta consideración se retoma en el capítulo 5 en el que se describen los procesos de evaluación realizados para OntoFoCE y para ObE Forensics.
- El valor del sub-atributo *by* de la última línea que se agrega (que en realidad es la primera que aparece en el archivo de la cabecera) corresponde a la ocurrencia de recepción.

Hay otros detalles del análisis para identificar la dirección IP del equipo/servidor que se indican en el apartado señalado como Recolectar Información de Ocurrencias, de la sección 4.3.1.1. del capítulo 4, en donde se explica el algoritmo implementado en la aplicación ObE Forensics para crear las instancias de los conceptos de OntoFoCE que representan la información contenida en la cabecera del correo.

Para entender el proceso de identificación de las ocurrencias, permítaseme tomar de la Figura 3-10 solo las líneas del *Received/X-Received* y destacar en cada una las distintas ocurrencias que se van identificando. Obsérvese que en la Figura 3-10 se han marcado los sub-atributos *by* y *from* con color rojo y azul respectivamente, indicando además el valor que contienen para este ejemplo.

Para la primera de estas líneas identificada como A en la Figura 3-10, se resalta en la Figura 3-11, el sub-atributo *by* con el valor de una dirección IP *2002:a2e:558c::*.

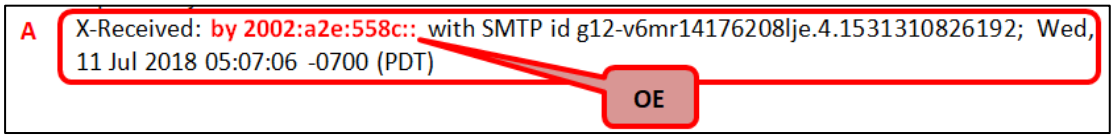


Figura 3-11: Línea A del parámetro *Received* de la cabecera CR1

Por ser esta la primera línea con el atributo *Received/X-Received*, entonces la dirección IP identificada en el *by* será la identificación del equipo que actúa como emisor, es decir, el equipo en el que reside la *OcurrenciaDeEmision (OE)*. En esta línea no se han identificado ningún otro dato que sea una dirección IP o un hostname, así que se sigue con la siguiente línea.

La línea identificada como B en la Figura 3-10, contiene únicamente el sub-atributo *by* con el valor del hostname *mail-lj1-f180.google.com*, que señala el primer servidor utilizado en el proceso de transmisión, es decir, este hostname señala al servidor que almacenará la primera ocurrencia de transmisión, identificada como OT1. Esta línea tampoco contiene el sub-atributo *from* (Ver Figura 3-12).

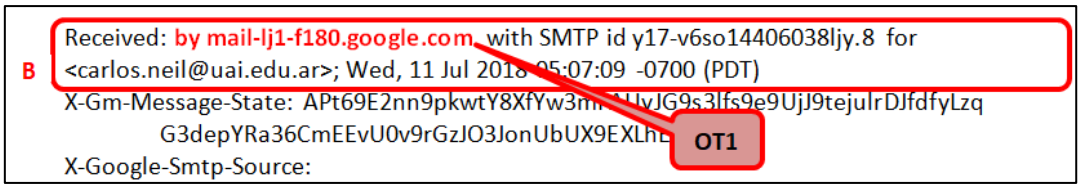


Figura 3-12: Línea B del parámetro *Received* de la cabecera CR1

Se sigue con la línea señalada en la Figura 3-10 como C, e ilustrada en la Figura 3-13. En esta última figura se observa que hay valores de direcciones IP/Hostname presentes en el sub-atributo *from*. El valor de este último sub-atributo *mail-lj1-*

f180.google.com coincide en parte con el valor del *by* de la línea anterior, en la que se identificó el servidor que almacena la ocurrencia OT1, pero también se observa que ese *from*, además del valor mencionado, enmascara una dirección IP, indicada entre paréntesis y corchete. Con lo cual puede suponerse que dicha dirección IP corresponde al equipo cuyo hostname es *mail-lj1-f180.google.com.* Pero hasta que no se obtenga la MAC Address del servidor cuyo hostname es *mail-lj1-f180.google.com* y la del servidor cuya dirección IP es *209.85.208.180*, no se puede aseverar con plena certeza que se trata del mismo equipo físico, entonces, se puede identificar que esta dirección IP enmascarada en el *from*, corresponde a un equipo que almacena una *nueva ocurrencia*, etiquetada como OT2 en la Figura 3-13, y en que toma la dirección IP *209.85.208.180* identificada en el sub-atributo from de la línea C como valor para el sub-atributo *by*.. Siguiendo con el análisis de la línea C, se observa el sub-atributo *by* que contiene el valor *mail.ucasal.edu.ar*, con lo cual se toma ese dato como identificación del servidor en el que se almacena la ocurrencia OT3.

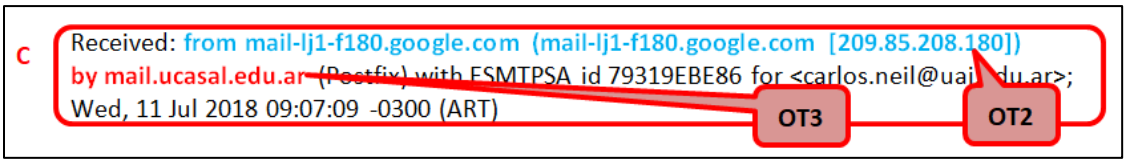


Figura 3-13: Línea C del parámetro *Received* de la cabecera CR1

La Figura 3-10 muestra la línea etiquetada como D, la cual se reproduce en la Figura 3-14. En esta última figura se observa que el sub-atributo *from* contiene el valor *mail.ucasal.edu.ar (mail.ucasal.edu.ar [127.0.0.1])*, y se tiene que el valor *mail.ucasal.edu.ar* corresponde al *by* de la línea C, pero como también enmascara una dirección IP, ésta será el identificador de equipo donde reside la ocurrencia OT4 (ver Figura 3-14). Luego, en la misma línea, el sub-atributo *by* contiene un identificador de equipo cuyo valor es *mail.ucasal.edu.ar (mail.ucasal.edu.ar [127.0.0.1])*, entonces se toma la dirección IP enmascarada entre corchetes, y esa será la que identifica el servidor en el que reside la ocurrencia OT5.

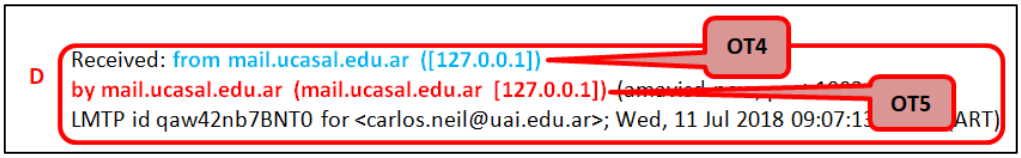


Figura 3-14: Línea D del parámetro *Received* de la cabecera CR1

Luego corresponde analizar la línea etiquetada como E en la Figura 3-10. La misma se reproduce en la Figura 3-15 donde se observa que el sub-atributo *from*

tiene un valor que es coincidente con el *by* de la línea D, pero al enmascarar una dirección IP, se toma *127.0.0.1* como identificadora del servidor que almacena una nueva ocurrencia, señalada como OT6 en esta última figura. Por su parte, el sub-atributo *by* de esta línea contiene el valor *mail.ucasal.edu.ar* el cual se tomará como identificador de equipo donde reside otra ocurrencia, denominada OT7.

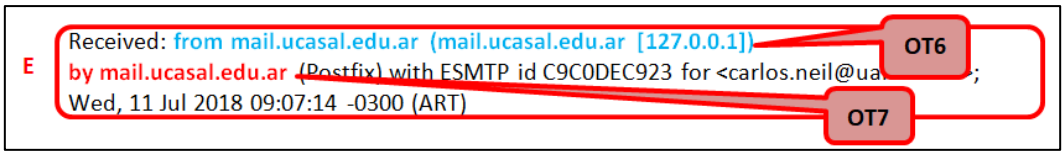


Figura 3-15: Línea E del parámetro *Received* de la cabecera CR1

En la línea identificada como F en la Figura 3-10 se observa que el sub-atributo *from* de esta línea coincide con el valor del sub-atributo *by* de la línea E pero enmascara una dirección IP (*200.10.180.1451).* Se toma dicha IP como identificador de equipo que contiene a la ocurrencia OT8. En la misma línea, el sub-atributo *by* contiene el valor *FNDEXCHG01.adm.vaneduc.edu.ar,* el cual se toma como identificador de equipo que guarda la ocurrencia OT9. Ambas ocurencias, se pueden observar en la Figura 3-16.

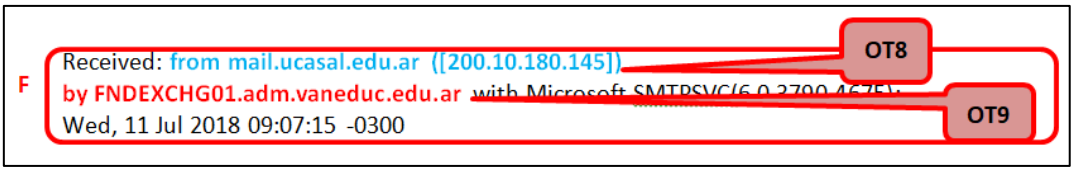


Figura 3-16: Línea F del parámetro *Received* de la cabecera CR1

Por último, se considera la línea que se etiqueta como G en la Figura 3-10. En la Figura 3-17, puede observarse que el sub-atributo *from* para dicha línea contiene un valor que coincide con el *by* de la línea E pero contiene además la dirección IP*10.1.100.15,* será éste último el identificador de equipo donde reside la ocurrencia OT10. En tanto, el sub-atributo *by* de la línea G contiene el hostname *FNDEXCHG01.adm.vaneduc.edu.ar* con la dirección IP *10.1.100.14* señalada entre corchetes. Por ser la línea etiquetada como G la última línea que comienza con el atributo *Received/X-Received,* esta IP será el identificador del último equipo por el que pasa el correo. Por lo tanto, la copia del correo que reside en dicho equipo será la última y se la identifica como ocurrencia de recepción OR.

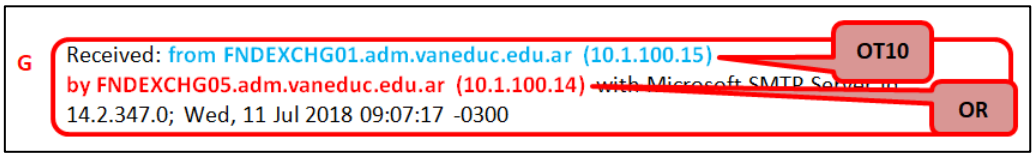


Figura 3-17: Línea G del parámetro *Received* de la cabecera CR1

En la Tabla 3-3 se resumen las ocurrencias identificadas en cada línea según el proceso descripto en los párrafos previos. Para cada línea, además de la ocurrencia identificada se muestra el identificador de equipo, así como la fecha y hora de cada una ocurrencia.

Tabla 3-3: Ocurrencias identificadas en cada Línea *Received* de la cabecera CR1

LINEA	by	FECHA Y HORA	OCURRENCIA IDENTIFICADA
A	2002:a2e:558c::	2018-07-11T12:07:06	OE
B	mail-lj1-f180.google.co	2018-07-11T12:07:09	OT1
C	209.85.208.180	2018-07-11T12:07:09	OT2
	mail.ucasal.edu.ar	2018-07-11T12:07:09	OT3
D	127.0.0.1	2018-07-11T12:07:13	OT4
	127.0.0.1	2018-07-11T12:07:13	OT5
E	127.0.0.1	2018-07-11T12:07:14	OT6
	mail.ucasal.edu.ar	2018-07-11T12:07:14	OT7
F	200.10.180.145	2018-07-11T12:07:15.	OT8
	FNDEXCHG01.adm.vaneduc.edu.ar	2018-07-11T12:07:15.	OT9
G	10.1.100.15	2018-07-11T12:07:17	OT10
	10.1.100.14	2018-07-11T12:07:17	OR

El proceso de transmisión del correo ejemplo se describe gráficamente en la Figura 3-18 donde se muestran las ocurrencias, los equipos de emisión/recepción y los servidores de paso de los correos; identificando además las direcciones IP de los mismos así como las fechas y hora de las ocurrencias.

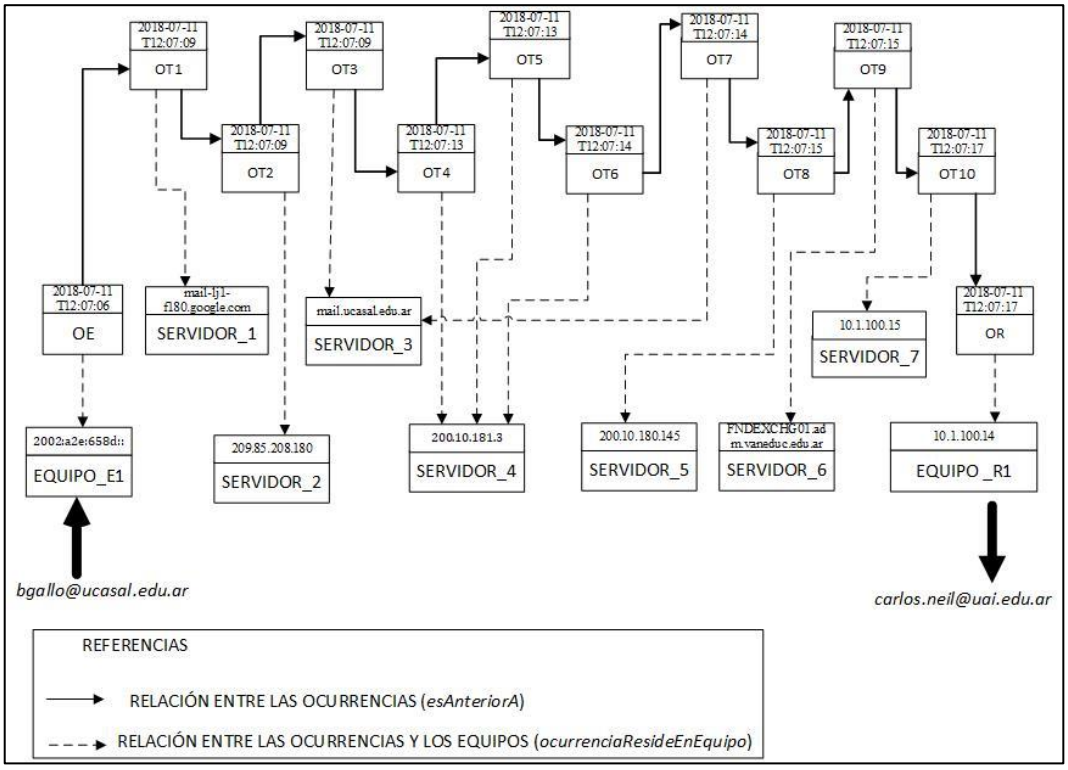


Figura 3-18: Proceso de Transmisión del correo CE1

Así, para el ejemplo de la cabecera CR1 detallada en la Figura 3-10, la Tabla 3-4 muestra la lista de *Ocurrencias* con el detalle de los valores de su atributo *fechaHoraOcurrencia*, y de los valores de los atributos *descripcionEquipo* e *identificadorEquipo* (sea éste una dirección IP o el nombre de un Hostname) para al equipo en el que la ocurrencia reside. La última columna de esta tabla indica si el identificador del equipo/servidor se encontró en el sub-atributo *by,* en cuyo caso se señala la ocurrencia como *ORIGINAL,* o si dicho identificador de equipo/servidor se encontró en el sub-atributo *from,* etiquetando la ocurrencia como *GENERADA.*

Tabla 3-4: Ocurrencias identificadas en el proceso de transmisión del correo CR1

OCURRENCIA	FECHA Y HORA	DESCRIPCIÓN EQUIPO	IDENTIFICACIÓN EQUIPO	TIPO OCURRENCIA
OE	2018-07-11T12:07:06	EQUIPO_E1	IP 2002:a2e:558c::	ORIGINAL
OT1	2018-07-11T12:07:09	SERVIDOR_1	mail-lj1-f180.google.com	ORIGINAL
OT2	2018-07-11T12:07:09	SERVIDOR_2	IP 209.85.208.180	GENERADA
OT3	2018-07-11T12:07:09	SERVIDOR_3	mail.ucasal.edu.ar	ORIGINAL
OT4	2018-07-11T12:07:13	SERVIDOR_4	IP 127.0.0.1	GENERADA
OT5	2018-07-11T12:07:13	SERVIDOR_4	IP 127.0.0.1	ORIGINAL
OT6	2018-07-11T12:07:14	SERVIDOR_4	IP 127.0.0.1	GENERADA
OT7	2018-07-11T12:07:14	SERVIDOR_3	mail.ucasal.edu.ar	ORIGINAL
O78	2018-07-11T12:07:15	SERVIDOR_5	IP 200.10.180.145	GENERADA
OT9	2018-07-11T12:07:15	SERVIDOR_6	FNDEXCHG01.adm.vaneduc.edu.ar	ORIGINAL
OT10	2018-07-11T12:07:17	SERVIDOR_7	IP 10.1.100.15	GENERADA
OR	2018-07-11T12:07:17	EQUIPO_R1	IP 10.1.100.14	ORIGINAL

Obsérvese en este ejemplo los siguientes detalles:

- Al no encontrarse en la cabecera CR1 el parámetro *Delivered-To* o *X-Original-To,* el nombre de la cuenta que actúa como receptora se debe buscar recorriendo las ocurrencias de abajo hacia arriba hasta encontrar un valor con formato de cuenta en el sub-parámetro *for* de la línea del *Received.* En el ejemplo en cuestión, esto se observa en la línea correspondiente a la OT1.
- Hay ocurrencias que son semejantes, tal como la OT4 y OT5, con lo cual podría descartase una de ellas y considerarlas *duplicadas.* Pero hasta tanto no se identifique que ambos identificadores del servidor, la dirección IP 127.0.0.1, efectivamente corresponde a un mismo equipo (mediante la obtención del identificador de MAC Address), no se puede asegurar que se trata del mismo servidor que tiene asignada la misma dirección IP en el mismo momento, y en consecuencia, no es posible aseverar con plena certeza que ambas ocurrencias son iguales.

- En las OT1, OT3, OT7 y OT9, la identificación del servidor corresponde a un nombre de dominio, con lo cual, si bien es posible identificar la dirección IP correspondiente mediante cualquier de los servicios de identificación de dominio[49], no se puede aseverar que la dirección IP que se encuentra cuando se realiza la consulta para identificar la conexión, sea la misma dirección IP que tenía asignado el equipo al momento del proceso de transmisión.
- La generación de ocurrencias identificadas a partir de los valores del sub-atributo *from*, le otorga más riqueza al análisis forense pues no se descarta ninguna dirección IP/Hostname. Esta característica refuerza la trazabilidad del proceso de transmisión, y como consecuencia, el carácter de no repudio de la evidencia digital.

Es cierto que puede ocurrir que la diferencia entre el valor del *by* de una línea con el *from* de la siguiente sea porque este primer valor es una dirección IP y el último sea un hostname, con lo cual, se estaría armando una nueva ocurrencia para el mismo servidor, y esto a lo sumo generaría redundancia. Cabe mencionar que –desde el punto de vista pericial- este proceso de *ordenamiento de ocurrencias* no altera la evidencia digital ya que no se agregan nuevas direcciones IP o hostname, más bien se genera una redundancia que no afecta el análisis pericial, por el contrario, lo fortalece.

Así, si se hiciera el análisis forense de la cabecera CR1 solo considerando el valor del sub-atributo *by* presente en cada línea, se obtendrían siete ocurrencias *originales.* Pero si se consideran los valores presentes en el sub-atributo *from* de cada línea, entonces es posible identificar otras cinco ocurrencias de transmisión *generadas* a partir del valor del sub-atributo *from.* De este modo se mantiene la vinculación entre todas las direcciones IP y los nombres de dominio, y no se pierde ningún valor correspondiente a los equipos/servidores que participaron en el proceso de transmisión.

A continuación se describe la instanciación de los datos del ejemplo en los correspondientes conceptos representados en OntoFoCE. Por cuestiones de visibilidad, se muestra la instanciación realizada en gráficos obtenidos utilizando el plugin Ontograf disponible en Protégé, el cual señaliza cada elemento de la ontología según diferentes formas y colores. Así, los rectángulos con un rombo

[49] Por ejemplo: http://ip-api.com

violeta son instancias, los que tienen un círculo naranja al inicio identifican las clases, las relaciones de color violeta vinculan cada clase con sus instancias, las relaciones de subclase se identifican con color celeste y las relaciones punteadas corresponden a las *objects properties* que se definen en la ontología. Para algunas de estas propiedades, se le agrega en la figura una etiqueta que indica de qué propiedad se trata.

Por una parte la Figura 3-19 muestra el gráfico de Ontograf para las instancias creadas a partir del ejemplo para las clases ***Correo, CuentaEmisor*** y ***CuentaReceptor,*** así como las relaciones que vinculan las clases y las correspondientes instancias *C1*, *CUENTA_E1 y CUENTA_R1*.

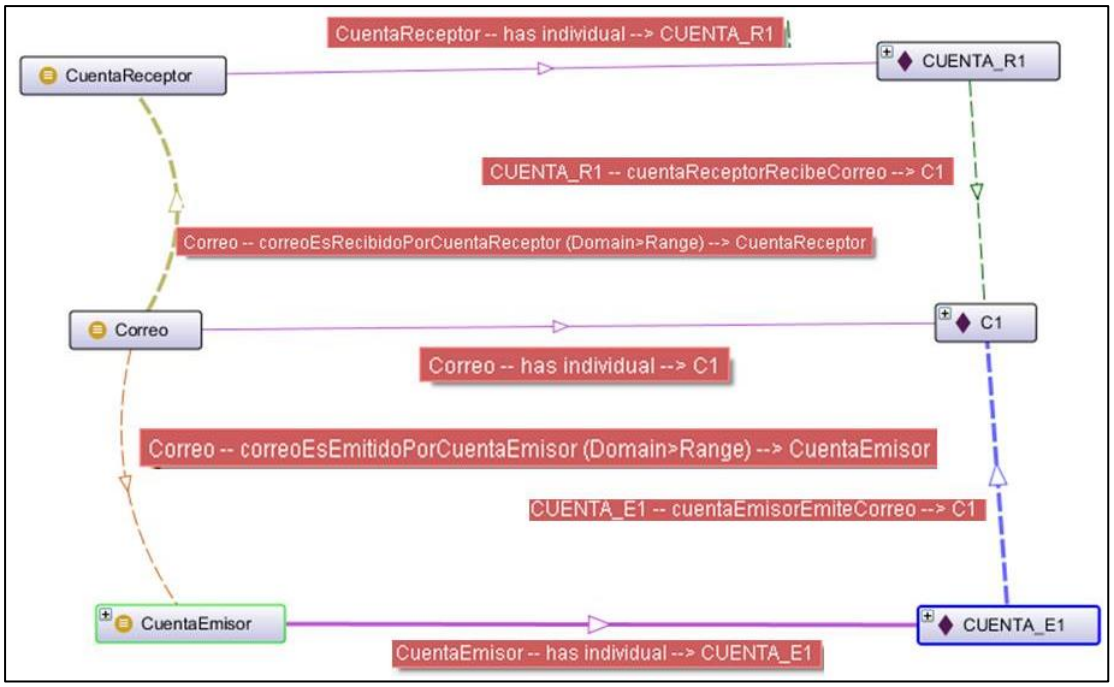


Figura 3-19: Instanciación de ***Correo, CuentaEmisor*** y ***CuentaReceptor***

Por otra parte la Figura 3-20 muestra los valores de los atributos correspondientes a lo graficado en la Figura 3-19.

En la Figura 3-19 se observa *CUENTA_E1* como la instancia que representa la cuenta emisora *bgallo@ucasal.edu.ar*, y *CUENTA_R1* que es la instancia que representa la cuenta receptora *carlos.neil@uia.edu.ar*. En la Figura 3-20 se muestran los atributos de ambas instancias con los valores correspondientes a los datos para ambas cuentas que se encuentran en la cabecera del correo. Obsérvese que en la cabecera no siempre se registran los datos de *aliasUsuario* ni *firmaUsuario*, ya que depende de que el usuario haya configurado estos datos en el gestor de cliente que está utilizando. Para el ejemplo que se está describiendo, de la cabecera se

obtuvieron los valores para instanciar en *aliasUsuario* y *cuentaCorreo,* figurando como nulo el valor del restante atributo.

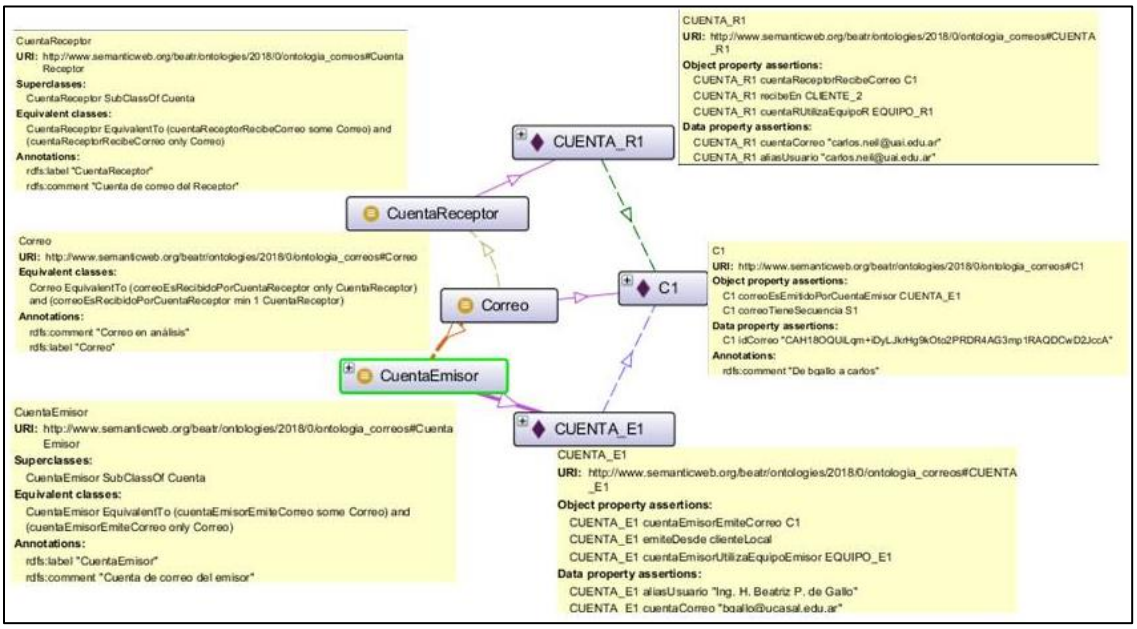


Figura 3-20: Detalle de las propiedades definidas para las instancias que se muestran en la figura 3-19

Se crea una instancia de la clase ***Correo*** llamada *C1* (ver Figura 3-19). Para esta instancia el valor que se carga en el atributo *idCorreo* es el que está señalado en la cabecera como *Message-ID: CAH18OQWH…*[50] (Ver Figura 3-20). Ambas instancias, de ***Cuenta*** y ***Correo***, se vinculan mediante la relación *correoEsEmitidoPorCuentaEmisor* expresando que el correo C1 se envió desde la cuenta CUENTA_E1.

De igual modo se puede mostrar la forma en que se ha instanciado la clase ***ClienteCorreo,*** en la Figura 3-21 se puede observar que CUENTA_E1, utiliza CLIENTE_1 para emitir el correo, y la relación *emiteDesde* representa el evento de envío realizado desde CUENTA_E1 mediante CLIENTE_1, poniendo en el valor del atributo *nombreClienteCorreo* los datos que el Perito ha relevado al momento de acceder al equipo emisor y observar que se trata de un gestor de correo que reside en el equipo, es decir es un cliente local denominado “ThunderBird”.

Asimismo, la Figura 3-21 muestra que CUENTA_R1, utiliza CLIENTE_2 para recibir el correo, y la relación *recibeEn* representa el evento de envío realizado desde CUENTA_R1 mediante CLIENTE_2.

En esta última instancia, el valor del atributo *nombreClienteCorreo* se completa con la información relevada por el Perito. Se observa que el CLIENTE_2 también es

[50] Por cuestiones de espacio, solo se incluyen los primeros caracteres del Message-ID del correo.

un gestor de correo que reside en el equipo (Cliente Local), pero es una aplicación distinta (Outlook).

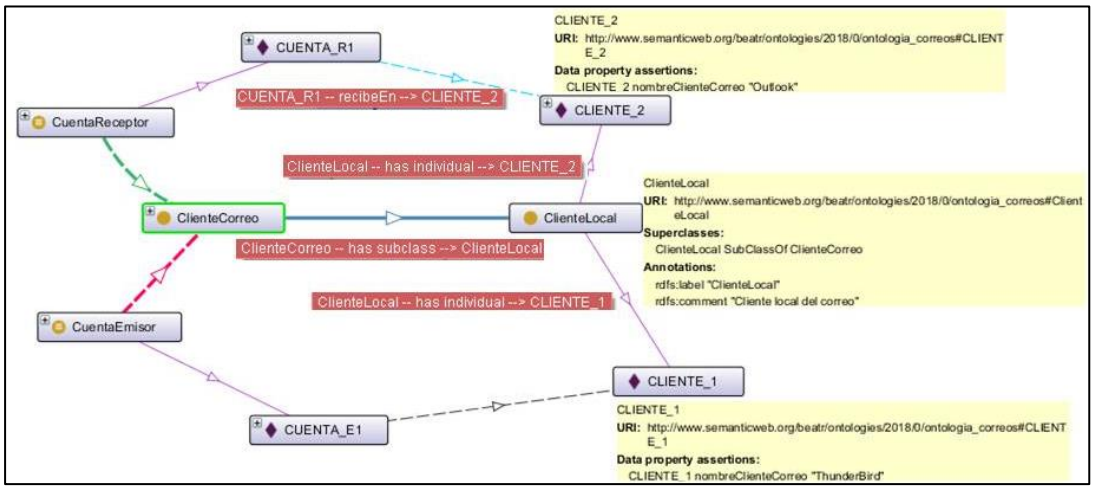


Figura 3-21: Instanciación de ***ClienteLocal***

En la Figura 3-22, se puede observar que, también a partir de datos relevados por el Perito durante la pericia, se crea el objeto EQUIPO_E1 como instancia de la clase ***EquipoEmisor***.

Para este objeto se define que "Notebook Beatriz" y "F0:E1:D2:C3:B4:A5" corresponden a los valores para los atributos *descripcionEquipo* y *macAddressEquipo.*

Una vez creada la instancia EQUIPO_E1 se lo relaciona con los objetos CLIENTE_1 y CUENTA_E1 mediante las correspondientes instancias de las relaciones *equipoEEjecutaCC*, y *cuentaEUtilizaEquipoE*.

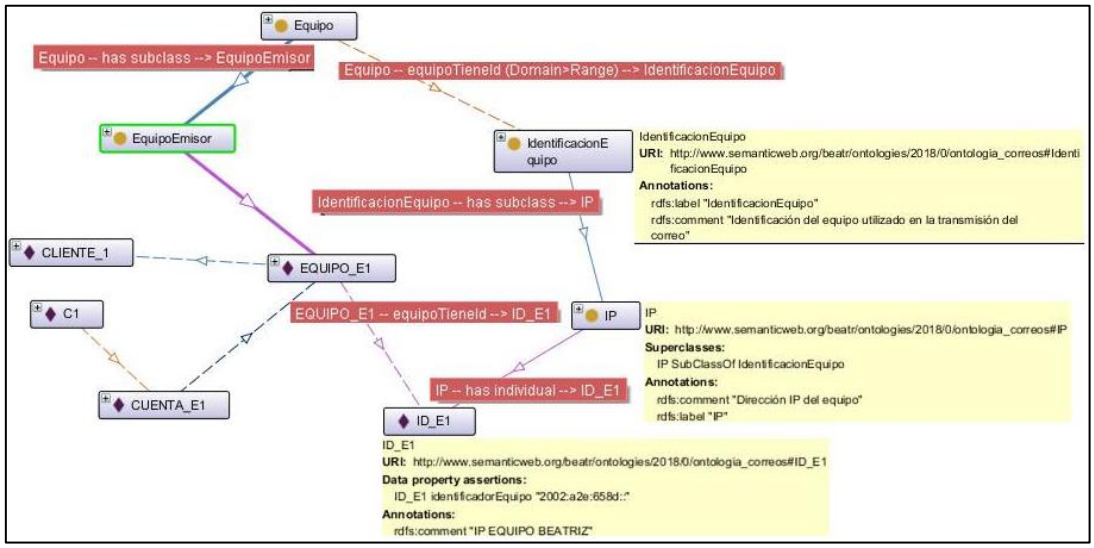


Figura 3-22: Instanciación de ***EquipoEmisor*** e ***IdentificacionEquipo***

De la cabecera se obtiene el valor del atributo *identificadorEquipo* para el objeto ID_E1 que se define como instancia de la clase *IdentificacionEquipo.*

Luego se vincula EQUIPO_E1 con ID_E mediante la relación ***equipoTieneId*** (ver Figura 3-22) y que corresponde a la dirección IP señalado en la Tabla 3-4 para el primer equipo identificado.

En este caso, hay atributos de ***IdentificacionEquipo*** que estarán nulos porque los datos referidos a la geolocalización, propietario, referente y proveedor ISP (Internet Service Provider) no se encuentran en la cabecera.

La Figura 3-23 muestra que C1, instancia de ***Correo,*** se vincula con S1, instancia de ***Secuencia,*** mediante la relación *correoTieneSecuencia.*

Obsérvese que el valor del atributo *idSecuencia* de ***Secuencia*** no se encuentra en la cabecera, sino que es asignado de manera secuencial, es decir, si fuera el caso que el correo tuviera más de una secuencia, el valor de *idSecuencia* será 1, 2, y así sucesivamente.

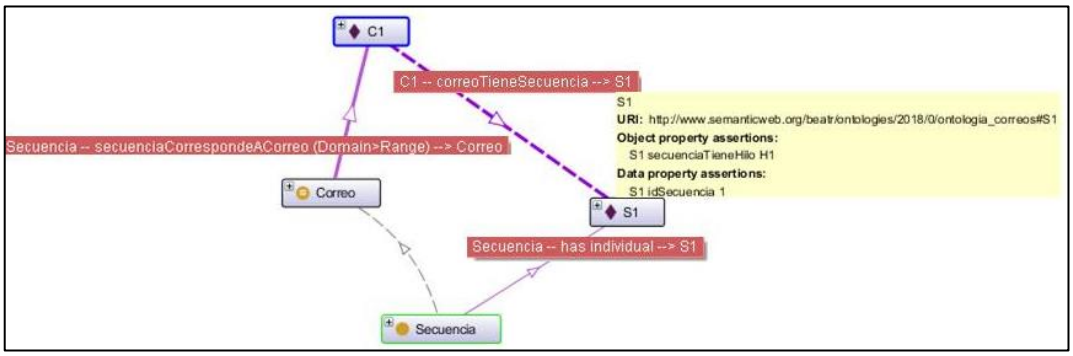


Figura 3-23: Instanciación de ***Secuencia***

Mediante la relación *secuenciaTieneHilo*, la instancia S1 se vincula con la instancia H1 correspondiente al hilo que asocia todas las ocurrencias que participan de un envío (ver Figura 3-24).

Aunque en este escenario sólo hay 1 hilo (ya que hay una sola cuenta receptora), el valor del atributo *idHilo* se maneja de manera similar al *idSecuencia* en caso de tener más hilos.

Obsérvese que en la figura citada, el objeto H1 se vincula, mediante instancias de la relación *hiloTieneOcurrencia* con las ocurrencias que se definen a partir de la cabecera del correo C1, y que se listan en la Tabla 3-4.

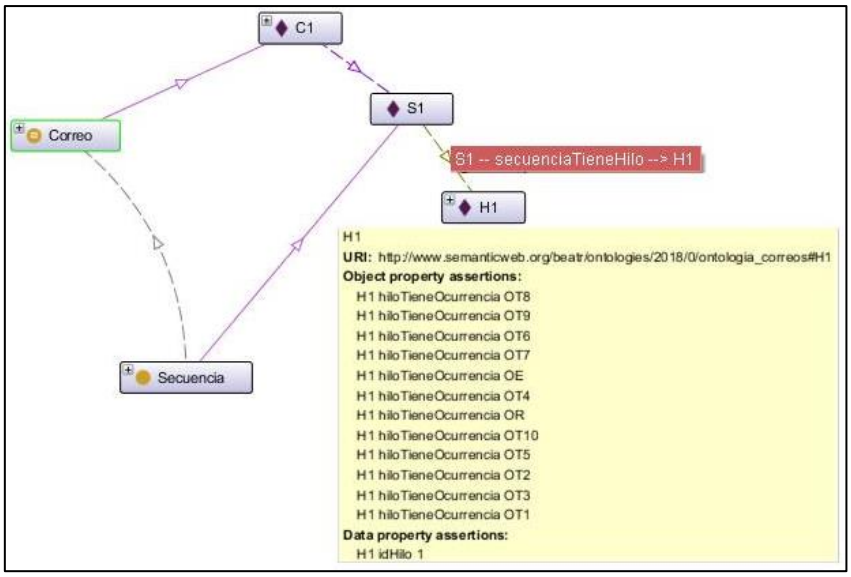


Figura 3-24: Instanciación de ***Hilo***

Para mayor claridad en la explicación, en la Figura 3-25 se ilustra una instancia de ***OcurrenciaDeEmision*** denominada OE que representa la primera copia del correo electrónico almacenada en el equipo emisor. Esta instancia tiene el valor "11 Jul 2018 12:07:06" en el atributo *fechaHoraOcurrencia*. Se observa también en la figura, que esta ocurrencia se relaciona con la ocurrencia OT1 mediante la propiedad *esAnteriorA*. Como se mencionara antes, esta relación permite establecer el camino de las copias del correo desde el emisor hasta el receptor. Las otras ocurrencias se instancian de manera similar y los valores de los atributos que corresponden a cada una se pueden ver en la Figura 3-18.

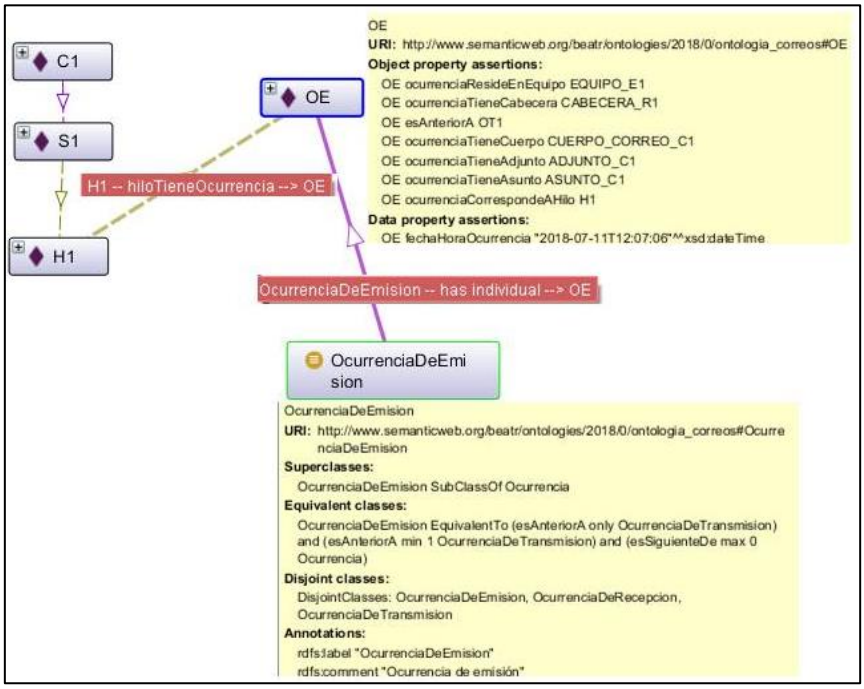


Figura 3-25: Instanciación de ***OcurrenciaDeEmision***

Por otra parte, OE se vincula con las correspondientes instancias de ***Asunto, Adjunto, Cuerpo y Cabecera*** mediante las correspondientes relaciones *ocurrenciaTieneAsunto, ocurrenciaTieneAdjunto, ocurrenciaTieneCuerpo* y *ocurrenciaTieneCabecera* (ver Figura 3-26).

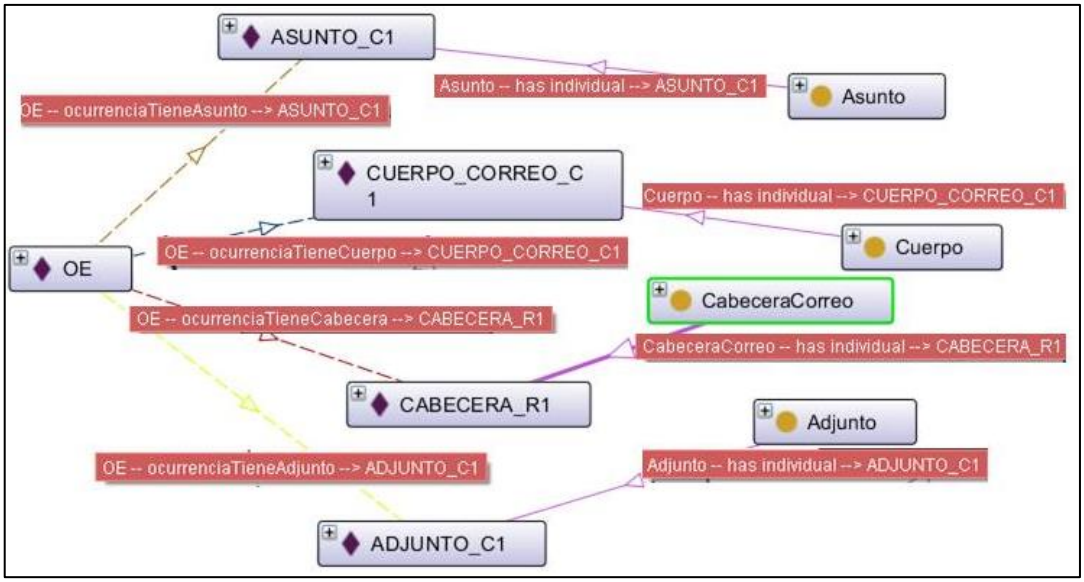


Figura 3-26: Instanciación de ***Asunto, Adjunto, Cuerpo y Cabecera***

En la Figura 3-27 se muestra la instancia ASUNTO_C1 que tiene el valor "Colaboración en…" para el atributo *contenidoAsunto,* mientras que la instancia CUERPO_CORREO_C1, asigna como valor del atributo *contenidoCuerpo* el texto completo del mensaje enviado.

En este caso, también por cuestiones de espacio, se muestra de manera reducida los primeros caracteres del mensaje real. En la misma Figura 3-27 se muestra las instancias ADJUNTO_C1 y CABECERA_C1 que tienen valores también reducidos en sus respectivos atributos *contenidoAdjunto* y *contenidoCabecera* respectivamente.

Con la instancia CABECERA_R1, corresponde hacer una aclaración. Como ya se dijo, durante el proceso de transmisión la cabecera va incrementándose con los datos del servidor por el que va pasando.

Así, si la cabecera muestra que el proceso de transmisión se realiza mediante 12 ocurrencias –como en el ejemplo en cuestión- entonces la cabecera de la ocurrencia de emisión OE se iniciará con los datos correspondientes a esa ocurrencia, es decir, la primera línea de la cabecera será la que comienza con "*X-Received: by 2002:a2e:558c…"* ya que al momento de *pasar* por el equipo emisor, la cabecera no contiene más datos que los correspondientes a esa primera ocurrencia.

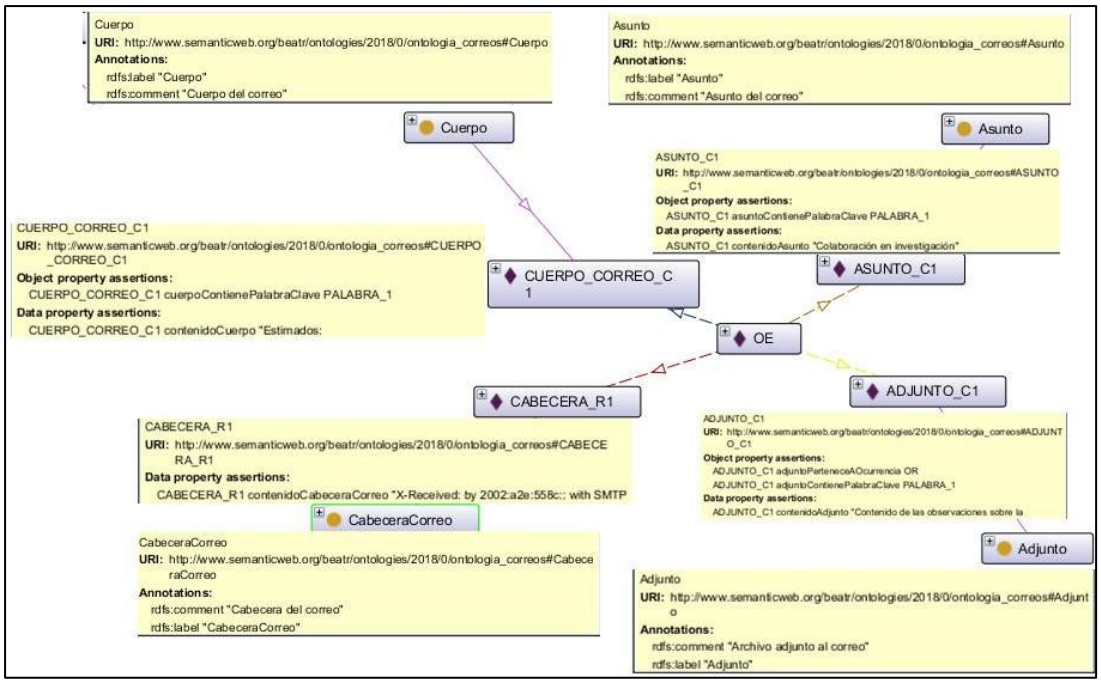


Figura 3-27: Atributos de clases e instancias ***Asunto, Adjunto, Cuerpo y Cabecera***

Por otra parte, en la Figura 3-28, se muestran las instancias de ***CuentaReceptor*** y ***Cliente*** que representan los valores identificados en la cabecera para la recepción del correo.

De forma similar a como se mostró en las figuras que grafican los valores referentes a la emisión del correo, se crea el objeto EQUIPO_R1 como instancia de la clase ***EquipoReceptor,*** que se relaciona con CLIENTE_2 y CUENTA_R1.

Obsérvese en la Figura 3-28 que se mantienen las relaciones ya señaladas en el caso de la instanciación de la cuenta de emisión descripta en los párrafos anteriores, con excepción de las correspondientes a la cuenta que actúa como receptora y al equipo receptor, es decir, aquí se aplican las relaciones *equipoREjecutaClienteC* que vincula EQUIPO_R1 con CLIENTE_2, *cuentaRUtilizaEquipoR* que vincula EQUIPO_R1 con CUENTA_R1, y *cuentaReceptorRecibeCorreo* que vincula C1 con CUENTA_R1.

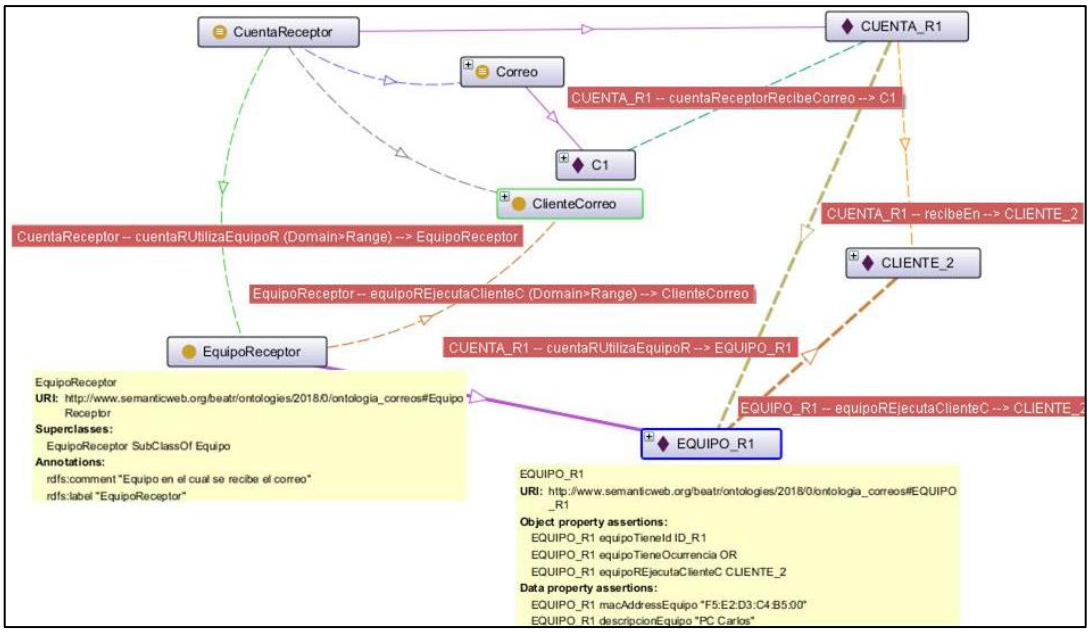


Figura 3-28: Instanciación de ***EquipoReceptor***

En la Figura 3-24 se mostró los objetos S1 y H1, instancias de ***Secuencia*** e ***Hilo*** respectivamente. Considerando ahora OR que es instancia de ***OcurrenciaDeRecepcion***, y representa la última copia del correo electrónico almacenada en el equipo receptor, la Figura 3-29 muestra esta instancia, así como el valor "11-07-2018 12:07:17" en el atributo *fechaHoraOcurrencia*, según se corresponde con lo indicado en la Figura 3-18 respecto de la última ocurrencia identificada.

Habiendo explicado el modo en que se instancian los valores correspondientes al envío/recepción del correo, se muestra en la Figura 3-29.a como las diferentes ocurrencias del hilo H1 se vinculan entre si. Considerando los valores de las ocurrencias de transmisión señaladas en la Tabla 3-4, en la Figura 3-29.a se observa que las ocurrencias se vinculan con la ocurrencia previa (que puede ser una *OcurrenciaDeEmisión* u otra *OcurrenciaDeTransmision*) a través de la relación *esAnteriorA.* Asimismo, en la Figura 3-29.b se muestra la relación *esSiguienteDe* de las distintas ocurrencias.

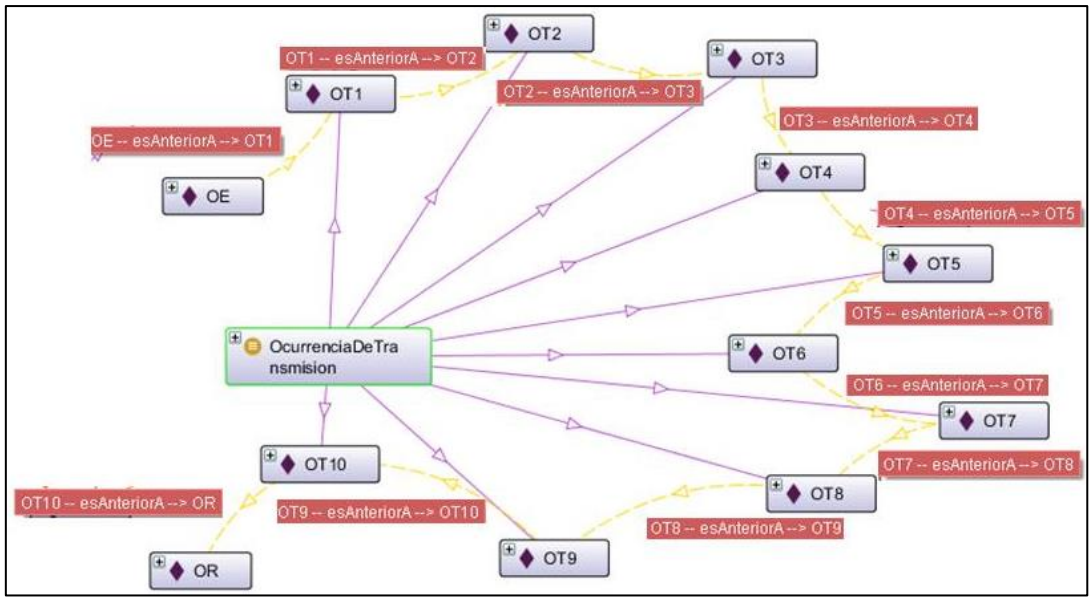


Figura 3-29.a: Instanciación de ***OcurrenciaDeTransmision*** y relación *esAnteriorA*

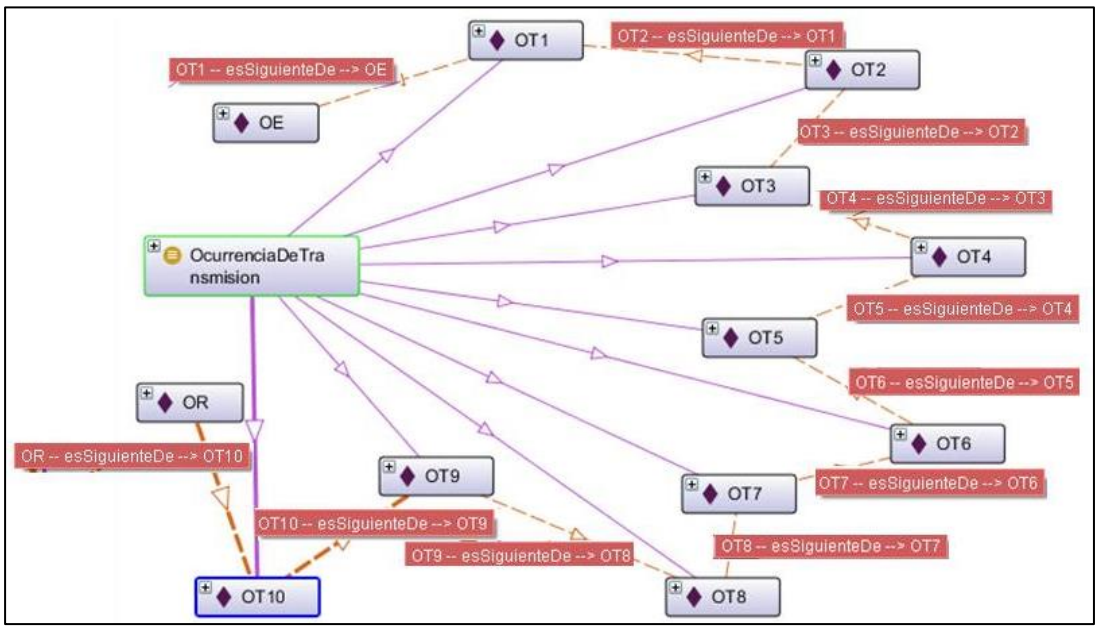


Figura 3-29.b: Instanciación de ***OcurrenciaDeTransmision*** y relación *esSiguienteDe*

Por otra parte, la Figura 3-30 muestra las instancias de la clase ***Servidor*** y la vinculación de estás con su respectiva instancia de ***IdentificacionEquipo*** a través de las relaciones *equipoTieneOcurrencia* y *equipoTieneId,* por una cuestión de espacio dicha figura solo ejemplifica éstas relaciones para el caso de SERVIDOR_4. Se puede observar en dicha figura, que en este servidor se almacenan 3 ocurrencias, a saber: OT4,OT5 y OT6.

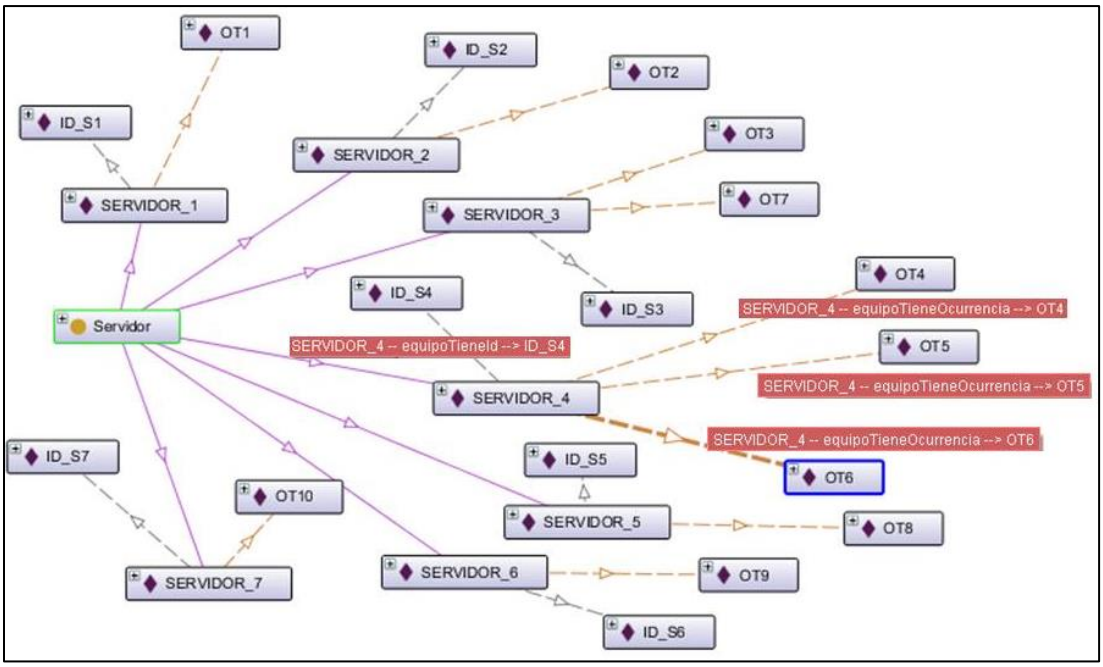


Figura 3-30: Instanciación de ***Servidor*** e ***IdentificacionEquipo***

Asimismo, cada instancia de ***OcurrenciaDeTransmision,*** se vincula con las correspondientes instancias de ***Asunto, Adjunto, Cabecera y Cuerpo*** respectivamente, tal lo indicado en la Figura 3-27.

Corresponde ahora probar las preguntas de competencia. Si bien todas pueden responderse desde cada escenario planteado en este capítulo, se seleccionan las más convenientes para cada escenario, así las preguntas PC01 a PC09, referidas al análisis forense de un correo en particular, se responden para el caso de estudio planteado en el Escenario 1, mientras que las restantes, referidos a dar respuesta cuando se procesa un conjunto de correos, se responden con los casos de estudio descriptos en el Escenario 2 y 3.

Cabe mencionar, que en el Capítulo 4, sección 4.7, se toman estos mismos escenarios como ejemplo para plantear las distintas fases del procedimiento pericial, así como la respuesta a puntos de pericia que se podrían solicitar sobre dichos ejemplos.

A continuación se describe la consulta SPARQL que formaliza cada pregunta de competencia y los resultados obtenidos según se muestran en el visor de consultas SPARQL de Protégé cuando se ejecutan. En la parte superior de cada figura se observa el código de la consulta SPARQL y, en la inferior el resultado obtenido al ejecutar la misma.

- **PC01: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de emisión?**

La consulta SPARQL que permite responder a esta pregunta de competencia es la siguiente:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?fechaEmision ?direccionIP
    WHERE { ?correo oc:correoTieneSecuencia ?s.
            ?s oc:secuenciaTieneHilo ?h.
            ?h oc:hiloTieneOcurrencia ?o.
            ?o rdf:type oc:OcurrenciaDeEmision.
            ?o oc:fechaHoraOcurrencia ?fechaEmision.
            ?o oc:ocurrenciaResideEnEquipo ?q.
            ?q oc:equipoTieneId?ip.
            ?ip oc:identificadorEquipo ?direccionIP.
           FILTER (?correo=oc:C1).}
```

SPARQL query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?fechaEmision ?direccionIP
        WHERE { ?correo oc:correoTieneSecuencia ?s.
                ?s oc:secuenciaTieneHilo ?h.
                ?h oc:hiloTieneOcurrencia ?o.
                ?o rdf:type oc:OcurrenciaDeEmision.
                ?o oc:fechaHoraOcurrencia ?fechaEmision.
                ?o oc:ocurrenciaResideEnEquipo ?q.
                ?q oc:equipoTieneId?ip.
                ?ip oc:identificadorEquipo ?direccionIP.
                FILTER (?correo=oc:C1).
}
```

fechaEmision	direccionIP
"2018-07-11T12:07:06"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"2002:a2e:658d::"

Figura 3-31: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC01

La Figura 3-31 muestra el Visor de Consultas SPARQL de Protégé con los resultados obtenidos en la pregunta de competencia PC01 para el caso del correo C1, señalando que dicho correo se envió el *11/07/18 a las 12:07:06* desde un equipo cuya dirección IP es *2002:a2e:658d::*.

- **PC02: ¿Cuál es la fecha, hora y dirección IP de recepción?**

Esta pregunta se responde a partir de la siguiente consulta SPARQL, y los resultados de esta consulta para el caso ejemplo, que se muestran en la Figura 3-32 indican que el correo se recibió el *11/07/18 a las 12:07:17* en un equipo cuya dirección IP es *10.1.100.14*.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?fechaDeRecepcion ?direccionIP
    WHERE { ?correo oc:correoTieneSecuencia ?s.
            ?s oc:secuenciaTieneHilo ?h.
            ?h oc:hiloTieneOcurrencia ?o.
            ?o rdf:type oc:OcurrenciaDeRecepcion.
```

?o oc:fechaHoraOcurrencia ?fechaDeRecepcion.
?o oc:ocurrenciaResideEnEquipo ?q.
?q oc:equipoTieneId ?ip.
?ip oc:identificadorEquipo ?direccionIP.
FILTER (?correo=oc:C1).}

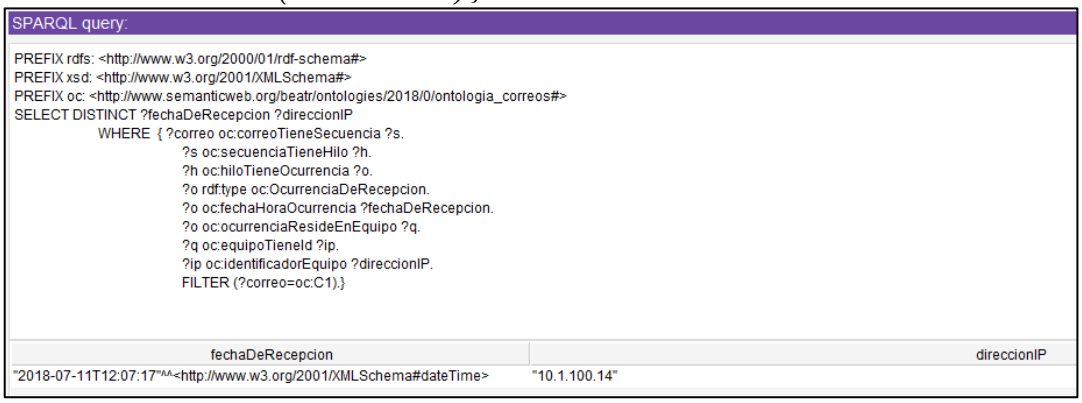


Figura 3-32: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC02

- **PC03: Dado un correo CE ¿A qué cuentas se remitió el correo?**

La consulta SPARQL correspondiente a esta pregunta es la indicada a continuación y su resultado se muestra en la Figura 3-34:

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT ?nombreUsuario ?direccionMail
WHERE { ?cuenta oc:cuentaReceptorRecibeCorreo ?correo.
?cuenta rdf:type oc:CuentaReceptor.
OPTIONAL {?cuenta oc:aliasUsuario ?nombreUsuario}.
?cuenta oc:cuentaCorreo ?direccionMail.
FILTER (?correo=oc:C1).}

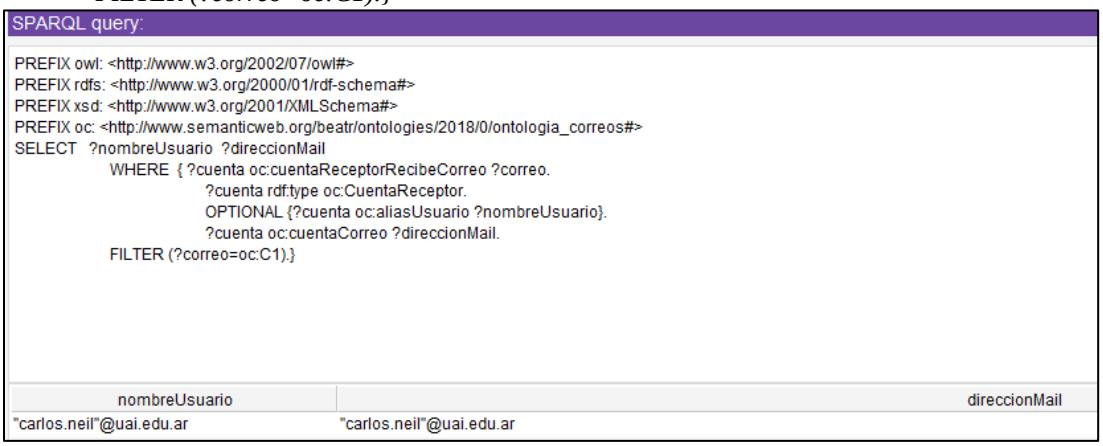


Figura 3-33: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC03

En la Figura 3-33 que muestra el visor con la consulta y los resultados obtenidos, se observa que el correo C1 fue remitido a la cuenta carlos.neil@uai.edu.ar.

- **PC04: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del emisor?**

  Para esta pregunta de competencia, la consulta SPARQL es la siguiente:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT   ?aliasUsuario  ?cuentaEmisora
    WHERE  { ?cuentaQueEmite oc:cuentaEmisorEmiteCorreo ?correo.
             ?cuentaQueEmite rdf:type oc:CuentaEmisor.
             OPTIONAL {?cuentaQueEmite oc:aliasUsuario ?aliasUsuario}.
             ?cuentaQueEmite oc:cuentaCorreo ?cuentaEmisora.
             FILTER (?correo=oc:C1).}
```

La Figura 3-34 muestra el visor de SPARQL e indica que el alias de usuario del emisor del correo C1 es *Ing. H. Beatriz P. de Gallo* y lo envió desde la cuenta *bgallo@ucasal.edu.ar*.

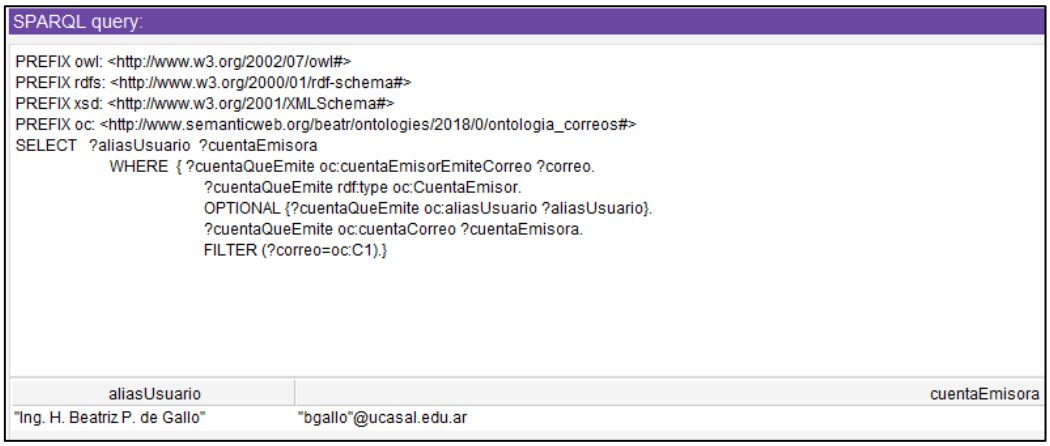


Figura 3-34: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC04

- **PC05: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del Receptor?**

  El código de la consulta SPARQL para esta pregunta de competencia se muestra a continuación:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT   ?aliasUsuario  ?cuentaReceptora
    WHERE  { ?cuentaQueRecibe oc:cuentaReceptorRecibeCorreo ?correo.
             ?cuentaQueRecibe rdf:type oc:CuentaReceptor.
             OPTIONAL {?cuentaQueRecibe oc:aliasUsuario ?aliasUsuario}.
             ?cuentaQueRecibe oc:cuentaCorreo ?cuentaReceptora.
             FILTER (?correo=oc:C1).}
```

Los resultados que se muestran en la Figura 3-35, indican que *carlos.neil@uai.edu.ar* es el alias definido para el receptor del correo C1, y que dicho correo se recibe en la cuenta *carlos.neil@uia.edu.ar*.

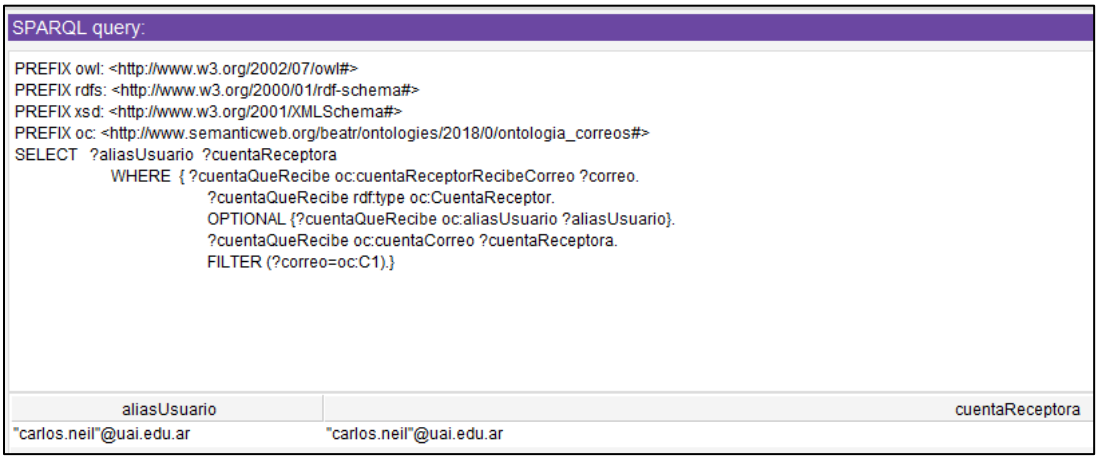


Figura 3-35: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC05

- **PC06: Dado un correo CE ¿Cuál fue el cliente de correo utilizado por cada usuario?**

A continuación se indica el código de la consulta SPARQL para esta pregunta de competencia:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?cuenta ?usuario ?cliente
    WHERE  {{ ?cuenta_receptor oc:cuentaReceptorRecibeCorreo ?correo.
              ?cuenta_receptor oc:cuentaCorreo ?cuenta.
              ?cuenta_receptor oc:recibeEn ?cli.
              ?cli oc:nombreClienteCorreo ?cliente.
              ?cuenta_receptor rdf:type oc:CuentaReceptor.
              OPTIONAL {?cuenta_receptor oc:aliasUsuario ?usuario}.}
    UNION
               { ?cuenta_emisor oc:cuentaEmisorEmiteCorreo ?correo.
              ?cuenta_emisor oc:cuentaCorreo ?cuenta.
              ?cuenta_emisor oc:emiteDesde ?cli.
              ?cli oc:nombreClienteCorreo ?cliente.
              ?cuenta_emisor rdf:type oc:CuentaEmisor.
              OPTIONAL {?cuenta_emisor oc:aliasUsuario ?usuario}.}
    FILTER (?correo=oc:C1).}
```

La pregunta se responde en la Figura 3-36, considerando el correo C1, para el cual se indica que el cliente de correo utilizado por la cuenta emisora es *ThunderBird* y el utilizado por la cuenta receptora es *Outlook*.

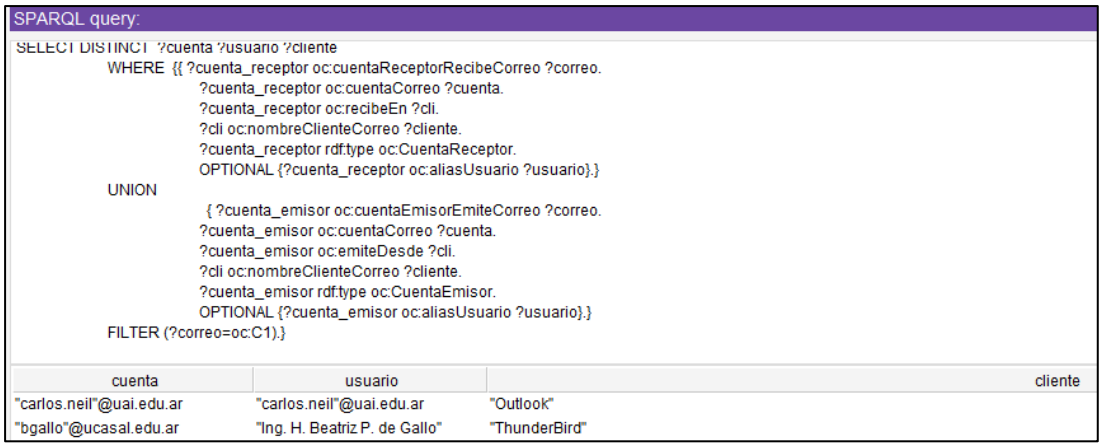


Figura 3-36: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC06

- **PC07: Dado un correo CE ¿Cuál fue el equipo desde el cual se emitió el correo?**

La consulta SPARQL que permite responder esta pregunta de competencia es la siguiente:

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?descripcionEquipo ?macAddress ?direccion
WHERE { ?correo oc:correoTieneSecuencia ?s.
?s oc:secuenciaTieneHilo ?h.
?h oc:hiloTieneOcurrencia ?o.
?o rdf:type oc:OcurrenciaDeEmision.
?o oc:ocurrenciaResideEnEquipo ?equipo.
?equipo oc:equipoTieneId ?IP.
?IP oc:identificadorEquipo ?direccion.
OPTIONAL {?equipo oc:macAddressEquipo ?macAddress}.
?equipo oc:descripcionEquipo ?descripcionEquipo.
FILTER (?correo=oc:C1).}

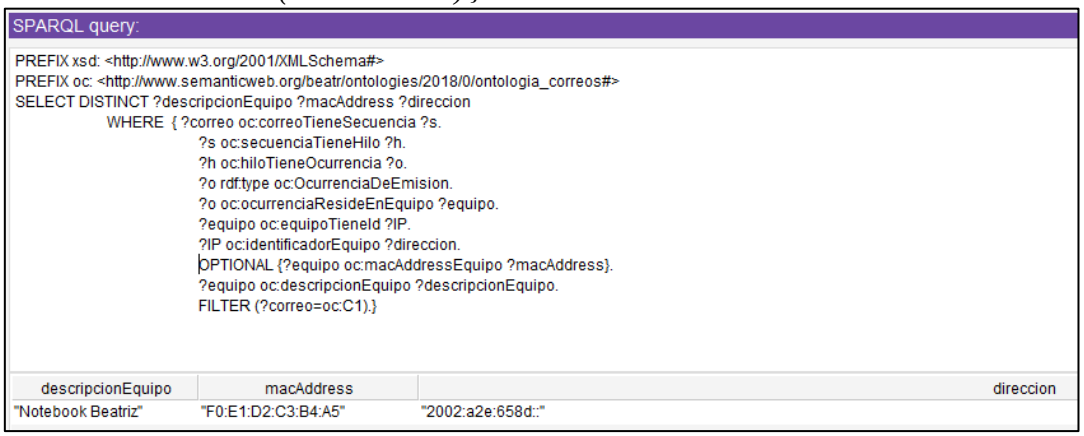


Figura 3-37: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC07

La Figura 3-37 señala que el correo C1 fue enviado desde el equipo identificado con el nombre de *Notebook Beatriz* identificado con la dirección IP 2002:a2e:658d:: y bajo la Mac Address *F0:E1:D2:C3:B4:A5*.

- **PC08: Dado un correo CE ¿Cuál fue el equipo en el que se recibió el correo?**

  La siguiente consulta SPARQL permite responder a la pregunta PC08.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?descripcionEquipo ?macAddress ?direccion
    WHERE  { ?correo oc:correoTieneSecuencia ?s.
             ?s oc:secuenciaTieneHilo ?h.
             ?h oc:hiloTieneOcurrencia ?o.
             ?o rdf:type oc:OcurrenciaDeRecepcion.
             ?o oc:ocurrenciaResideEnEquipo ?equipo.
             ?equipo oc:macAddressEquipo ?macAddress.
             ?equipo oc:equipoTieneId ?IP.
             ?IP oc:identificadorEquipo ?direccion.
             ?equipo oc:descripcionEquipo ?descripcionEquipo.
             FILTER (?correo=oc:C1).}
```

En el visor de la consulta SPARQL que se muestra en la Figura 3-38 de la pregunta de competencia PC08, se observa que el correo C1 ha sido recibido en el equipo denominado *PC Carlos* identificado con la Mac Address *F5:E2:D3:C4:B5:00* cuya dirección IP es *10.1.100.14*.

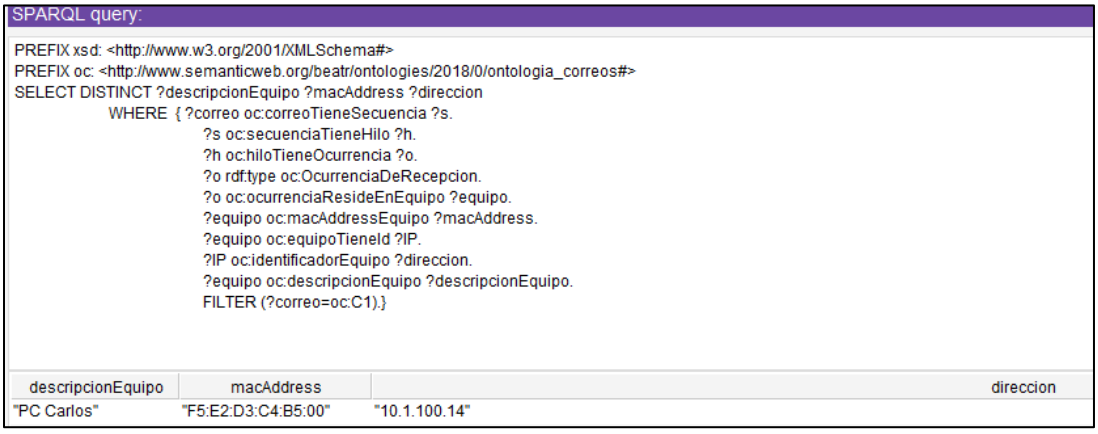


Figura 3-38: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC08

- **PC09: Dado un correo, un emisor y un receptor ¿cuál es la secuencia de equipos por los que pasó ese correo?**

  La consulta SPARQL que permite responder a la pregunta de competencia PC09 es la siguiente:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT   ?ocurrencias ?fecha  ?direccionIP  ?descripcionEquipo
WHERE {     ?emisor oc:cuentaEmisorEmiteCorreo ?correo.
    ?emisor oc:cuentaEUtilizaEquipoE ?equipoE.
    ?receptor oc:cuentaReceptorRecibeCorreo ?correo.
    ?receptor oc:cuentaRUtilizaEquipoR ?equipo.
    ?o oc:ocurrenciaResideEnEquipo ?equipo.
    ?correo oc:correoTieneSecuencia ?s1.
    ?s1 oc:secuenciaTieneHilo ?hilo1.
```

?hilo1 oc:hiloTieneOcurrencia ?ocurrencias.
?ocurrencias oc:ocurrenciaResideEnEquipo ?equipo2.
?equipo2 oc:equipoTieneId ?IP.
?equipo2 oc:descripcionEquipo ?descripcionEquipo.
?IP oc:identificadorEquipo ?direccionIP.
?ocurrencias oc:fechaHoraOcurrencia ?fecha.
FILTER (?receptor=oc:CUENTA_R1 && ?emisor=oc:CUENTA_E1 && ?correo=oc:C1).
} order by ?fecha

El visor de consulta SPARQL para esta pregunta se muestra en la Figura 3-39, en la que se muestran todas las ocurrencias del correo C11, indicando para cada una la fecha y hora del paso del correo por cada equipo/servidor y la dirección IP correspondiente.

SPARQL query:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?ocurrencias ?fecha  ?direccionIP  ?descripcionEquipo
WHERE {    ?emisor oc:cuentaEmisorEmiteCorreo ?correo.
           ?emisor oc:cuentaEUtilizaEquipoE ?equipoE.
           ?receptor oc:cuentaReceptorRecibeCorreo ?correo.
           ?receptor oc:cuentaRUtilizaEquipoR ?equipo.
           ?o oc:ocurrenciaResideEnEquipo ?equipo.
           ?correo oc:correoTieneSecuencia ?s1.
           ?s1 oc:secuenciaTieneHilo ?hilo1.
           ?hilo1 oc:hiloTieneOcurrencia ?ocurrencias.
           ?ocurrencias oc:ocurrenciaResideEnEquipo ?equipo2.
           ?equipo2 oc:equipoTieneId ?IP.
           ?equipo2 oc:descripcionEquipo ?descripcionEquipo.
           ?IP oc:identificadorEquipo ?direccionIP.
```

ocurrencias	fecha	direccionIP	descripcionEquipo
OE	"2018-07-11T12:07:06"^^<http://www.w3.org/2001/XMLSchema#dateTi	"2002:a2e:658d::"	"Notebook Beatriz"
OT1	"2018-07-11T12:07:09"^^<http://www.w3.org/2001/XMLSchema#dateTi	"mail-lj1-f180.google.com"	"servidor 1"
OT3	"2018-07-11T12:07:09"^^<http://www.w3.org/2001/XMLSchema#dateTi	"mail.ucasal.edu.ar"	"servidor 3"
OT2	"2018-07-11T12:07:09"^^<http://www.w3.org/2001/XMLSchema#dateTi	"209.85.208.180"	"servidor 2"
OT5	"2018-07-11T12:07:13"^^<http://www.w3.org/2001/XMLSchema#dateTi	"127.0.0.1"	"servidor 4"
OT4	"2018-07-11T12:07:13"^^<http://www.w3.org/2001/XMLSchema#dateTi	"127.0.0.1"	"servidor 4"
OT7	"2018-07-11T12:07:14"^^<http://www.w3.org/2001/XMLSchema#dateTi	"mail.ucasal.edu.ar"	"servidor 3"
OT6	"2018-07-11T12:07:14"^^<http://www.w3.org/2001/XMLSchema#dateTi	"127.0.0.1"	"servidor 4"
OT9	"2018-07-11T12:07:15"^^<http://www.w3.org/2001/XMLSchema#dateTi	"FNDEXCHG01.adm.vaneduc.edu.ar"	"servidor6"
OT8	"2018-07-11T12:07:15"^^<http://www.w3.org/2001/XMLSchema#dateTi	"200.10.180.145"	"servidor 5"
OR	"2018-07-11T12:07:17"^^<http://www.w3.org/2001/XMLSchema#dateTi	"10.1.100.14"	"PC Carlos"
OT10	"2018-07-11T12:07:17"^^<http://www.w3.org/2001/XMLSchema#dateTi	"10.1.100.15"	"servidor7"

Figura 3-39: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC09

El resto de las preguntas de competencia, identificadas como PC10 a PC21, se ejemplifican con datos del Escenario 2, que se detalla a continuación.

3.5.2 Análisis Forense de un Correo Electrónico enviado a Varios Receptores

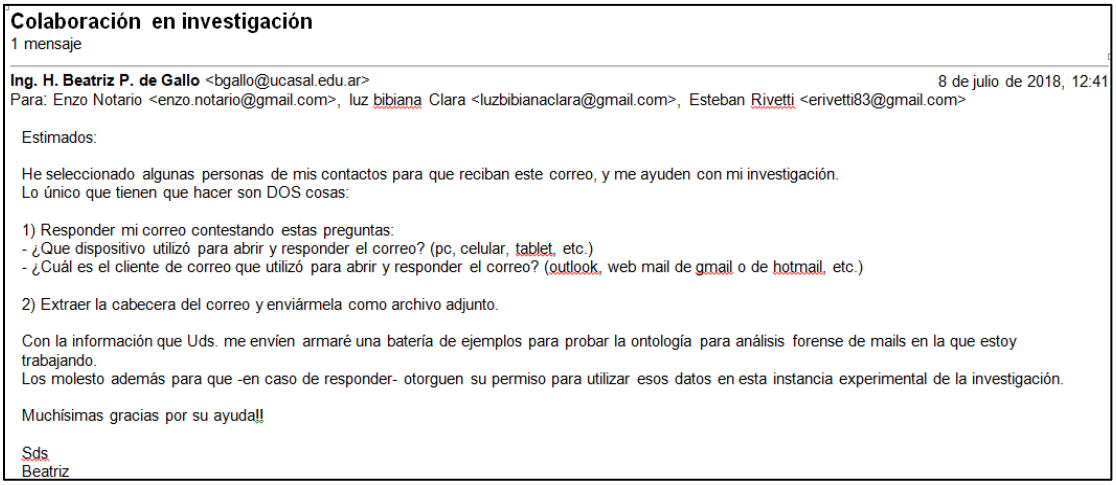
En este caso, el perito cuenta con más de un correo recibido, y todos ellos fueron emitidos por el mismo emisor. Usualmente es el caso de un correo que se emite a varias cuentas bajo la opción “con copia” o por “lista de distribución”.

También se parte de los correos recibidos. Es decir, el perito tiene acceso a los correos recibidos de aquellas cuentas que participan en la copia o en la lista de distribución.

En la práctica, esto significa que el perito tiene acceso a distintos dispositivos (tantos como correos recibidos tenga que analizar), y para cada dispositivo busca la

cabecera del correo y toda la información complementaria necesaria (equipo receptor, cliente de correo remoto/local, etc.).

Supongamos que el correo planteado como ejemplo del Escenario 1, se envía a 3 cuentas diferentes. En este caso, el correo sería el que muestra la Figura 3-40.



Colaboración en investigación
1 mensaje

Ing. H. Beatriz P. de Gallo <bgallo@ucasal.edu.ar> 8 de julio de 2018, 12:41
Para: Enzo Notario <enzo.notario@gmail.com>, luz bibiana Clara <luzbibianaclara@gmail.com>, Esteban Rivetti <erivetti83@gmail.com>

Estimados:

He seleccionado algunas personas de mis contactos para que reciban este correo, y me ayuden con mi investigación.
Lo único que tienen que hacer son DOS cosas:

1) Responder mi correo contestando estas preguntas:
- ¿Que dispositivo utilizó para abrir y responder el correo? (pc, celular, tablet, etc.)
- ¿Cuál es el cliente de correo que utilizó para abrir y responder el correo? (outlook, web mail de gmail o de hotmail, etc.)

2) Extraer la cabecera del correo y enviármela como archivo adjunto.

Con la información que Uds. me envíen armaré una batería de ejemplos para probar la ontología para análisis forense de mails en la que estoy trabajando.
Los molesto además para que -en caso de responder- otorguen su permiso para utilizar esos datos en esta instancia experimental de la investigación.

Muchísimas gracias por su ayuda!!

Sds
Beatriz

Figura 3-40: Correo ejemplo C2

Este correo fue enviado desde la cuenta bgallo@ucasal.edu.ar a tres usuarios, identificados en las siguientes cuentas receptoras: enzo.notario@gmail.com, luzbibianaclara@gmail.com y erivetti83@gmail.com.

Supuesto que el Perito realiza la extracción de las cabeceras correspondientes en las 3 cuentas receptoras siguiendo el procedimiento indicado para realizar la pericia, en las Figuras 3-41, 3-42 y 3-43 se muestran dichas cabeceras. En cada figura se han señalado los datos de interés, que luego permitirán generar las instancias de clases en OntoFoCE.

También se debe aclarar que, con el objetivo de destacar solo aquellos datos de la cabecera que interesan para el análisis pericial, en las Figuras 3-41, 3-42 y 3-43, se han recortado de la imagen, algunos tramos de la cabecera referente a parámetros y valores no pertinentes.

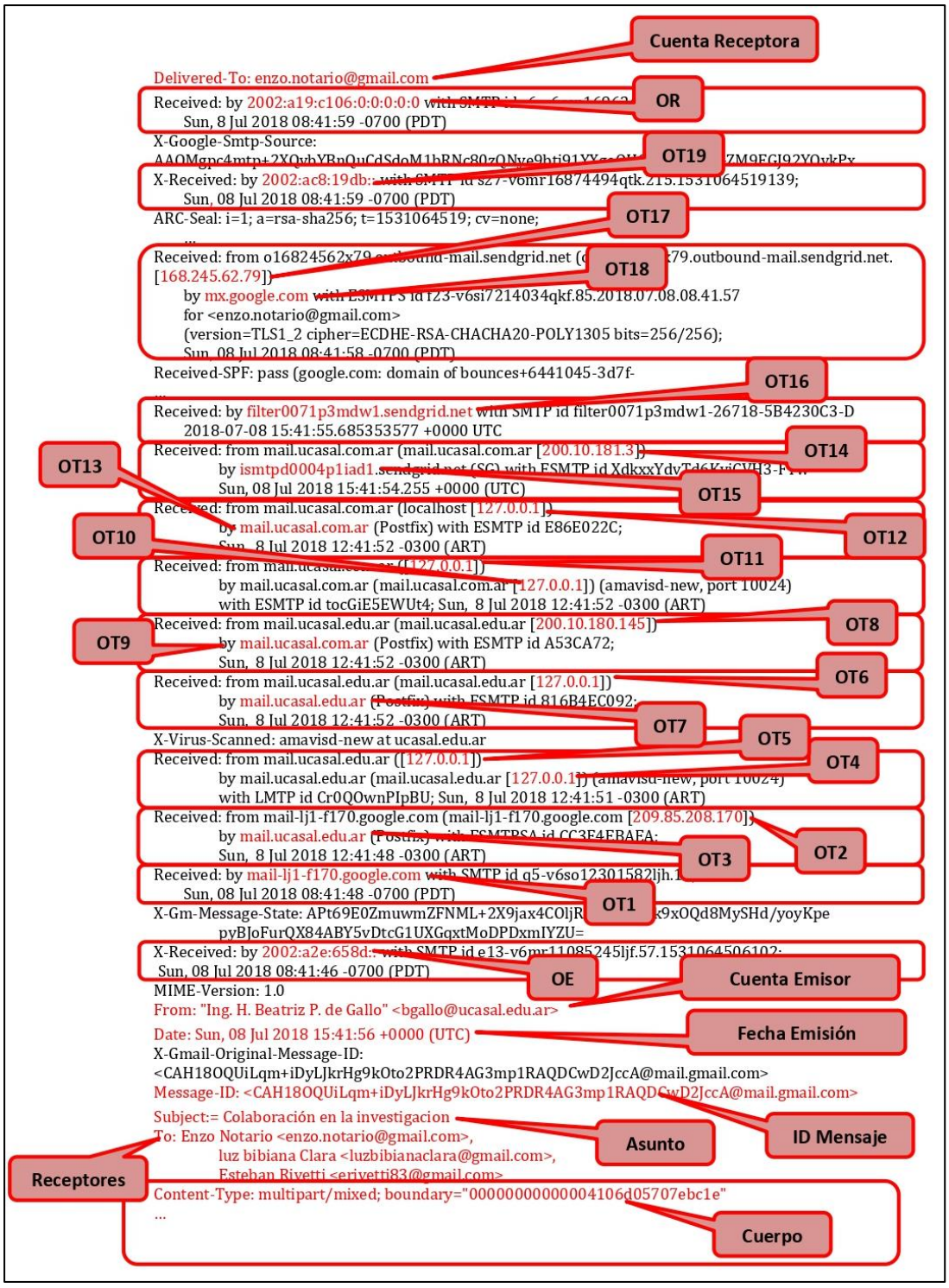


Figura 3-41: Cabecera CABECERA_R1 recibida en la cuenta enzo.notario@gmail.com

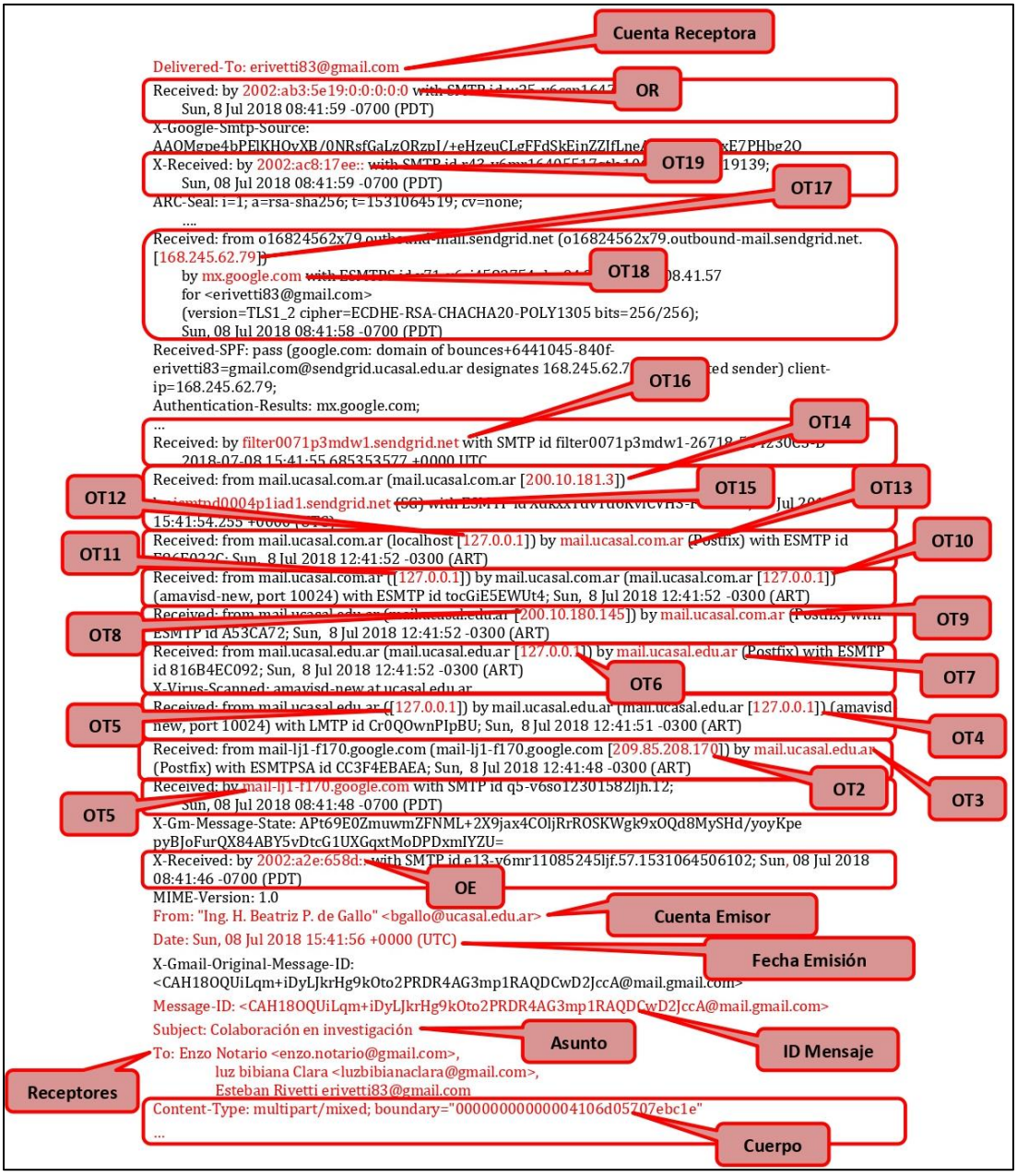


Figura 3-42: Cabecera CABECERA_R2 recibida en la cuenta erivetti83@gmail.com

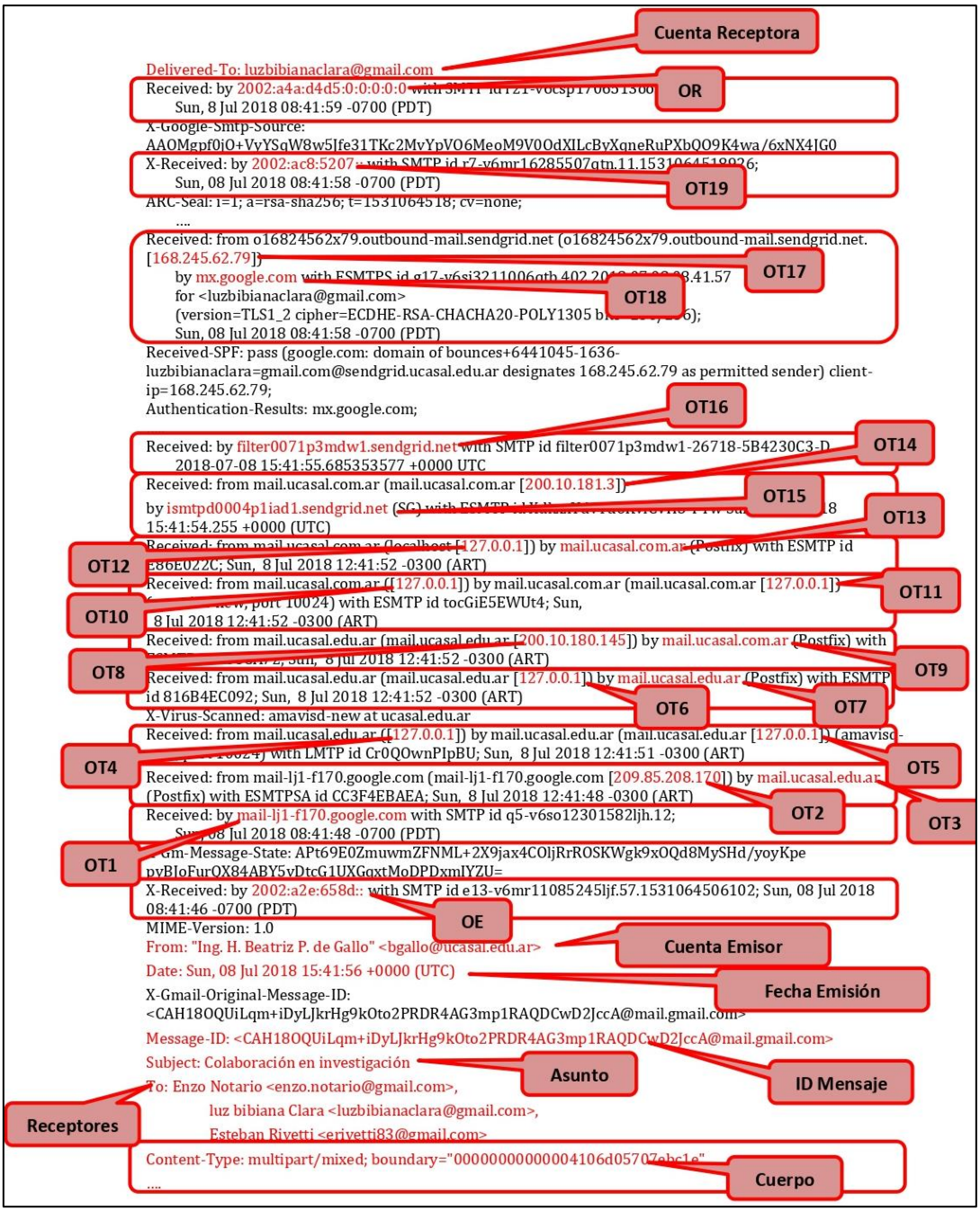


Figura 3-43: Cabecera CABECERA_R3 recibida en la cuenta luzbibianaclara@gmail.com

Se toma entonces una a una esas cabeceras, y se crean las instancias de clases siguiendo el proceso señalado en la sección 3.5.1 de este mismo capítulo. Aquí también se tiene en cuenta el criterio de no descartar ninguna dirección IP/Hostname que figura en la cabecera en el sub-atributo *from*, siguiendo el mismo procedimiento que el explicado para el caso del Escenario 1. Estas tres cabeceras, al corresponder al mismo envío, tienen tres hilos, y a manera de ejemplo, en la Figura 3-44 se muestra el hilo identificado como H3 con todas las ocurrencias que lo componen, los equipos/servidores y las identificaciones correspondientes a la CABECERA_R3.

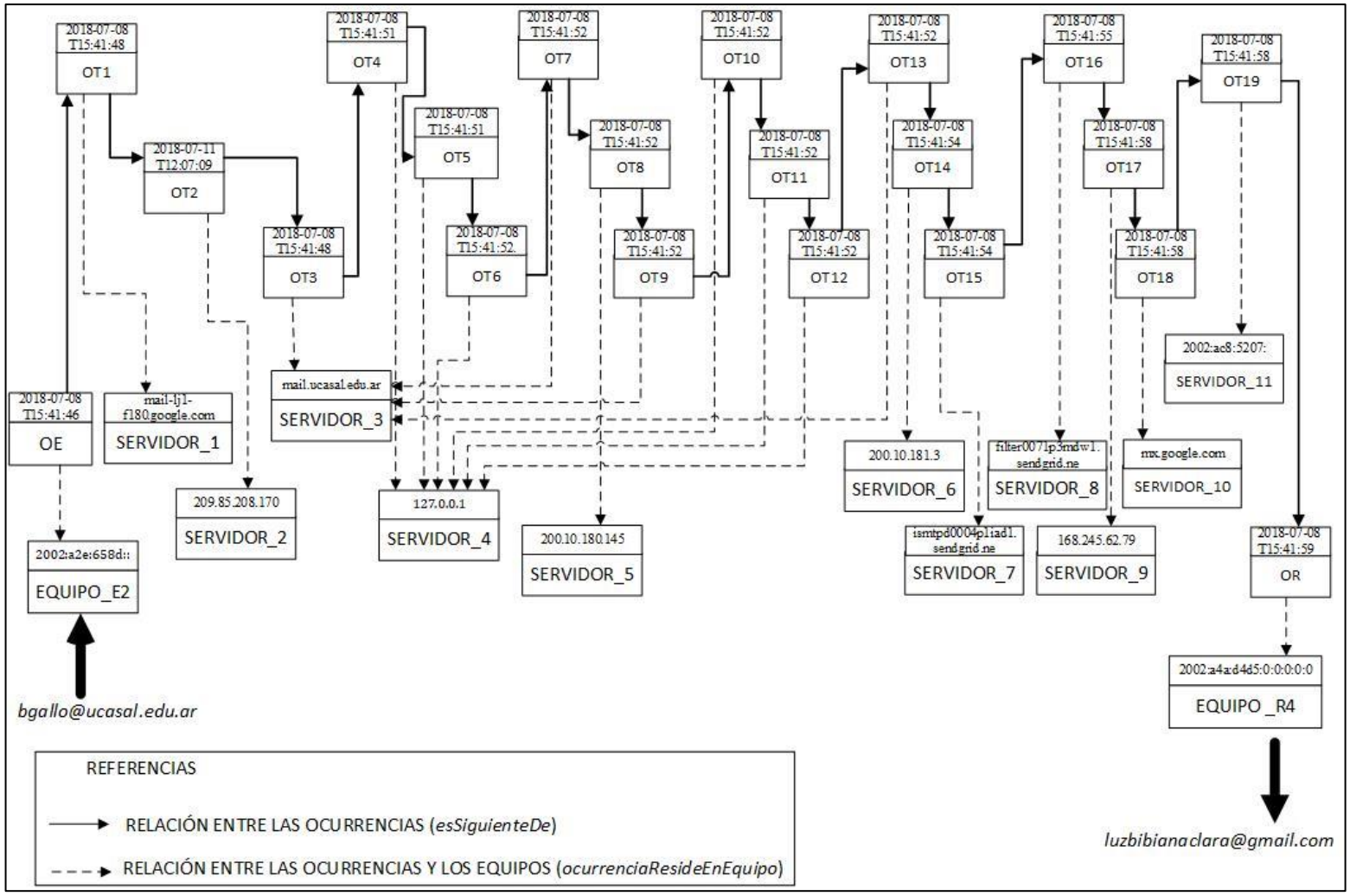


Figura 3-44: Esquema del proceso de Transmisión de las cabeceras CABECERA_R3

Analizando la Figura 3-44 en detalle, se observa que algunas ocurrencias de transmisión se almacenan en un mismo servidor. Tal es el caso del SERVIDOR_2 cuya identificación es el nombre de domino *mail.ucasal.edu.ar* que almacena 3 ocurrencias de transmisión de dicha cabecera (OT7, OT9 Y OT13), mientras que el SERVIDOR_4 cuya identificación es la dirección IP *127.0.0.1*, almacena 6 ocurrencias de transmisión de dicha cabecera (OT4, OT5, OT6, OT10, OT11, OT12). El proceso de transmisión de los hilos H1 y H2, son similares al indicado en la Figura 3-44.

A continuación se muestran algunos aspectos distintos de las instanciaciones realizadas para las tres cabeceras del Escenario 2. La instanciación de la cuenta de recepción y del proceso de transmisión es similar a los incluidos en la sección anterior, referida al caso de estudio del Escenario 1. Por esa razón no se repiten, y se muestran a continuación los componentes distintos de este escenario.

Al analizar las tres cabeceras en su conjunto, comienzan a aparecer datos relevantes, como por ejemplo: la ocurrencia de emisión de todos los correos recibidos es la misma, lo que asegura que fueron emitidos en el mismo momento desde la cuenta emisora. Esto se muestra en la Figura 3-45, resaltando con un recuadro en rojo las instancias de *hiloTieneOcurrencia* que vincula el objeto OE con H1, H2 y H3, que representan los tres hilos generados para las cabeceras en análisis.

Figura 3-45: Instanciación de ***Hilo*** y de ***OcurrenciaDeEmision***

La Figura 3-46 muestra el gráfico de Ontograf con la instanciación de ***Ocurrencia*** para CABECERA_R3, en la que se muestran la relación *esAnteriorA* para todas ellas. Por razones de espacio solo se etiquetaron las instancias de la citada relación que vinculan OT5-3 y OT6-3, así como la que enlaza OT7-3 con OT8-3.

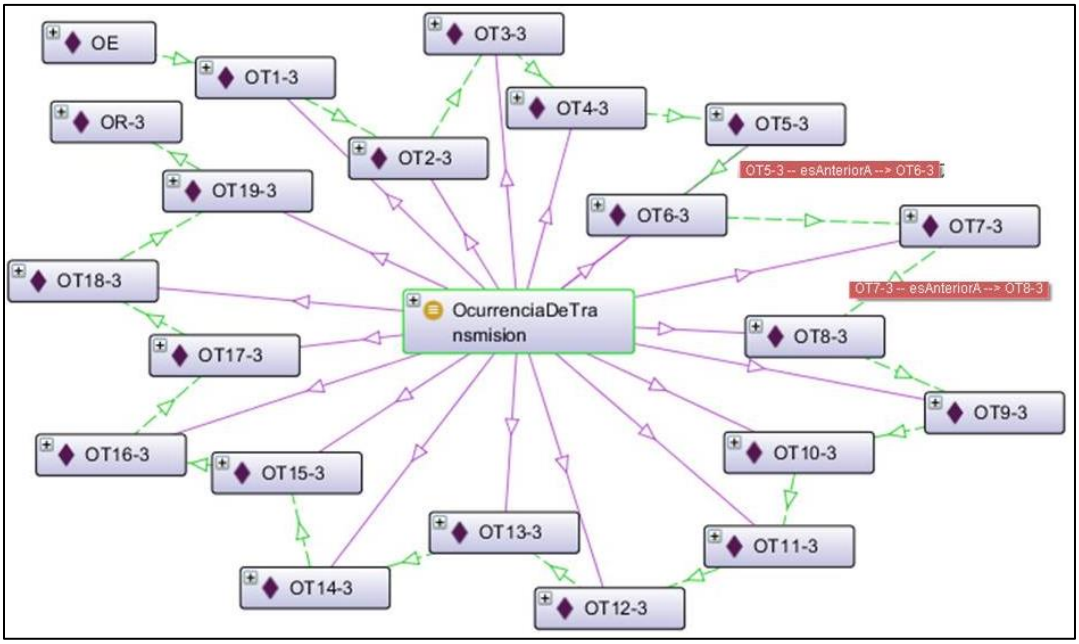


Figura 3-46: Instanciación de ***Ocurrencias*** y relación *esAnteriorA* de la cabecera CABECERA-R3

El resto de los valores instanciados para las clases ***Servidor, IdentificacionEquipo, Asunto, Cuerpo, Adjunto*** y ***Cabecera*** son similares a los indicados para el ejemplo del Escenario 1 en la sección anterior.

En particular, este ejemplo se distingue del Escenario 1, en que hay instancias de ***Ocurrencias*** correspondientes a distintos hilos que se vinculan a una misma instancia de ***Servidor.*** Esto, bajo el supuesto que se ha podido comprobar que la MAC Address del equipo en cuestión (SERVIDOR_2 por ejemplo), es siempre la misma en los 3 hilos, con lo cual, se puede asumir que se trata del mismo equipo físico.

A modo de ejemplo, se presenta un diagrama de Ontograf (Figura 3-47), en el que se observa la vinculación de las instancias de *Ocurrencia* con las correspondientes instancias de *Equipo* y de *IdentificacionEquipo,* para ello se toma como ejemplo la instancia que identifica al SERVIDOR_2, que es compartido por las instancias de los tres hilos que se están analizando.

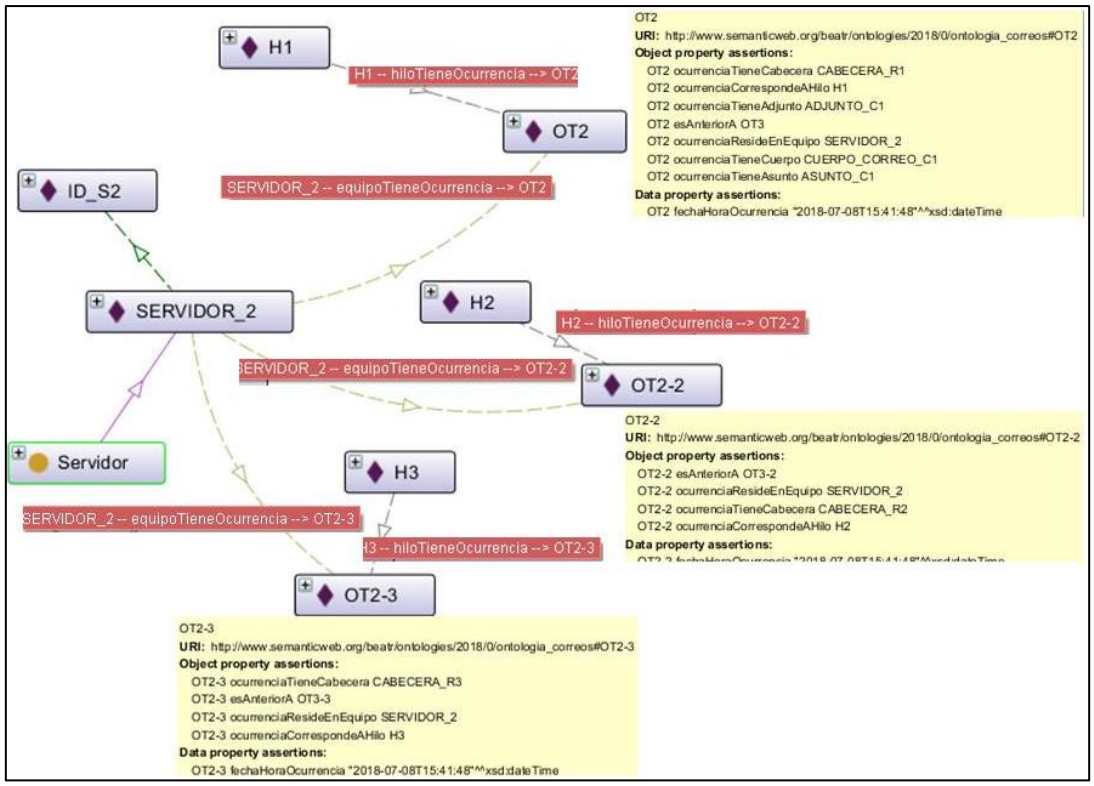


Figura 3-47: Instanciación del equipo SERVIDOR_2 con las ocurrencias que almacena

Sobre cada una de estas cabeceras, CABECERA_R1, CABECERA_R2 y CABECERA_R3, se pueden consultar las preguntas de competencia del caso anterior, es decir aquellas dirigidas a obtener información de un correo específico e identificadas como PREGUNTAS DE COMPETENCIA PC01 a PC09, y que por razones de espacio, no se repetirán en esta sección.

Se tomarán las cabeceras CABECERA_R1, CABECERA_R2 y CABECERA_R3 para responder las preguntas de competencia PC10 a PC21.

- **PC10: Dado una cuenta C ¿cuáles son los correos que emitió?**

Para el ejemplo del escenario 2, la pregunta sería "Dada la cuenta bgallo@ucasal.edu.ar ¿cuáles con los correos que emitió?", y la consulta SPARQL que permite responder a esta pregunta de competencia es la siguiente:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT   ?correo ?asunto ?fechaEmision ?nombreCuentaReceptor
    WHERE  { ?cuenta oc:cuentaEmisorEmiteCorreo ?correo.
             ?cuentaRep oc:cuentaReceptorRecibeCorreo ?correo.
             ?cuentaRep oc:cuentaCorreo ?nombreCuentaReceptor.
             ?correo oc:correoTieneSecuencia ?s.
             ?s oc:secuenciaTieneHilo ?h.
```

```
?h oc:hiloTieneOcurrencia ?o.
?o oc:ocurrenciaTieneAsunto ?a.
?a oc:contenidoAsunto ?asunto.
?o rdf:type oc:OcurrenciaDeEmision.
?o oc:fechaHoraOcurrencia ?fechaEmision.
FILTER (?cuenta=oc:CUENTA_E1)}
```

La Figura 3-48 muestra el visor de consultas SPARQL que indica los resultados encontrados, al considerar la cuenta bgallo@ucasal.edu.ar como cuenta emisora, observando que hay un correo C1 para el cual se muestra el asunto, la fecha de emisión y la cuenta la que lo recibió, todo ello obtenido de las correspondientes cabeceras CABECERA_R1, CABECERA_R2 y CABECERA_R3 analizadas.

SPARQL query:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo ?asunto ?fechaEmision ?nombreCuentaReceptor
        WHERE  { ?cuenta oc:cuentaEmisorEmiteCorreo ?correo.
                 ?cuentaRep oc:cuentaReceptorRecibeCorreo ?correo.
                 ?cuentaRep oc:cuentaCorreo ?nombreCuentaReceptor.
                 ?correo oc:correoTieneSecuencia ?s.
                 ?s oc:secuenciaTieneHilo ?h.
                 ?h oc:hiloTieneOcurrencia ?o.
                 ?o oc:ocurrenciaTieneAsunto ?a.
                 ?a oc:contenidoAsunto ?asunto.
                 ?o rdf:type oc:OcurrenciaDeEmision.
                 ?o oc:fechaHoraOcurrencia ?fechaEmision.
                 FILTER (?cuenta=oc:CUENTA_E1)}
```

correo	asunto	fechaEmision	nombreCuentaReceptor
C1	"Colaboración en investigación"	"2018-07-08T15:41:46"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"luzbibianaclara"@gmail.com
C1	"Colaboración en investigación"	"2018-07-08T15:41:46"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"Enzo Notario"
C1	"Colaboración en investigación"	"2018-07-08T15:41:46"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"erivetti83"@gmail.com

Figura 3-48: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC09

- **PC14: Dada una dirección IP ¿cuál sería la localización geográfica del equipo asociado?**

En este escenario también se puede verificar la pregunta de competencia PC14 para el caso particular de la instancia ID_S2, que almacena la dirección IP 209.85.208.170 correspondiente al SERVIDOR_2. El código SPARQL de la consulta es el siguiente:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT  ?direccion ?localizacion
    WHERE  { ?IP oc:geoLocalizacionIP ?localizacion.
             ?IP oc:identificadorEquipo ?direccion.
             FILTER (?IP=oc:ID_S2).}
```

En este caso, la Figura 3-49 muestra el visor de consultas SPARQL que indica como resultado que la dirección ID_S2 cuyo valor es 209.85.208.170 está localizada geográficamente según las coordenadas *Latitud 39.0438 y Longitud -77.4874*.

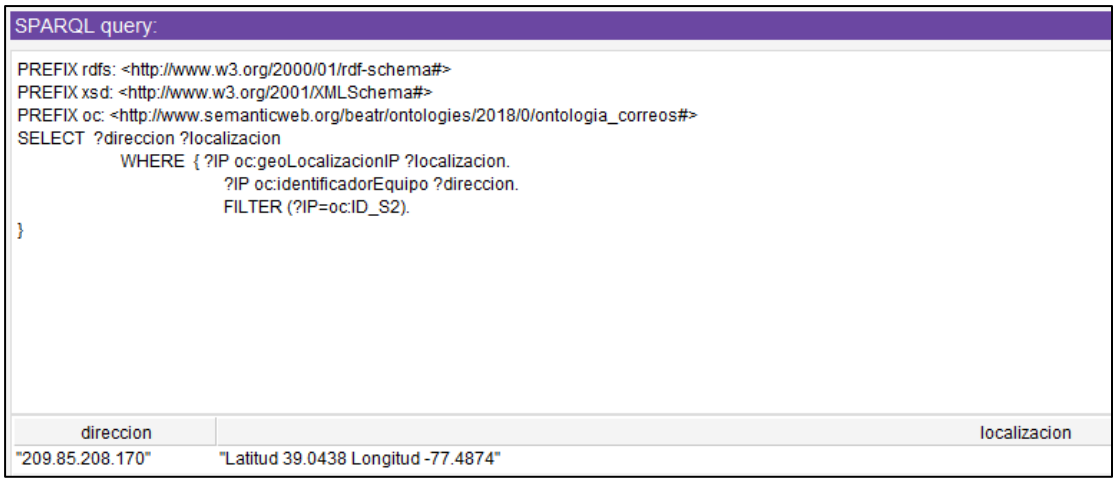


Figura 3-49: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC14

Cabe mencionar que los datos referidos a la geolocalización de la dirección IP no se obtuvieron de las cabeceras, como se explicó en el apartado 2.5.4 del Capítulo 2 al describir el procedimiento de análisis forense, sino que se consulta a los directorios de Direcciones IP habituales, sitios a los que se accede directamente desde ObE Forensics, y se presentan al Perito. Los datos instanciados en OntoFoCE que se indican en la pantalla de la Figura 3-49 fueron obtenidos de manera manual desde el sitio http://ip-api.com/docs/api:json, el cual es consultado de manera automática desde la aplicación ObE Forensics al responder esta pregunta de competencia.

- **PC15: ¿Cuáles son los correos que han pasado por el dispositivo que posee una IP dada?**

Considerando la Figura 3-47 en la que se muestra la Instanciación de SERVIDOR2 y todos los objetos de OCURRENCIA vinculados, se puede aprovechar el ejemplo para mostrar los resultados de la pregunta de competencia PC15. En este caso, la pregunta sería "¿Cuáles son los correos que han pasado por el dispositivo que posee la dirección IP 209.85.208.170?" siendo la dirección IP de consulta el valor del atributo *identificadorEquipo* y el código SPARQL es el siguiente

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?equipo ?Cabecera
        WHERE { ?equipo oc:equipoTieneId ?IP.
                ?ocurrencia oc:ocurrenciaResideEnEquipo ?equipo.
                ?hilo oc:hiloTieneOcurrencia ?ocurrencia.
                ?s oc:secuenciaTieneHilo ?hilo.
```

?correo oc:correoTieneSecuencia ?s.
?e oc:cuentaEmisorEmiteCorreo ?correo.
?IP oc:identificadorEquipo ?direccion.
?ocurrencia oc:ocurrenciaTieneCabecera ?Cabecera.
FILTER regex(?direccion,"209.85.208.170").}

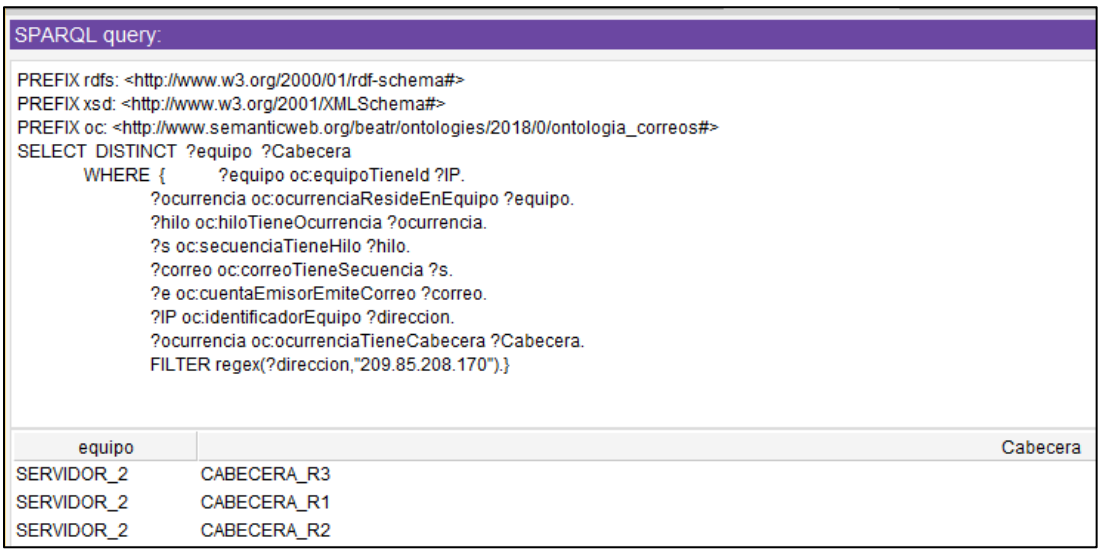


Figura3-50: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC15

La Figura 3-50 muestra los datos resultantes de la consulta para el servidor identificado como SERVIDOR_2, indicando que por ese equipo pasaron tres cabeceras denominadas CABECERA_R1, CABECERA_R2 y CABECERA_R3.

Con el ejemplo del Escenario 2 se probaron las preguntas de competencia referidas a una cuenta desde la cual se envía un correo a varios receptores, a continuación se completa el análisis de las preguntas de competencia restantes, considerando el caso de un conjunto de correos referidos a la misma cuenta.

3.5.3 Análisis Forense de un Conjunto de Correos

Este caso corresponde al Escenario 3, señalado en la Figura 3-8: Escenarios de Pericias de Correos Electrónicos. En este caso, la pericia debe realizarse sobre un conjunto de correos intercambiados entre las cuentas peritadas.

Supongamos que se toma los correos analizados en el escenario 2 a los que se agregan otros de un segundo emisor, denominado JUAN, dirigidos a los receptores ya indicados. De este modo se puede ejemplificar el análisis de un conjunto de correos en los que intervienen dos cuentas emisores y tres cuentas receptoras,

mediante el intercambio epistolar entre los correos emisores y receptores que señala la Figura 3-36:

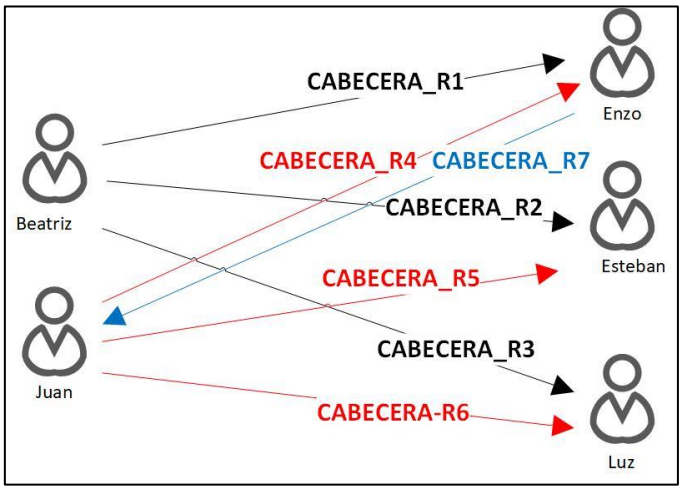


Figura 3-51: Esquema de Correos Enviados (CE) y Recibidos (CR) entre cuentas del Escenario 3

En la Figura 3-51 se señalan las cabeceras CABECERA_R1, CABECERA_R2 y CABECERA_R3 que ya se trabajaron en el Escenario 2, a los que se agregan los correos nuevos señalados como CABECERA_R4, CABECERA_R5 y CABECERA_R6, emitidos por JUAN a cada uno de los receptores del ejemplo. La última cabecera señalada CABECERA_R7, permite simular el caso de cuentas de correo que actúan como emisor y receptor a la vez en el mismo espacio de la pericia que se está realizando.

El Perito realiza el procedimiento definido para la realización del análisis forense, siguiendo todas las fases señaladas hasta obtener las cabeceras de correos recibidos y se generan las instancias respectivas.

Cabe mencionar que a fin de darle variación al escenario, las cabeceras identificadas como CABECERA_R4, CABECERA_R5 y CABECERA_R6 no son cabeceras provenientes de un único correo emitido a los tres receptores, sino que provienen de sendos correos enviados por Juan a cada receptor separadamente. Así, en este escenario habrá dos cuentas emisoras y tres cuentas receptoras, y se analizarán seis cabeceras de correo, las tres ya analizadas en el escenario 2 y las nuevas identificadas como CABECERA_R4, CABECERA_R5 y CABECERA_R6 en la Figura 3-51. A continuación, se muestran las cabeceras que se agregan en el análisis en este escenario.

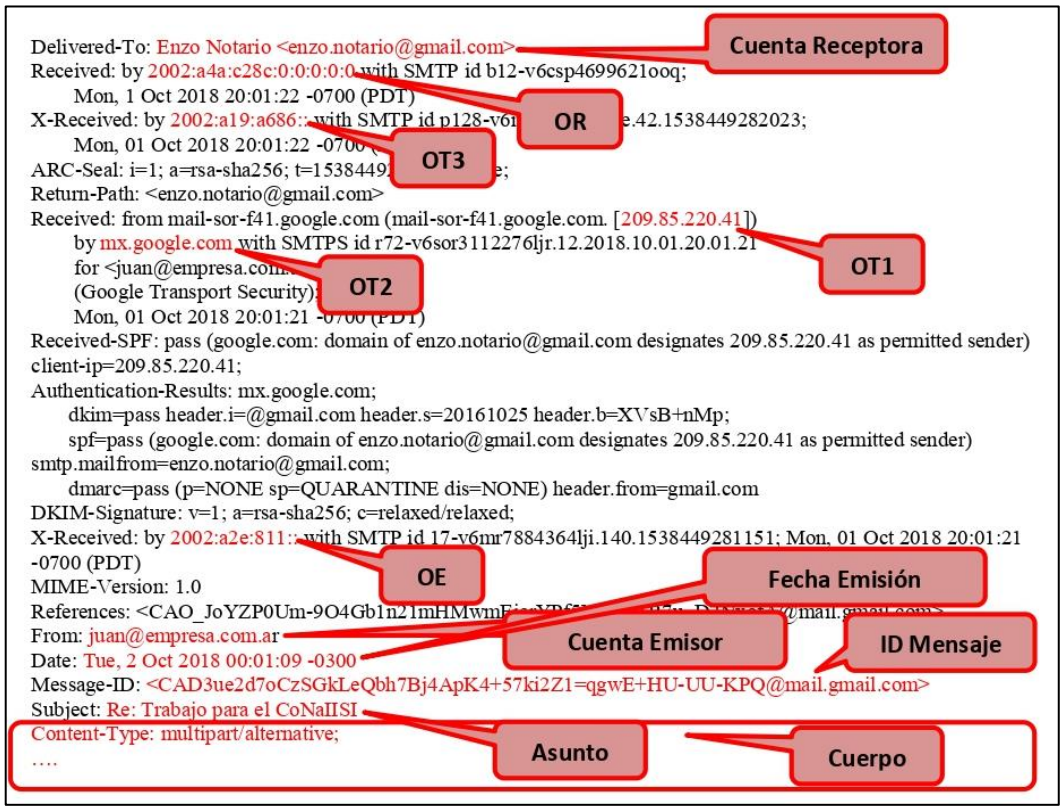


Figura 3-52: Cabecera identificada como CABECERA_R4

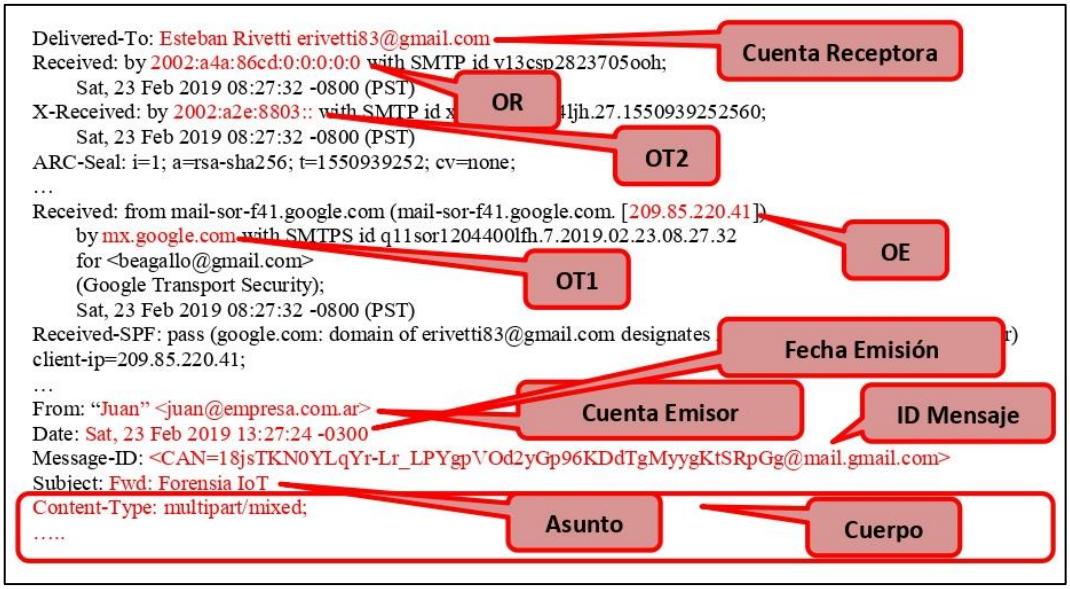


Figura 3-53: Cabecera identificada como CABECERA_R5

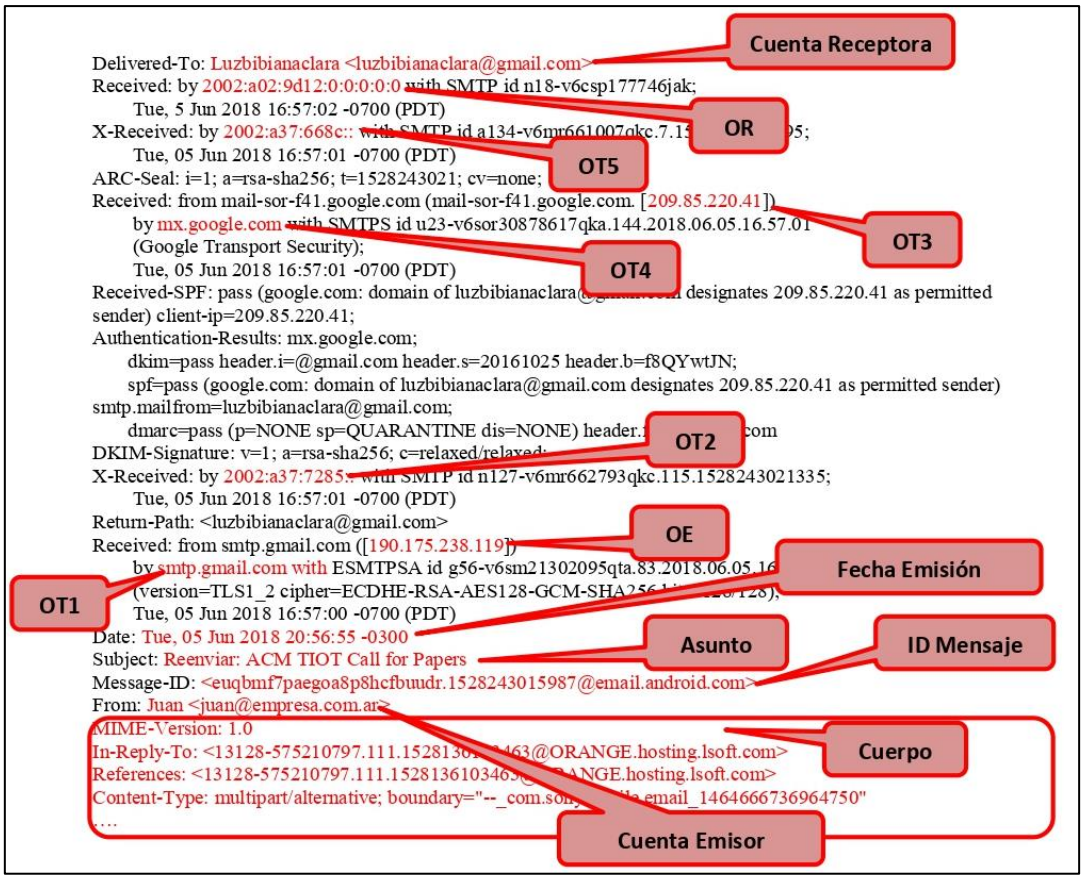


Figura 3-54: Cabecera identificada como CABECERA_R6

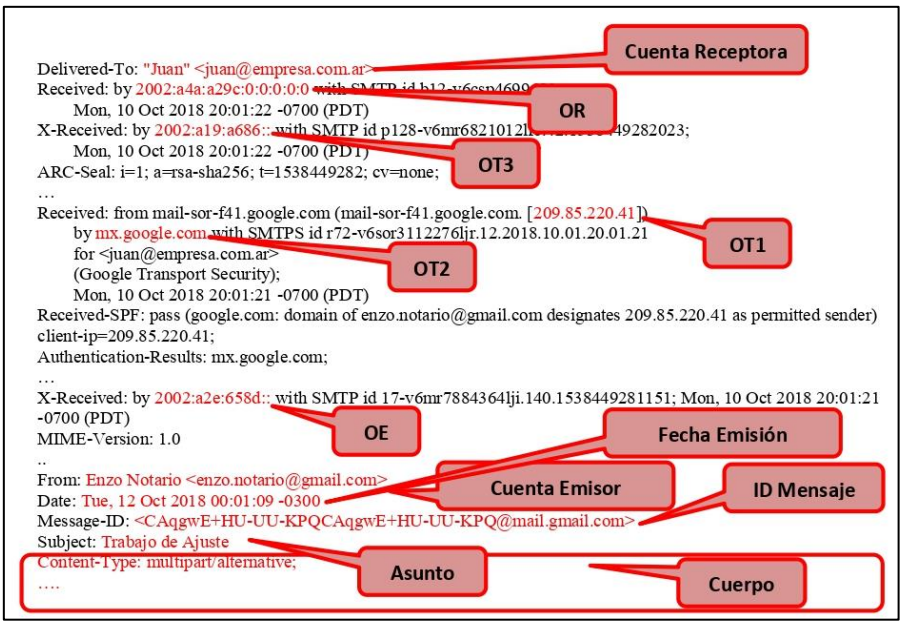


Figura 3-55: Cabecera identificada como CABECERA_R7

Para cada cabecera se realiza la instanciación correspondiente de las cuentas de emisión/recepción y del proceso de transmisión de manera similar a lo indicado en la sección 3.5.1. Por esa razón, y en el afán de mostrar los componentes distintos del Escenario 3, no se repiten, y se muestran a continuación los componentes distintos de este escenario.

El gráfico de Ontograf que se muestra en la Figura 3-56 muestra las secuencias e hilos que vinculan las cuentas emisoras y receptoras del ejemplo al crear las instancias correspondientes.

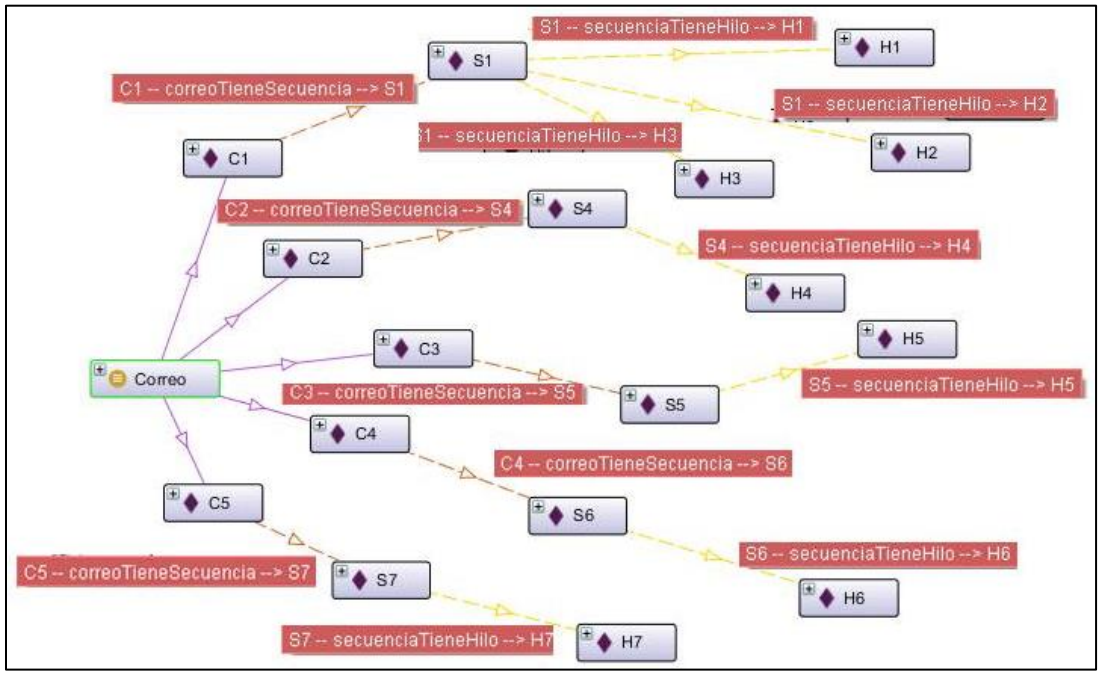


Figura 3-56: Instanciación de ***Secuencia*** e ***Hilo*** para los correos del Escenario 3

Para los cinco correos que se muestran en la Figura 3-56 se muestra su instanciación en la Figura 3-57, en la que se puede observar las instancias de ***Cuenta*** correspondientes.

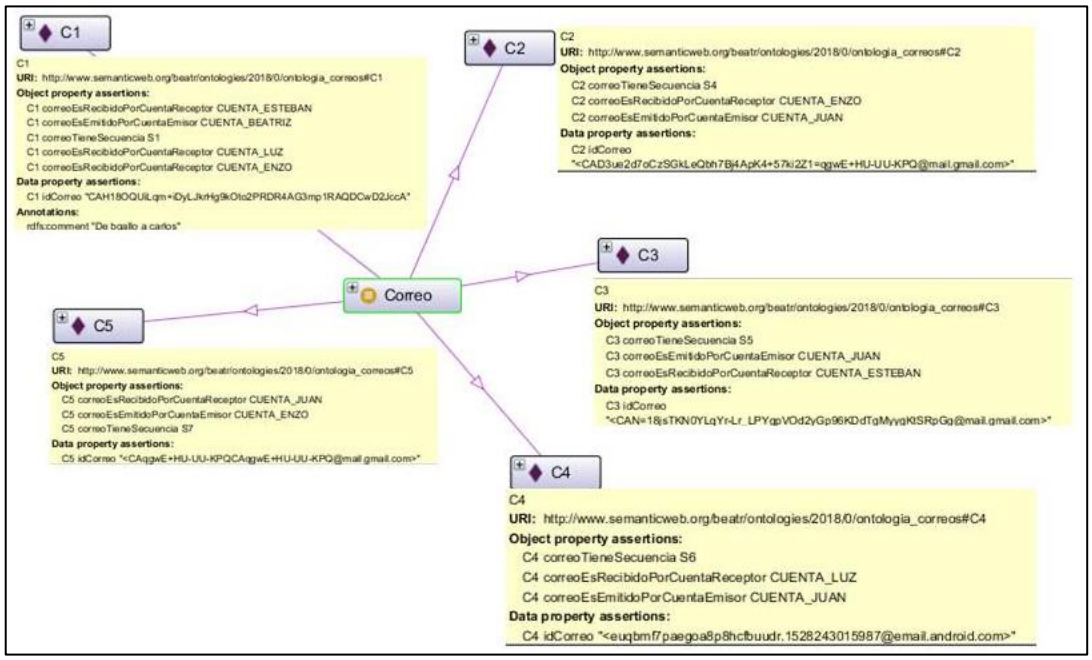


Figura 3-57: Instanciación de ***Correo*** para los 5 correos del Escenario 3

Obsérvese que la Figura 3-57 muestra el caso de los correos identificados como C2 y C5, que se vinculan a CUENTA_ENZO y CUENTA_JUAN y que actúan como cuenta emisora/receptora en el correo C2 y a la inversa en el correo C5. Esto se observa con más claridad en la Figura 3-58, en la que se señala (remarcado con un recuadro rojo) que el individuo CUENTA_ENZO está definido explícitamente como una instancia de la clase ***CuentaReceptor*** (la clase de la que es instancia se ve negritas y fondo blanco)**,** porque así se vinculó en la primera instanciación (aquella realizada en el Escenario 2) pero que además el razonador la infiere como instancia de ***CuentaEmisor*** (el nombre de la clase está resaltado con un fondo amarillo y con una fuente normal, sin negrita)**,** ya que ese es rol que le corresponde en el Escenario 3**.**

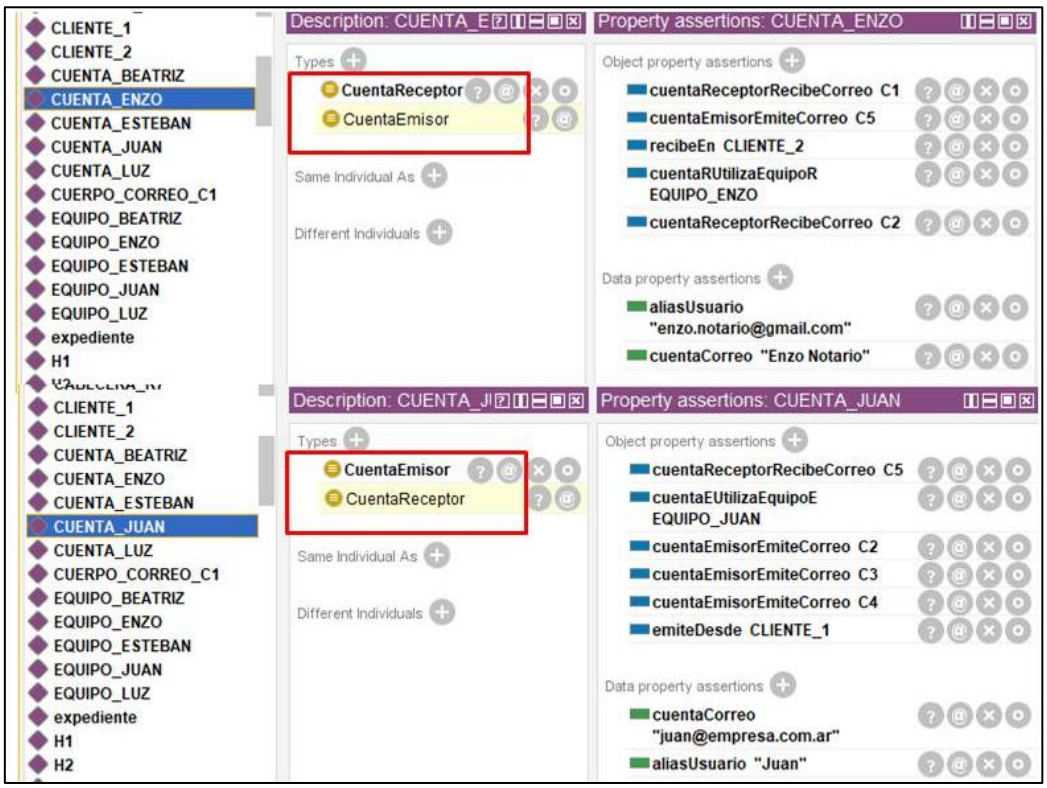


Figura 3-58: Instanciación de ***Correo*** para los 5 correos del Escenario 3

Las preguntas de competencia que se pueden trabajar con este conjunto de correos son las siguientes.

- **PC11: Dado una cuenta C ¿cuáles son los correos que recibió?**

Esta pregunta de competencia se aplicaría al ejemplo del escenario 3 para averiguar cuáles son los correos recibidos por la cuenta *enzo.notario@gmail.com*. El

código SPARQL de la pregunta de competencia P11 para la cuenta indicada es el siguiente:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo ?asunto ?fechaEmision  ?cuentaEmisora
    WHERE  { ?cuenta oc:cuentaReceptorRecibeCorreo ?correo.
             ?correo oc:correoTieneSecuencia ?s.
             ?correo oc:correoEsEmitidoPorCuentaEmisor ?ctaEmisor.
             ?ctaEmisor oc:cuentaCorreo ?cuentaEmisora.
             ?s oc:secuenciaTieneHilo ?h.
             ?h oc:hiloTieneOcurrencia ?o.
             ?o oc:ocurrenciaTieneAsunto ?a.
             ?a oc:contenidoAsunto ?asunto.
             ?o rdf:type oc:OcurrenciaDeRecepcion.
             ?o oc:fechaHoraOcurrencia ?fechaEmision.
             FILTER (?cuenta=oc:CUENTA_ENZO).}
```

El visor de consultas, que se muestra en la Figura 3-59, señala que en la cuenta enzo.notario@gmail.com se recibieron dos correos identificados como C1 y C2, para los cuales se muestra la fecha de emisión, el asunto y el emisor del correo.

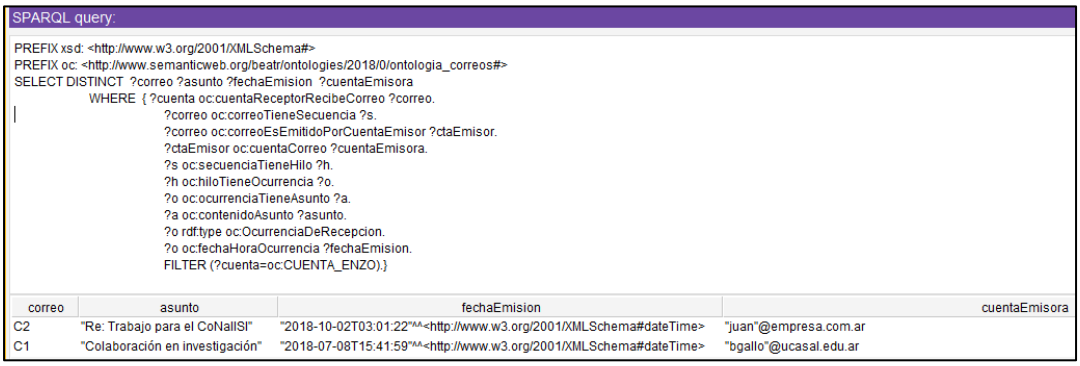


Figura 3-59: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC11

- **PC12: Dado una cuenta C1 ¿se ha emitido un correo hacia la cuenta C2?**

La pregunta de competencia PC12 se puede utilizar para verificar si desde la cuenta juan@empresa.com.ar se emitió un correo a la cuenta erivetti83@gmail.com.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT   ?correo ?asunto ?fecha
    WHERE  { ?cuentaC1 oc:cuentaEmisorEmiteCorreo ?correo.
             ?cuentaC2 oc:cuentaReceptorRecibeCorreo ?correo.
             ?cuentaC1 oc:cuentaCorreo ?cuentaEmisora.
             ?cuentaC2 oc:cuentaCorreo ?cuentaReceptor.
             ?correo oc:correoTieneSecuencia ?s.
             ?s oc:secuenciaTieneHilo ?h.
             ?h oc:hiloTieneOcurrencia ?o.
             ?o rdf:type oc:OcurrenciaDeEmision.
             ?o oc:ocurrenciaTieneAsunto ?a.
             ?o oc:fechaHoraOcurrencia ?fecha.
             ?a oc:contenidoAsunto ?asunto.
    FILTER (?cuentaC1=oc:CUENTA_JUAN &&  ?cuentaC2=oc:CUENTA_ESTEBAN).}
```

En la Figura 3-60 se muestra los resultados de la consulta, señalando que desde la cuenta *juan@empresa.com.ar* se emitió un correo a la cuenta *erivetti83@gmail.com* identificado como correo C3 cuyo asunto es *Fwd: Forensia IoT* y fue enviado el *23 de Feb de 2019 16:27:32*.

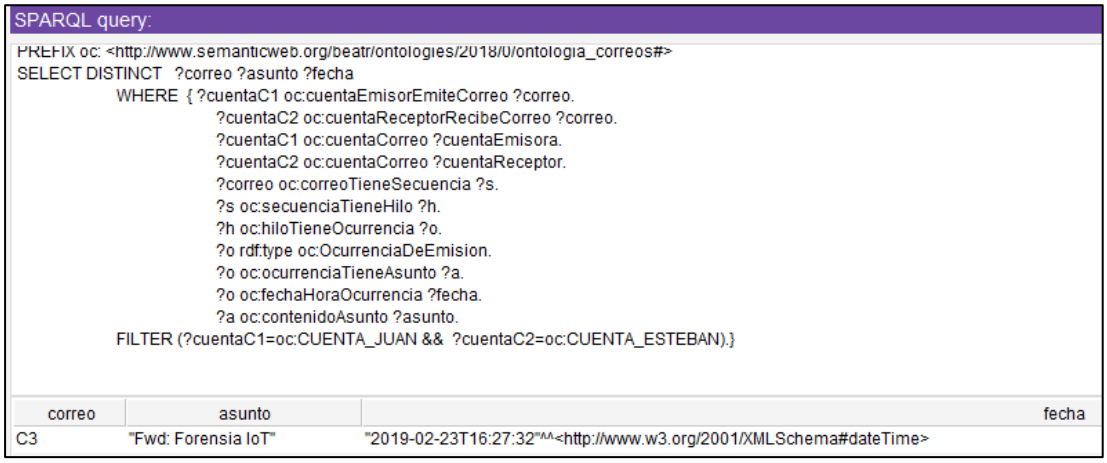


Figura 3-60: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC12

- **PC13: Dado una cuenta C1 ¿se ha recibido un correo desde la cuenta C2?**

Para ejemplificar esta pregunta de competencia se puede indagar si en la cuenta *luzbibianaclara@gmail.com* se ha recibido un correo enviado desde la cuenta *juan@empresa.com.ar*. El código SPARQL de esa consulta es el siguiente:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?correo ?asunto ?fecha
        WHERE { ?cuentaC1 oc:cuentaReceptorRecibeCorreo ?correo.
                ?cuentaC2 oc:cuentaEmisorEmiteCorreo ?correo.
                ?correo oc:correoTieneSecuencia ?s.
                ?s oc:secuenciaTieneHilo ?h.
                ?h oc:hiloTieneOcurrencia ?o.
                ?o rdf:type oc:OcurrenciaDeRecepcion.
                ?o oc:fechaHoraOcurrencia ?fecha.
                ?o oc:ocurrenciaTieneAsunto ?a.
                ?a oc:contenidoAsunto ?asunto.
        FILTER (?cuentaC1=oc:CUENTA_LUZ &&  ?cuentaC2=oc:CUENTA_JUAN).}
```

Los resultados de esta consulta se observan en la Figura 3-61, allí se indica que en la cuenta *luzbibianaclara@gmail.com* se recibió un correo desde la cuenta *juan@empresa.com.ar* identificado como C4 cuyo asunto es *Reenviar: ACM TIOT Call for Papers* y fue emitido el *5 de Jun 2018 23:57:02*.

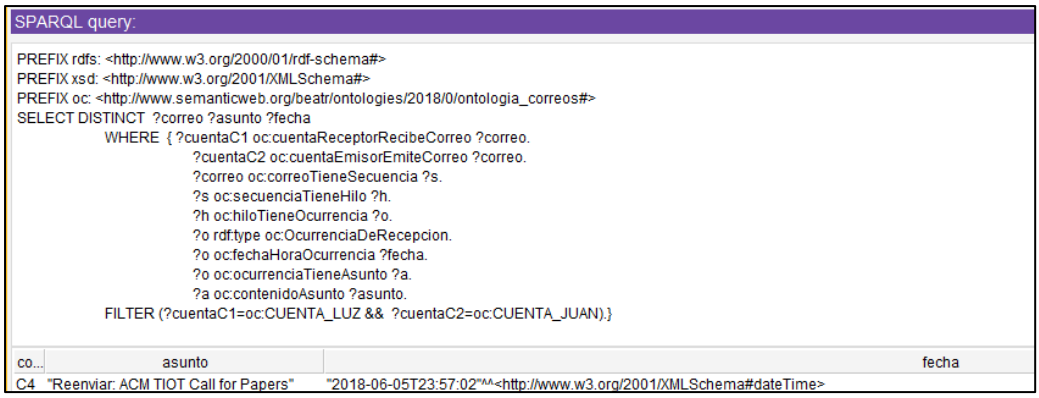


Figura 3-61: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC13

- **PC16: ¿Cuáles son los mails enviados desde una determinada cuenta en una fecha dada?**

Esta pregunta de competencia filtra los correos enviados desde una cuenta según una fecha determinada, se puede tomar como ejemplo para seleccionar y mostrar los correos enviados desde la cuenta *juan@empresa.com.ar* en una fecha dada.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?correo ?cuentaReceptora ?asunto ?fecha
WHERE { ?cuenta rdf:type oc:CuentaEmisor.
?cuenta oc:cuentaEmisorEmiteCorreo ?correo.
?correo oc:correoTieneSecuencia ?s.
?s oc:secuenciaTieneHilo ?h.
?Ocurrencia oc:ocurrenciaCorrespondeAHilo ?h.
?Ocurrencia oc:ocurrenciaTieneAsunto ?a.
?a oc:contenidoAsunto ?asunto.
?Ocurrencia rdf:type oc:OcurrenciaDeEmision.
?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
?cuentaR oc:cuentaReceptorRecibeCorreo ?correo.
?cuentaR oc:cuentaCorreo ?cuentaReceptora.
FILTER(("2018-10-02T03:01:21"^^xsd:dateTime=?fecha) && (?cuenta=oc:CUENTA_JUAN)).}

Se toma como parámetro de fecha el valor "*2 oct 2018 03:01:21"* para el cual el resultado señala que se emitió un correo identificado como C2 (ver Figura 3-62).

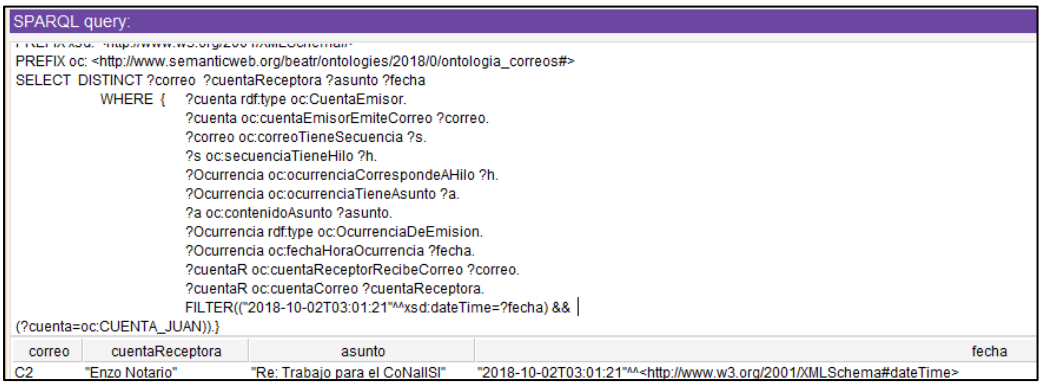
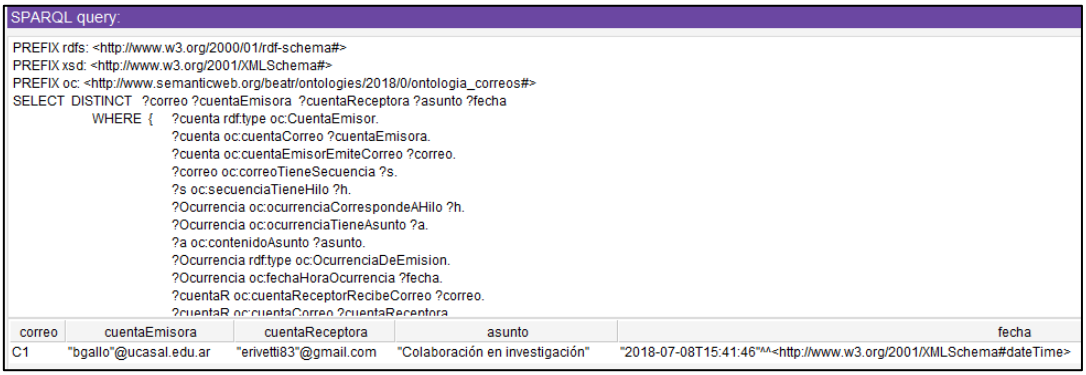


Figura 3-62: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC16

- **PC17: ¿Cuáles son los mails recibidos por una determinada cuenta en una fecha dada?**

La pregunta de competencia PC17 también requiere un parámetro de fecha, como por ejemplo el valor 8 Jul 2018 15:41:46 y una cuenta, por ejemplo, la de *erivetti83@gmail.com* para el cual el resultado señala que se emitió un correo identificado como C1 (Figura 3-63). El código SPARQL de la consulta para esta pregunta de competencia es el siguiente:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT  DISTINCT   ?correo ?cuentaEmisora  ?cuentaReceptora ?asunto?fecha
        WHERE  {        ?cuenta rdf:type oc:CuentaEmisor.
                ?cuenta oc:cuentaCorreo ?cuentaEmisora.
                ?cuenta oc:cuentaEmisorEmiteCorreo ?correo.
                ?correo oc:correoTieneSecuencia ?s.
                ?s oc:secuenciaTieneHilo ?h.
                ?Ocurrencia oc:ocurrenciaCorrespondeAHilo ?h.
                ?Ocurrencia oc:ocurrenciaTieneAsunto ?a.
                ?a oc:contenidoAsunto ?asunto.
                ?Ocurrencia rdf:type oc:OcurrenciaDeEmision.
                ?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
                ?cuentaR oc:cuentaReceptorRecibeCorreo ?correo.
                ?cuentaR oc:cuentaCorreo ?cuentaReceptora.
                FILTER(("2018-07-08T15:41:46"^^xsd:dateTime=?fecha) &&
                (?cuentaR=oc:CUENTA_ESTEBAN)).}
```


SPARQL query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT  DISTINCT  ?correo ?cuentaEmisora  ?cuentaReceptora ?asunto ?fecha
        WHERE  {    ?cuenta rdf:type oc:CuentaEmisor.
                    ?cuenta oc:cuentaCorreo ?cuentaEmisora.
                    ?cuenta oc:cuentaEmisorEmiteCorreo ?correo.
                    ?correo oc:correoTieneSecuencia ?s.
                    ?s oc:secuenciaTieneHilo ?h.
                    ?Ocurrencia oc:ocurrenciaCorrespondeAHilo ?h.
                    ?Ocurrencia oc:ocurrenciaTieneAsunto ?a.
                    ?a oc:contenidoAsunto ?asunto.
                    ?Ocurrencia rdf:type oc:OcurrenciaDeEmision.
                    ?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
                    ?cuentaR oc:cuentaReceptorRecibeCorreo ?correo.
                    ?cuentaR oc:cuentaCorreo ?cuentaReceptora.
```

correo	cuentaEmisora	cuentaReceptora	asunto	fecha
C1	"bgallo"@ucasal.edu.ar	"erivetti83"@gmail.com	"Colaboración en investigación"	"2018-07-08T15:41:46"^^<http://www.w3.org/2001/XMLSchema#dateTime>

Figura 3-63: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC17

Las preguntas referidas a la búsqueda de palabras claves, PC18, PC19 y PC20, se pueden comprobar si es que previamente se avanza en el desarrollo del algoritmo de búsqueda de dichos términos en las instancias generadas para el *Cuerpo* de cada correo que usualmente se guardan como datos encriptados en la cabecera, además de resolver la problemática de analizar un archivo adjunto, si se considera que el mismo puede responder a diferentes formatos (imagen, audio, salida de programas específicos como AutoCAD, Excel, entre otros). No obstante, para avanzar con la

descripción de las preguntas de competencia, se va a suponer que estos pasos están resueltos y que se cuenta con la información del *Cuerpo* y el *Adjunto* en formato texto.

Para responder las preguntas de competencia PC18, PC19 y PC20 vamos a suponer que el Juez solicita que se busquen aquellos correos que hagan referencia a una palabra clave determinada, supongamos *Colaboración.* Es posible realizar la búsqueda en los 3 elementos del correo que puede contener texto: asunto, cuerpo y adjunto. A continuación se trabaja particularmente las preguntas PC18, PC19 y PC20 para la palabra clave *Colaboración.*

- **PC18: Dada una palabra clave ¿Figura en el asunto de un correo?**

El código SPARQL de la consulta para esta pregunta de competencia es el siguiente:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo  ?asunto ?fecha
        WHERE  { ?correo oc:correoTieneSecuencia ?s.
                 ?s oc:secuenciaTieneHilo ?h.
                 ?h oc:hiloTieneOcurrencia ?Ocurrencia.
                 ?Ocurrencia rdf:type oc:OcurrenciaDeRecepcion.
                 ?Ocurrencia oc:ocurrenciaTieneAsunto ?a.
                 ?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
                 ?a oc:contenidoAsunto ?asunto.
                 ?a oc:asuntoContienePalabraClave ?palabra.
                 ?palabra oc:contenidoPalabraClave ?pc.
                 FILTER regex(?pc,"Colaboración").}
```

La Figura 3-64 muestra el resultado del visor de consultas SPARQL para los datos ingresados, identificando el correo C1, el asunto y la fecha de recepción del mismo.

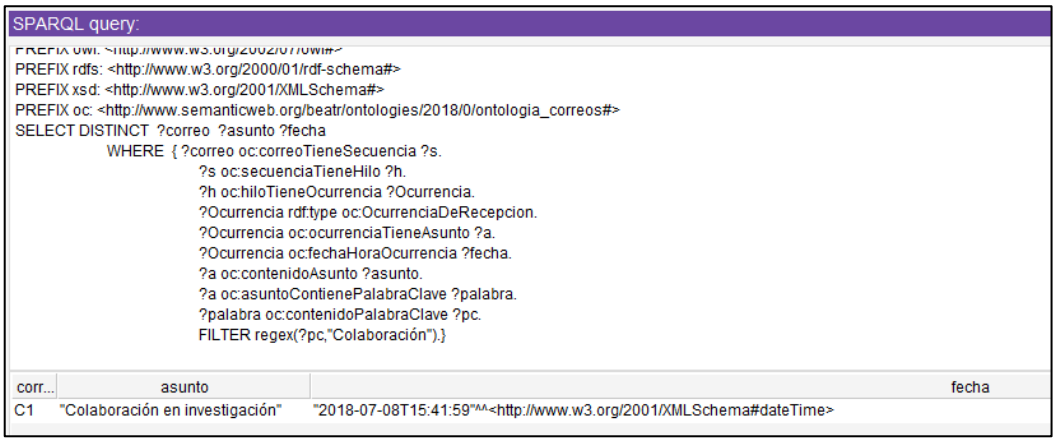


Figura 3-64: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC18

- **PC19: Dada una palabra clave ¿Figura en el cuerpo de un correo?**

En este caso la consulta para esta pregunta se realiza en base al siguiente código SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo  ?cuerpo ?fecha
        WHERE  { ?correo oc:correoTieneSecuencia ?s.
                 ?s oc:secuenciaTieneHilo ?h.
                 ?h oc:hiloTieneOcurrencia ?Ocurrencia.
                 ?Ocurrencia rdf:type oc:OcurrenciaDeRecepcion.
                 ?Ocurrencia oc:ocurrenciaTieneCuerpo ?c.
                 ?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
                 ?c oc:contenidoCuerpo ?cuerpo.
                 ?c oc:cuerpoContienePalabraClave ?palabra.
                 ?palabra oc:contenidoPalabraClave ?pc.
                 FILTER regex(?pc,"Colaboración").}
```

En la Figura 3-65 se muestran los datos resultantes de la consulta: el correo identificado como C1 emitido con fecha *8 de jul de 2018 15:41:50* contiene la palabra clave *Colaboración* en el cuerpo.

SPARQL query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo  ?cuerpo ?fecha
        WHERE  { ?correo oc:correoTieneSecuencia ?s.
                 ?s oc:secuenciaTieneHilo ?h.
                 ?h oc:hiloTieneOcurrencia ?Ocurrencia.
                 ?Ocurrencia rdf:type oc:OcurrenciaDeRecepcion.
                 ?Ocurrencia oc:ocurrenciaTieneCuerpo ?c.
                 ?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
                 ?c oc:contenidoCuerpo ?cuerpo.
                 ?c oc:cuerpoContienePalabraClave ?palabra.
                 ?palabra oc:contenidoPalabraClave ?pc.
                 FILTER regex(?pc,"Colaboración").}
```

correo	cuerpo	fecha
C1	"Estimados: He seleccionado algunas personas	"2018-07-08T15:41:59"^^<http://www.w3.org/2001/XMLSchema#dateTime>

Figura 3-65: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC19

- **PC20: Dada una palabra clave ¿Figura en el adjunto de un correo?**

Esta pregunta de competencia se responde desde el siguiente código SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo  ?contenidoAdjunto  ?fecha
        WHERE  { ?correo oc:correoTieneSecuencia ?s.
                 ?s oc:secuenciaTieneHilo ?h.
                 ?h oc:hiloTieneOcurrencia ?Ocurrencia.
                 ?Ocurrencia rdf:type oc:OcurrenciaDeRecepcion.
                 ?Ocurrencia oc:ocurrenciaTieneAdjunto ?ad.
```

?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
?ad oc:contenidoAdjunto ?contenidoAdjunto.
?ad oc:adjuntoContienePalabraClave ?palabra.
?palabra oc:contenidoPalabraClave ?pc.
FILTER regex(?pc,"Colaboración").}

En la Figura 3-66 el visor SPARQL muestra como resultado a esta pregunta de competencia los datos referidos a la identificación del correo, el contenido del archivo adjunto y la fecha de recepción del correo.

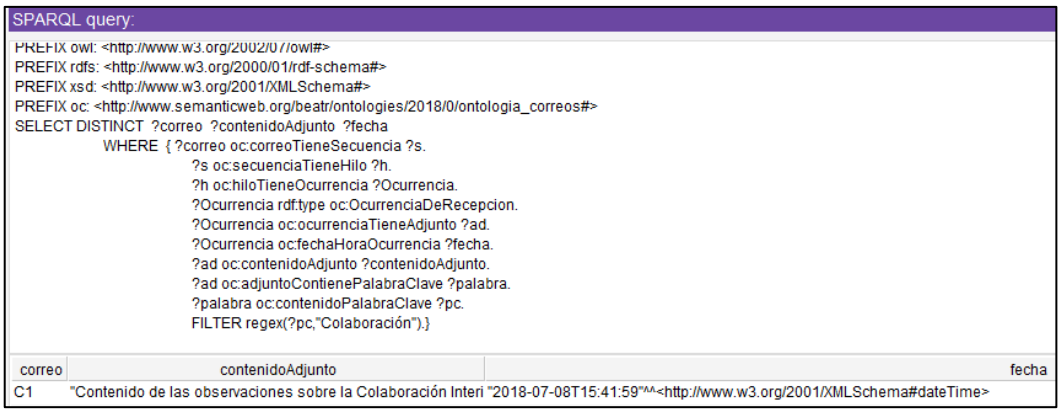


Figura 3-66: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC20

- **PC21: ¿Cuáles son los correos intercambiados entre las cuentas C1 y C2 en un rango de fechas dado?**

Esta pregunta de competencia permite filtrar los correos intercambiados entre dos cuentas. Tómese como ejemplo las cuentas *juan@empresa.com.ar* y *enzo.notario@gmail.com*, para lo cual se ha instanciado una cabecera identificada como C7.

Cabe mencionar que el sentido de esta pregunta es *cruzar correos* entre dos cuentas, considerando que desde cualquiera de ellas se pueden haber enviado o recibido correos, por esa razón, la consulta SPARQL incluye la función UNION, de la selección de aquellos correos que fueron enviados desde la cuenta C1 a la cuenta C2, más la selección de los correos recibidos en C1 y que provienen de la cuenta C2. Así, el código SPARQL de la consulta es el siguiente:

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?CtaEmisor1 ?CuentaReceptor1 ?asunto1 ?fechaEmision1 ?CtaEmisor2 ?CtaReceptor2 ?asunto2 ?fechaEmision2
WHERE {{ ?cuentaC1 oc:cuentaEmisorEmiteCorreo ?correo1.
?cuentaC1 oc:cuentaCorreo ?CtaEmisor1.
?cuentaC2 oc:cuentaReceptorRecibeCorreo ?correo1.
?cuentaC2 oc:cuentaCorreo ?CuentaReceptor1.
?correo1 oc:correoTieneSecuencia ?s.

```
        ?s oc:secuenciaTieneHilo ?h.
        ?h oc:hiloTieneOcurrencia ?o.
        ?o oc:ocurrenciaTieneAsunto ?a.
        ?a oc:contenidoAsunto ?asunto1.
        ?o rdf:type oc:OcurrenciaDeEmision.
        ?o oc:fechaHoraOcurrencia ?fechaEmision1.
        FILTER (?cuentaC1=oc:CUENTA_JUAN && ?cuentaC2=oc:CUENTA_ENZO).
        FILTER("2018-06-05"^^xsd:dateTime <= ?fechaEmision1 && ?fechaEmision1 <
        "2019-02-24"^^xsd:dateTime) }
    UNION {
        ?cuentaC3 oc:cuentaEmisorEmiteCorreo ?correo2.
        ?cuentaC3 oc:cuentaCorreo ?CtaEmisor2.
        ?cuentaC4 oc:cuentaReceptorRecibeCorreo ?correo2.
        ?cuentaC4 oc:cuentaCorreo ?CtaReceptor2.
        ?correo2 oc:correoTieneSecuencia ?s2.
        ?s2 oc:secuenciaTieneHilo ?h2.
        ?h2 oc:hiloTieneOcurrencia ?o2.
        ?o2 oc:ocurrenciaTieneAsunto ?a2.
        ?a2 oc:contenidoAsunto ?asunto2.
        ?o2 rdf:type oc:OcurrenciaDeEmision.
        ?o2 oc:fechaHoraOcurrencia ?fechaEmision2.
        FILTER (?cuentaC3=oc:CUENTA_JUAN &&  ?cuentaC4=oc:CUENTA_ENZO).
        FILTER("2018-06-05"^^xsd:dateTime <= ?fechaEmision2 && ?fechaEmision2 <
        "2019-02-24"^^xsd:dateTime) }}
```

La Figura 3-67 muestra los resultados de la consulta realizada, indicando que son 2 los correos intercambiados por ambas cuentas, y señala para cada una la fecha de emisión y el asunto.

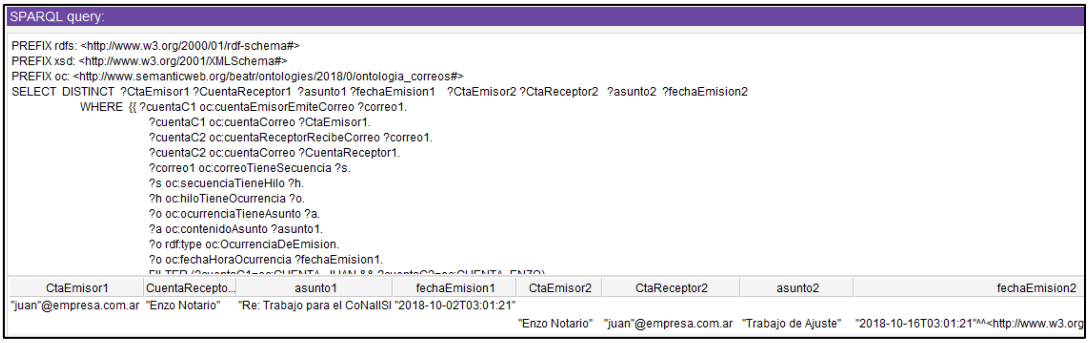


Figura 3-67: Visor de Consulta SPARQL para la PREGUNTA DE COMPETENCIA PC21

3.6 Conclusiones del Capítulo

Con lo hasta aquí descripto, se observa que OntoFoCE permite responder al criterio de autenticidad del correo electrónico que dice:

> *Un correo electrónico es auténtico cuando se identifican los datos del remitente (cuenta de correo y dirección IP), la trazabilidad del mismo (diferentes dispositivos que intervienen en la transmisión) y los datos del destinatario (cuenta de correo y dirección IP).*

Con el ejemplo de instanciación desarrollado para el Escenario 1, se observa que se cumple con la identificación de los datos del remitente, del destinatario y de la trazabilidad del correo electrónico. Esta condición puede replicarse sin inconveniente alguno en los correos cuyas cabeceras se instanciaron para el Escenario 2 y el Escenario 3 propuestos en la sección anterior de este mismo capítulo.

Particularmente, el ejemplo trabajado en el Escenario 1, mostrando el proceso de generación de *nuevas ocurrencias,* describe uno de los principales aportes de esta tesis, cual es, la recuperación de todas las direcciones IP y hostname que pueda contener la cabecera analizada.

El uso de la ontología por parte de los peritos y de manera manual, tal como se realizó para instanciar los casos de ejemplo de cada escenario, resulta una tarea laboriosa, por esa razón, a partir del modelo ontológico desarrollado, se avanzó en la implementación del mismo en una aplicación web que pone OntoFoCE a disposición del perito para realizar el análisis forense de correos electrónicos.

En el capítulo 4 se aborda detalladamente el proceso de desarrollo de esa aplicación web, denominada *ObE Forensic*, que contempla el algoritmo que toma como entrada la cabecera del correo electrónico y realiza la instanciación en la ontología, para que luego el perito pueda interactuar con la aplicación seleccionando las preguntas de competencia que responden a los puntos periciales.

CAPÍTULO 4. ObE Forensics UNA HERRAMIENTA PARA EL ANÁLISIS FORENSE DE CORREOS ELECTRÓNICOS

4.1 Introducción

En el capítulo anterior se introdujo el modelo conceptual de OntoFoCE una ontología que permite la representación de correos electrónicos y su proceso de trasmisión. A fin de que ésta pueda ser usada por un perito para realizar el análisis forense, es necesaria la construcción de una herramienta que ponga la ontología a disposición del mismo. Teniendo en cuenta esta necesidad, se ha desarrollado la herramienta ObE Forensics (Ontology based E-mail Forensics), la cual se describe en las siguientes secciones.

Cuando los datos que figuran en la cabecera de un correo electrónico en análisis son pocos, la población de la ontología puede hacerse de manera manual, estudiando uno a uno los datos y asociándolos a los conceptos y relaciones de la ontología. Este proceso manual de instanciación no es confiable pues depende de la experticia y conocimiento del perito para asociar convenientemente los valores a los correspondientes conceptos y relaciones. En respuesta a esto, la ontología detallada en el capítulo anterior, OntoFoCE, se utiliza como base de una aplicación web denominada ObE Forensic, que actúa como interface de comunicación para la realización del análisis forense a partir de la carga de la cabecera del correo electrónico por parte del perito informático.

Esta aplicación web brinda soporte para: i) la carga de la cabecera del correo electrónico y de los datos complementarios necesarios para el análisis pericial, ii) la instanciación de los datos obtenidos de las cabeceras como conceptos y relaciones en la ontología y iii) la generación de respuestas a los puntos de pericia mediante las preguntas de competencia de la ontología.

El trabajo de revisión de (Blandón Andrade, 2018) describe diversas metodologías para la instanciación de las ontologías. Algunas basadas en técnicas de procesamiento de lenguaje natural, que permiten realizar el análisis, la representación y la generación de textos, a partir de un conjunto de herramientas computacionales para el procesamiento lingüístico a nivel morfológico, sintáctico y semántico. Otros métodos recurren al machine learning, con algoritmos que realizan

el reconocimiento de patrones a partir de un conjunto de datos iniciales de ejemplos, desde el cual el algoritmo aprende a identificar los datos que luego poblarán la ontología. Hay métodos que utilizan un conjunto de reglas manuales que permiten identificar porciones del texto que luego serán los conceptos y relaciones a instanciar. Otros autores también señalan métodos híbridos que surgen como combinación de las metodologías existentes y que se organizan de la mejor manera para dar una respuesta más eficiente al problema de identificar la información para la instanciación de la ontología.

En su trabajo sobre análisis de contenido automatizado (Arcila-Calderón, Barbosa-Caro, & Cabezuelo-Lorenzo, 2016) destacan los beneficios de los métodos automatizados frente al análisis tradicional de textos por medio de la lectura y revisión, ya que el uso de la tecnología permite ganar en confiabilidad pues se disminuye notablemente los sesgos producto de la interpretación del contenido por parte del analista de textos, de modo que se puede replicar los estudios más acertadamente y en diferentes escalas. Pero además señalan el riesgo de bajar el nivel de confiabilidad de los procesos automatizados, si éstos no se consideran estándares de validez que garanticen la veracidad de los resultados a obtener.

Otra cuestión a la que se debe dar la debida atención es a las características de la fuente de información, con datos estructurados, semiestructurados o no estructurados. Si el conjunto de datos está estructurado generalmente no se requiere una análisis previo para identificar el concepto o relación sobre la cual se crea la instancia en la ontología, pero cuando se abordan fuentes de información de tipo semiestructurada o no estructurada –generalmente textos e imágenes- se requiere un proceso previo de identificación de los datos para realizar la extracción que luego se utilizará para la instanciación. Y en estos casos es necesario definir cuál será la estrategia que se va a seguir para realizar un procesamiento que permita la extracción, transformación y carga de datos en la ontología, ya sea mediante métodos automáticos o semiautomáticos. En estos casos es posible generar algoritmos de análisis de textos utilizando técnicas de map-reduce para distribuir las tareas de análisis en diferentes nodos (Map) y luego juntar los resultados en un único archivo (Reduce). También se puede recurrir a los métodos ETL (Extracción, Transformación y Carga) para la construcción de almacenes de datos. Existen estrategias y esquemas de trabajo para realizar la selección e identificación de las porciones de texto que luego se instanciarán en la ontología.

Tomando en cuenta estas consideraciones, se planteó un método propio para la extracción, transformación y carga de las cabeceras de correos electrónicos en OntoFoCE, a partir de los conceptos vistos en las diversas técnicas de instanciación estudiadas. Así, el proceso ETL trabajado en este subcomponente se basa en los siguientes criterios:

- Se requiere del análisis morfoléxico de la cabecera del correo, aunque no desde el punto de vista del *significado* de las palabras, sino más bien de la ubicación en el texto respecto de la estructura de la cabecera señalada por la norma RFC 822;
- Es importante dotar al procedimiento de la máxima automatización, en virtud del volumen de datos que puedan considerarse durante el análisis forense de las cabeceras de una cuenta de correo electrónico.

Este capítulo se organizó de la siguiente manera: en la sección 4.2 se describe el uso de ObE Forensics durante el proceso pericial; en la sección 4.3 se aborda detalladamente las funcionalidades de la aplicación describiendo los distintos subprocesos que están involucrados en el análisis forense de correos electrónicos; mientras que en la sección 4.4 se detallan las tecnologías utilizadas para la implementación de ObE Forensics en una aplicación web; y por último la sección 4.5 muestra el funcionamiento de la aplicación a partir de los tres casos ejemplos utilizado en el Capítulo 3 para explicar el modelo lógico de OntoFoCE, y se agrega un cuarto escenario para mostrar las ventajas de procesamiento de múltiples correos. Por último, en la sección 4.6 se describen las conclusiones de este capítulo.

4.2 Uso de ObE Forensics durante el Procedimiento Pericial

El procedimiento de realización de la pericia incluye una serie de pasos a ejecutar para lograr resultados exitosos que se detallaron exhaustivamente en la sección 2.5 del capítulo 2, se pueden sintetizar indicando las fases que incluye:

- ***Fase de Relevamiento*** incluye las actividades referidas a la toma de conocimiento del caso, identificación del correo electrónico y de los puntos de pericia solicitados.
- ***Fase de Recolección*** comprende las acciones necesarias tendientes a lograr el acceso al equipo receptor del correo a peritar, así como al cliente de correo utilizado.

- ***Fase de Preparación*** involucra las actividades técnicas en las que se prepara el ambiente de trabajo del informático forense, para el caso del correo electrónico, se organiza el acceso a un espacio de trabajo adecuado, considerando la ubicación física del equipo receptor.
- ***Fase de Extracción y Análisis*** comprende las tareas forenses de extracción de la cabecera de los correos a peritar, con las correspondientes acciones de resguardo de la prueba y de control de acceso solo a lo autorizado por el Juez.
- ***Fase de Presentación*** consiste en el armado de los informes necesarios y la presentación del caso ante el Juez.

Durante la FASE DE EXTRACCIÓN Y ANÁLISIS es cuando se utiliza ObE Forensics, tomando como insumo la cabecera del correo electrónico más los datos complementarios relevados. A partir de este conjunto de información, la aplicación busca e identifica en la cabecera del correo en análisis, el conjunto de datos que luego permitirán la instanciación de los conceptos, según su representación en OntoFoCE, permitiendo con ello dar respuesta a las preguntas de competencia. Es aquí en donde se observan las ventajas de esta herramienta respecto de otras propuestas para el análisis forense de correos electrónicos, ya que automatiza el análisis de la cabecera del correo y en función de la instanciación realizada sobre la ontología, permite responder a los puntos de pericia mediante los resultados de las preguntas de competencia. Por último, ObE Forensics permite emitir un informe impreso sobre los resultados del análisis forense que el perito adjunta a su Informe Pericial en la FASE DE PRESENTACIÓN DE RESULTADOS.

Como se indicó en el Capítulo 2 existen herramientas para el análisis forense de correos electrónicos, tales como *Aid4Mail*, *EmailTrackerPro*, *MailNavigator*, *OSForensics*, *E-mail Examiner*, *MailXaminer*, *IEFAF* y otras más robustas como *EnCase Forensic* que trabajan con diferentes formatos de evidencia digital, o las orientadas específicamente a teléfonos celulares como *TULP2G*, *MOBILedit FORENSIC* y *Elcomsoft Phone Breaker*.

La mayoría de ellas cumplen con diversas características de versatilidad pero además presentan problemas relacionados con la falta de orientación a la evidencia, posibilidades de tratamiento selectivo de los datos recabados, problemas estructurales que las hacen rígidas para procesar datos reservados o sensibles, o porque presentan pocas posibilidades de trabajar la información haciendo foco en un individuo. Atendiendo a estos inconvenientes, se presenta ObE Forensics como una

propuesta superadora, una aplicación que tiene como principal componente la ontología OntoFoCE diseñada expresamente para el análisis forense de correos electrónicos.

ObE Forensics es una herramienta de análisis forense de correos electrónicos desarrollada para cumplir con los criterios definidos por (Garfinkel, 2010) acerca de las herramientas forenses:

- El diseño de ObE Forensics está orientado a la evidencia, particularmente a la obtenida en casos en donde dicha evidencia es un correo electrónico.
- El modelo de visibilidad, filtro e informe de ObE Forensics le permite al experto forense establecer vínculos o relaciones de prioridad entre los datos encontrados a través de una interface de comunicación sencilla.
- ObE Forensics es una aplicación dedicada, es decir, diseñada y desarrollada específicamente para el análisis forense de correos electrónicos, y responde a las necesidades propias de este tipo de evidencia digital.
- La arquitectura de procesamiento de ObE Forensics está conformada por herramientas no propietarias, y su diseño se ajusta a criterios de abstracción y modularidad que garantizan la mejora del algoritmo inicial a partir de las sucesivas pruebas realizadas con el prototipo de la aplicación puesta a disposición de usuarios expertos.
- ObE Forensics considera todos los datos necesarios para enfocar el análisis en la identidad del individuo, al tomar los valores referidos a nombres de cuentas, identificación de equipos que pueden asociarse al usuario participante del correo electrónico analizado.

Si bien es posible encontrar herramientas disponibles para el análisis forense de correos electrónicos que permiten procesar un conjunto de cabeceras de correos electrónicos, la mayoría de ellas se agotan en mostrar los datos de la cabecera. Estas herramientas permiten responder puntos de pericia simples (como por ejemplo cuales son los datos de emisión/recepción del correo), dejando a consideración del perito la respuesta a los puntos periciales complejos como por ejemplo, establecer la trazabilidad del correo, identificar correos enviados y recibidos en rangos de fechas, o buscar información de correos asociadas entre un conjunto de cuentas de correo, entre otros.

Una particularidad de la aplicación propuesta es que, si bien en ella se han considerado 46 puntos de pericia diferentes (los surgidos del relevamiento mostrado en el ANEXO II: PUNTOS DE PERICIA), es posible incorporar a la aplicación nuevos puntos de pericia que pudieran surgir con posterioridad. Para dar respuesta a estos nuevos puntos de pericia, sólo es necesaria la definición de una o más nuevas preguntas de competencia, su formalización en una consulta SPARQL e implementación en *ObE Forensics*. Se espera una definición e implementación rápida de tal consulta dado que el modelo conceptual de OntoFoCE incluye todos los conceptos para representar un correo electrónico y su trazabilidad, dando un soporte al análisis forense de correos electrónicos. Aun así, si fuera necesario incorporar nuevos conceptos, la actualización de la ontología no supondría un gran esfuerzo debido a la capacidad de representación que presentan estas herramientas.

4.3 Funcionalidades de ObE Forensics

Se pone a disposición del Perito la ontología presentada en el capítulo anterior, dándole soporte para obtener las respuestas a los puntos de pericia, a través de tres actividades: i) la generación de instancias de la ontología OntoFoCE para representar el o los correos sobre los cuales debe realizarse el análisis forense, ii) la generación de instancias a partir de los datos complementarios cargados y iii) la obtención de respuestas a los puntos de pericia mediante las preguntas de competencia de la ontología.

El diagrama de casos de uso de la Figura 4-1 muestra las funcionalidades que ofrece ObE Forensics al perito forense, las cuales se listan a continuación:

- Ingresar Archivo de Texto Plano: que consiste en seleccionar la cabecera de uno o varios correos electrónicos a través de una ventana de selección de directorios de trabajo y archivos y cargarlos en ObE Forensics.
- Ingresar Datos Complementarios: aquí se pueden ingresar datos referidos al caso en análisis, tales como datos de identificación de los equipos utilizados en el proceso de transmisión, gestores de correos utilizados e información de identificación sobre el expediente judicial.
- Instanciar los datos en OntoFoCE: este proceso es transparente al usuario, y se explica en detalle en las secciones siguientes.

- Responder las preguntas de competencia sobre un único correo: con esta opción el perito puede acceder a las preguntas de competencia relacionadas con una única cabecera.
- Responder las preguntas de competencia sobre un conjunto de correos: se accede por esta opción al resto de las preguntas de competencia, para procesar los datos que vinculan 2 o más correos en el análisis forense.
- Visualizar e imprimir los datos procesados: le permite al perito emitir un documento impreso como anexo técnico en la respuesta que el perito debe entregar al Juez.
- Limpiar el formulario de ingreso: esta opción está disponible para el perito si es que desea procesar un nuevo conjunto de cabeceras.

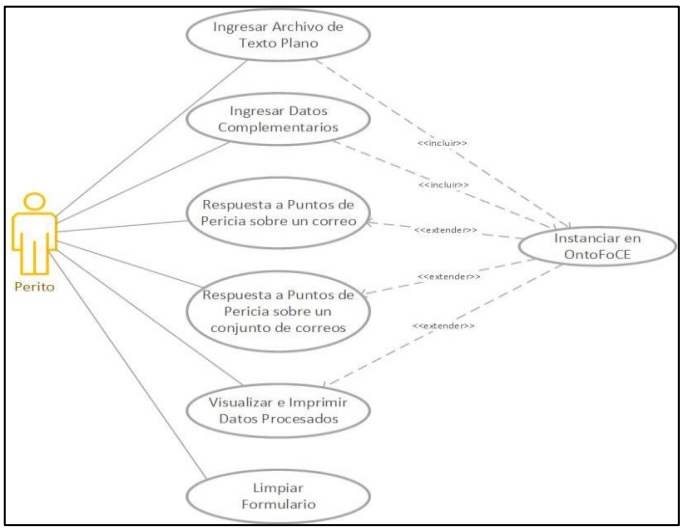


Figura 4-1: Diagrama de Casos de Uso de ObE Forensics

Para brindar estas funcionalidades, la aplicación integra en un mismo ambiente cuatro componentes, que se muestran en la Figura 4-2: a) el Gestor de Instancias de la Ontología, responsable de crear las instancias; b) el Analizador de Puntos de Pericia, responsable de mostrar las respuestas de las preguntas de competencia; c) la ontología OntoFoCE que representa el dominio de conocimiento para el análisis forense de correos electrónicos; y d) el servicio *SPARQL Endpoint*[51] que permite almacenar las instancias y realizar consultas sobre las mismas.

[51] SPARQL Endpoint es un punto de presencia en una red HTTP que es capaz de recibir y procesar solicitudes de consulta SPARQL.

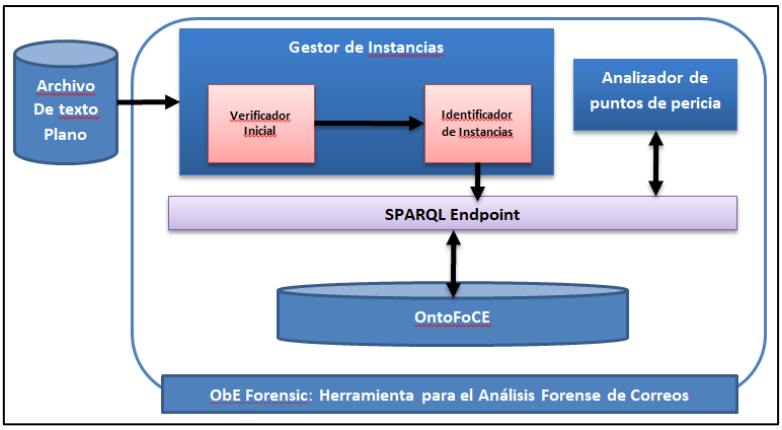


Figura 4-2: Principales componentes de ObE Forensics

A modo de breve explicación de la Figura 4-2, puede decirse que ObE Forensics toma el archivo de texto plano que conforma la evidencia digital de la cabecera de uno o un conjunto de correos que se ingresan mediante una interface diseñada a tal fin. Si bien ObE Forensics toma como insumo la cabecera del correo electrónico para realizar el análisis forense, en realidad, el archivo de texto plano que contiene la cabecera del correo también contiene el *cuerpo del mensaje*. Sobre la cabecera en sí se analiza el proceso de transmisión del correo, por esa razón, es necesario *separar* el archivo de texto plano identificando con claridad la cabecera y el cuerpo del correo electrónico en análisis.

El *Gestor de Instancias* se encarga de crear las instancias de las clases de OntoFoCE. Para lograr ese cometido, las funcionalidades de este proceso se distribuyen en dos subprocesos más específicos: el *Verificador Inicial* y el *Identificador de Instancias.* El primero de ellos es responsable de recorrer el texto del archivo plano que contiene la evidencia digital a analizar, y separar las dos partes principales: cabecera y cuerpo del correo, haciéndose cargo también de verificar que la cabecera obtenida sea factible de peritar, es decir, cumpla con los requisitos mínimos para que un correo electrónico pueda ser analizado. La función del *Identificador de Instancias* es la de definir las instancias de las clases de la ontología que representan los valores resultantes del proceso anterior y almacenar las instancias en una base de tripletas TDB[52] de Apache Fuseki[53], sobre la que se ejecutan las consultas a través del SPARQL Endpoint para dar respuesta a las preguntas de competencia. Cabe mencionar que el almacenamiento de las tripletas es

[52] https://jena.apache.org/documentation/tdb/

[53] Apache Fuseki es un servidor SPARQL que proporciona una API sobre HTTP, permitiendo realizar inserciones y consultas de manera fácil y desde cualquier lenguaje de programación (se puede consultar en la siguiente página. https://jena.apache.org/documentation/fuseki2/).

temporal; es decir, atendiendo a cuestiones de reserva y privacidad de los datos que se procesan (datos que conforman evidencia digital en causas judiciales), el almacenamiento de las tripletas se mantiene mientras el usuario tenga la sesión habilitada, en cuando se desconecta de la aplicación, se realiza un proceso de borrado total de la base de tripletas TDB utilizada.

En el Capítulo 3 se describió detalladamente OntoFoCE. A continuación, se describe el proceso que se realizar para ejecutar cada una de las tres funcionalidades indicadas para ObE Forensics: i) Generación de instancias a partir de los archivos de texto plano de las cabeceras de los correos electrónicos, ii) Generación de instancias a partir de los datos complementarios cargados y iii) Respuestas a los puntos de pericia.

4.3.1 Generación de Instancias a partir de las Cabeceras de Correos Electrónicos

El Gestor de Instancias es el componente responsable de crear las instancias a partir de los archivos de texto plano que ingresa el Perito. Para ilustrar este proceso se recurre al diagrama de actividades indicado en la Figura 4-3. En dicha figura se observa que el Perito inicia el proceso con el ingreso de los archivos de texto plano que corresponden a las cabeceras de los correos a analizar, dando inicio al procesamiento de los datos, que comienza con el *Gestor de Instancias*, este componente consta de dos módulos: el *Verificador Inicial*, que comprende las primeras actividades (separación de la cabecera y cuerpo del texto plano y verificación de la validez de la cabecera) y el *Identificador de Instancias* que lleva adelante la extracción de los atributos, las transformaciones necesarias y la creación de las instancias de clases de OntoFoCE que los representan.

A continuación se describe cada uno de estos módulos.

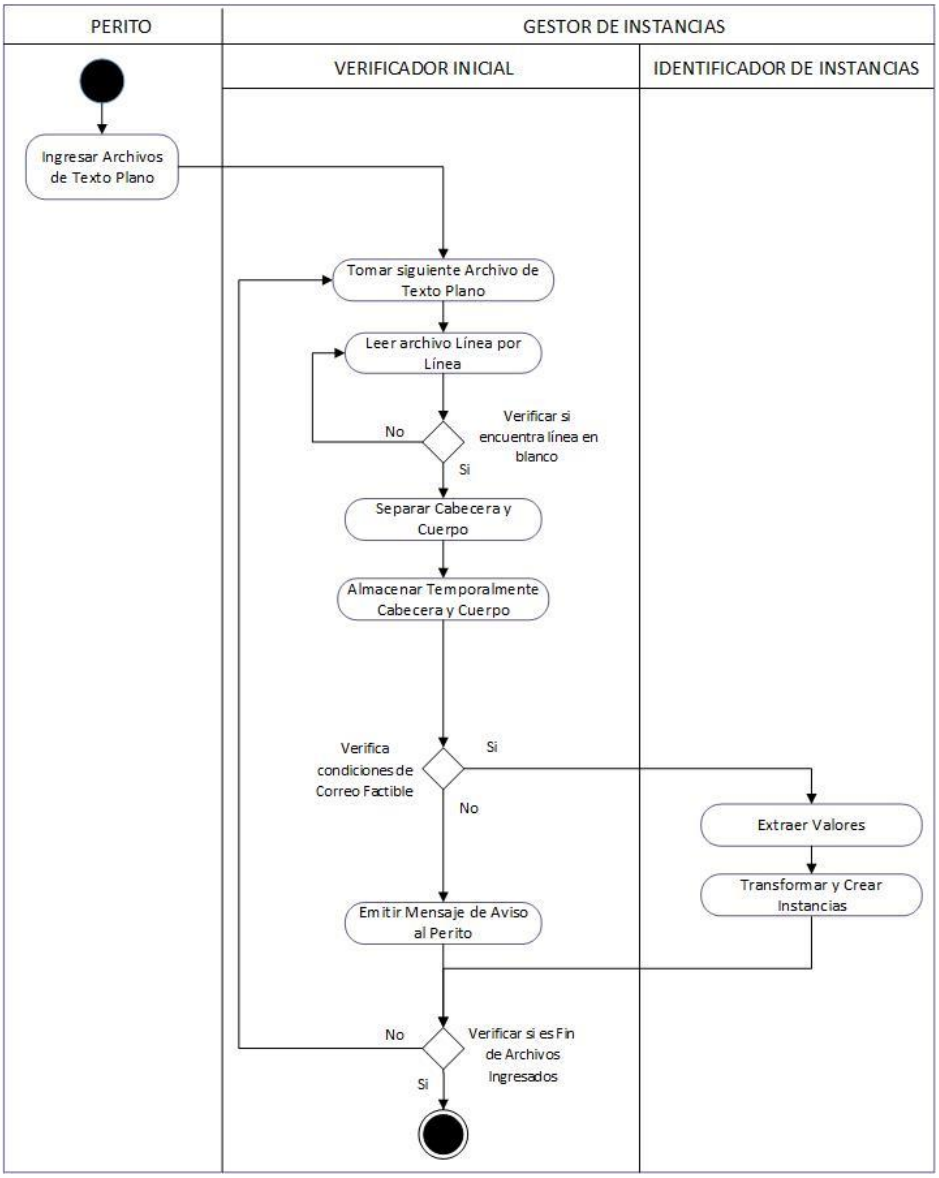


Figura 4-3: Diagrama de Actividades de GENERACIÓN DE INSTANCIAS A PARTIR DE LA CABECERA

Verificador Inicial

Tal como se describe en la Figura 4-3, una vez cargados los archivos de texto plano, se inicia el procesamiento individual de cada archivo mediante el *Verificador Inicial*. Este módulo de ObE Forensics, tiene como primera función la identificación de la parte del archivo de texto plano ingresado que corresponde a la cabecera del correo, separándolo del cuerpo.

Según la norma RFC 822, la cabecera y el cuerpo están separados por una línea en blanco. Teniendo en cuenta esta premisa, el proceso recorre el contenido del correo línea por línea desde el inicio hasta encontrar la primera línea en blanco. Cuando la encuentra almacena temporalmente en memoria todo lo anterior como la

cabecera del correo y a partir de esa posición y hasta el final del archivo, se considera ese texto como el *cuerpo del correo,* y también es almacenado temporalmente.

La segunda función del *Verificador Inicial* es comprobar si la cabecera cumple con los requisitos suficientes para realizar el análisis forense. Por ello cada cabecera se valida y se descartan aquellos que no cumplen con al menos una de las siguientes condiciones:

- Se debe contar con la cabecera del correo a analizar
- Se debe identificar en ella la dirección IP del equipo emisor y
- Se debe identificar en ella la dirección IP del equipo receptor

Estas tres condiciones dan origen a la regla "*Un correo es factible de peritar cuando tiene Cabecera, y en ella figuran la IP del Equipo Emisor y la IP del Equipo Receptor"*, expresada en la sección 3.4 del capítulo 3.

Para cumplir con esta condición de factibilidad para la realización de la pericia de un correo electrónico, el proceso de *Verificador Inicial* realiza la búsqueda de la dirección IP correspondiente al equipo emisor, si se verifica que existe, se realiza la búsqueda de la dirección IP del equipo receptor. En caso de no encontrar alguna de estas dos direcciones IP, el proceso emite un mensaje de aviso dirigido al perito. Cabe mencionar que este mensaje de advertencia se muestra al momento de seleccionar el correo y visualizar los datos de la cabecera, de modo que el proceso de verificación avanza con el siguiente archivo de texto plano, hasta terminar el lote ingresado.

Identificador de Instancias

Cuando la cabecera es factible de peritar el procesamiento queda a cargo del *Identificador de Instancias,* éste funciona como un proceso ETL (Extracción, Transformación y Carga) realizando una serie de pasos necesarios para cargar la ontología con los datos provenientes de la cabecera del correo electrónico:

a) Extracción de los datos que están representados como conceptos en la ontología,

b) Generación de las Instancias en las clases de OntoFoCE, realizando procesos de transformación de los datos cuando corresponda.

En la siguiente sección se describe el módulo correspondiente al *Identificador de Instancias*.

4.3.1.1 Identificador de Instancias

Este proceso se inicia con la identificación de las diversas partes del correo que figuran en la cabecera en análisis y consta de dos módulos que se describen a continuación.

Extraer Valores

En el proceso anterior, referido como *Verificador Inicial*, se realizó la separación de la cabecera y cuerpo del correo y el almacenamiento temporal en memoria. El módulo *Extraer Valores* toma el archivo de la cabecera así obtenida, y vuelve a realizar una lectura línea por línea para buscar los datos que son de interés para crear las instancias de OntoFoCE (ver Figura 4-4).

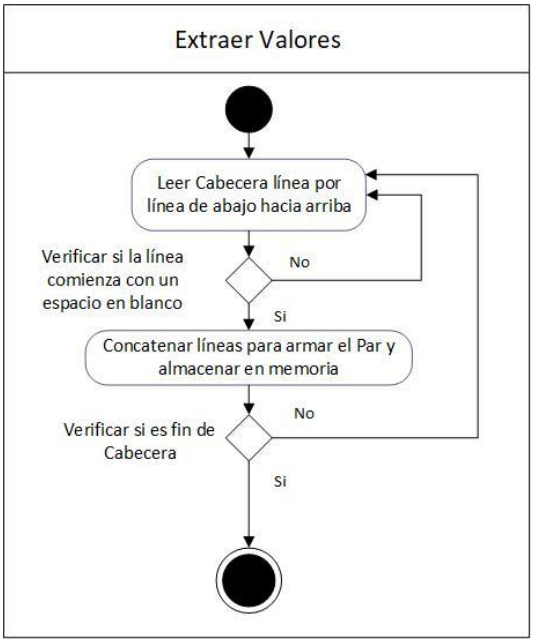


Figura 4-4: Módulo EXTRAER VALORES

La norma RFC 822 señala que cada línea de la cabecera del correo consta de dos partes: un *Parámetro* que se encuentra al principio de la línea, y un *Valor,* que corresponde al resto del texto de esa línea, separando ambos elementos con el signo de ":". El formato es el siguiente:*{Parámetro}:{Valor}.*

A veces el par *{Parámetro}:{Valor}* se encuentra presente en múltiples líneas de la cabecera, entonces es necesario identificar las líneas que en realidad son continuación de una anterior. Durante las pruebas realizadas se pudo identificar que aquellas líneas que son una continuación de la anterior, comienzan con al menos un espacio en blanco, por lo tanto: son concatenadas al valor de la línea anterior. Así,

las líneas se van concatenando hasta que se encuentra una línea que no comienza con un espacio en blanco, indicando con ello que se está en presencia de un atributo *{Parámetro}:{Valor}* distinto al que se está concatenando, esta situación se ilustra en la Figura 4-5 en la que se muestran todas las líneas identificadas en la cabecera que ilustra el caso de estudio del Escenario 1.

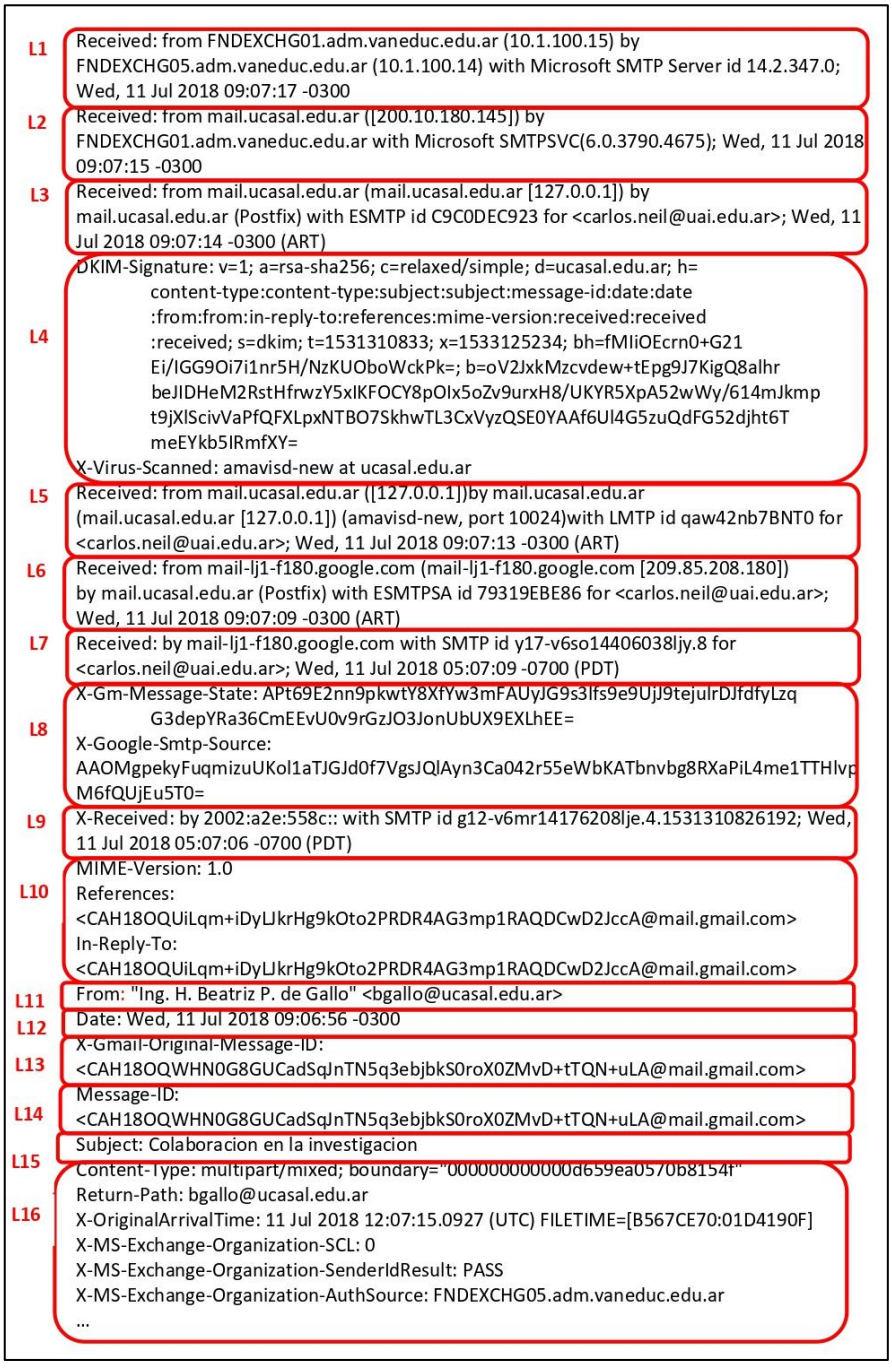


Figura 4-5: Líneas {Parámetro}:{Valor} identificadas en una cabecera

En la Figura 4-5 cada línea que se identifica como de interés para los conceptos representados en OntoFoCE se almacena temporalmente en memoria. Este proceso

de identificación y concatenación de líneas se realiza iterativamente hasta recorrer toda la cabecera. Considerando el correo ejemplo utilizado en el escenario 1 del Capítulo 3, la Figura 3-5 indica las líneas *{Parámetro}:{Valor}* identificadas por el proceso *Extraer Valores.*

Cabe mencionar que en la cabecera pueden figurar pares *{Parámetro}:{Valor}* que no son de interés para el análisis forense, como por ejemplo, las líneas identificadas en la Figura 4-5 como L4, L8, L10 y L13 que referencian parámetros internos y de seguridad del proceso de transmisión. Previamente, el módulo de *Verificación Inicial* identificó las líneas L1 a L15 como el texto correspondiente a la *cabecera* del correo propiamente dicha mientras que la línea L16 es el *cuerpo* del mismo.

Para relacionar el ejemplo con los parámetros que se van identificando en los módulos siguientes, se acuerda indicar el parámetro y a continuación entre paréntesis el número de línea de la Figura 4-5 que le corresponde, por ejemplo: *Subject (L15).*

Transformar Valores y Generar Instancias

Hay conceptos que se toman sin realizar ningún procesamiento ni transformación de los datos, se trata de aquellos que usualmente son datos de carácter descriptivo.

En estos casos, se genera la instancia de la clase correspondiente de manera directa. Tal es el caso de: *Asunto, Correo y CuentaEmisor* para los que se crean directamente las instancias para las clases respectivas a partir de los valores de los parámetros *Subject(L15), Message-ID(L14) y From(L11)* respectivamente, según se indica en la Figura 4-6. Cabe aclarar que, por cada valor que se asignará como atributo, se crea previamente la instancia de clase correspondiente, así como las instancias de las distintas relaciones requeridas por OntoFoCE para dichas clases.

En el caso de la creación de la instancia correspondiente a la clase ***CuentaEmisor*** a partir de los datos del parámetro *From,* es posible que dicho parámetro no contenga información sobre el *aliasUsuario y/o firmaUsuario*. Que esté o no presente dicha información depende de la configuración realizada por el usuario en su cuenta de correo, y en ese caso, estos atributos contendrán valores nulos.

Obsérvese que, en este módulo, se aprovecha la recorrida completa que se realiza sobre las líneas *{Parámetro}:{Valor}* que se encuentran almacenadas temporalmente para realizar otros procesos de análisis y preparación de los datos.

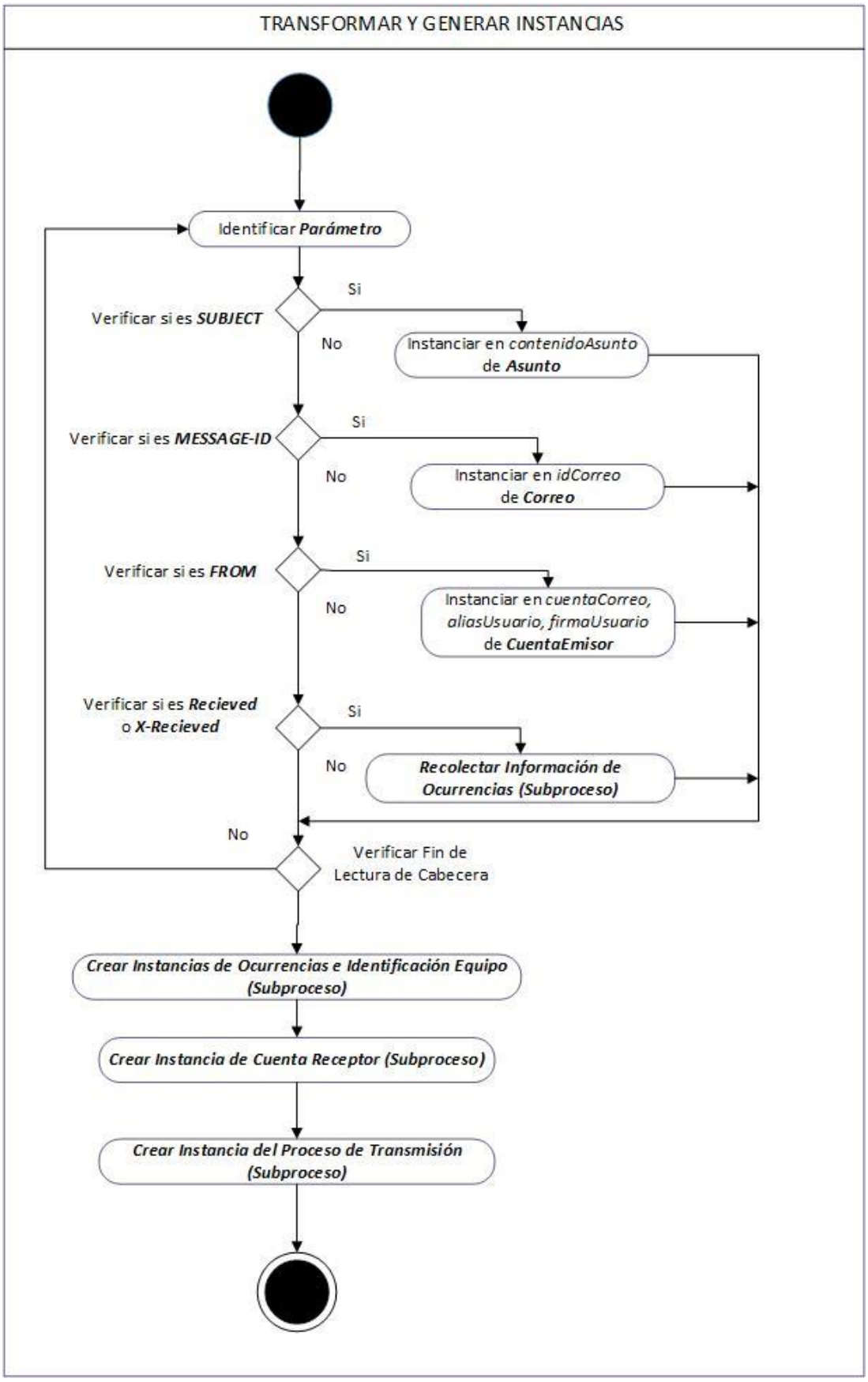


Figura 4-6: Diagrama de Actividades del Módulo TRANSFORMAR VALORES Y GENERAR INSTANCIAS

En el diagrama de la Figura 4-6 se señalan cuatro subprocesos que es importante explicar con mayor detalle:

- *Recolectar Información de Ocurrencias*: en el que se establece el tipo de cada ocurrencia y la vinculación entre ellas mediante los sub-atributos *From y By* de la línea *Received/X-Received* de la cabecera que se esté analizando.
- *Crear Instancias de Ocurrencias e Identificación de Equipos:* se ejecuta luego de finalizar el recorrido de las líneas del proceso anterior tomando los valores identificados y almacenados en ese proceso.
- *Crear Instancia de Cuenta Receptor:* para generar la instancia del concepto *CuentaReceptor* que requiere de un análisis propio.
- *Crear Instancia del Proceso de Transmisión:* aquí se generan las instancias correspondientes a las clases *Secuencia e Hilo* para representar el proceso de transmisión.

A continuación se detallan estos subprocesos, considerando la descripción del procedimiento y la referenciación al par *{Parámetro}:{Valor}* de las líneas correspondientes.

Recolectar Información de Ocurrencias

Este subproceso analiza las líneas cuyos parámetros son *Received/X-Received* y se encarga de identificar los tipos de ocurrencia, es decir, si es una ocurrencia de emisión, de transmisión o de recepción.

Además, identifica los valores que se toman de cada línea de los parámetros *Received (L1, L2, L3, L5, L6 Y L7) o X-Received (L9)*, o sea la fecha y hora en que el correo se almacena en el equipo/servidor, así como la identificación del mismo que también figura como valor de dichos parámetros.

Todos estos datos –tomados de cada línea analizada- se van almacenando en un vector de memoria, del cual luego se recuperan los valores requeridos para crear las instancias que representan las ocurrencias, según se describe en los siguientes apartados. La Figura 4-7 muestra el diagrama de actividades correspondiente.

Como se explicó en la sección 3.5.1 del capítulo 3, cuando se abordó la descripción del proceso de instanciación manual de las ocurrencias para el caso de estudio del Escenario 1, la actividad se inicia cuando el proceso de nivel superior encuentra el parámetro *Received/X-Received* que, a su vez, contienen los tres sub-

atributos que contienen los valores para identificar la ocurrencia (*from, by* y *timestamp).*

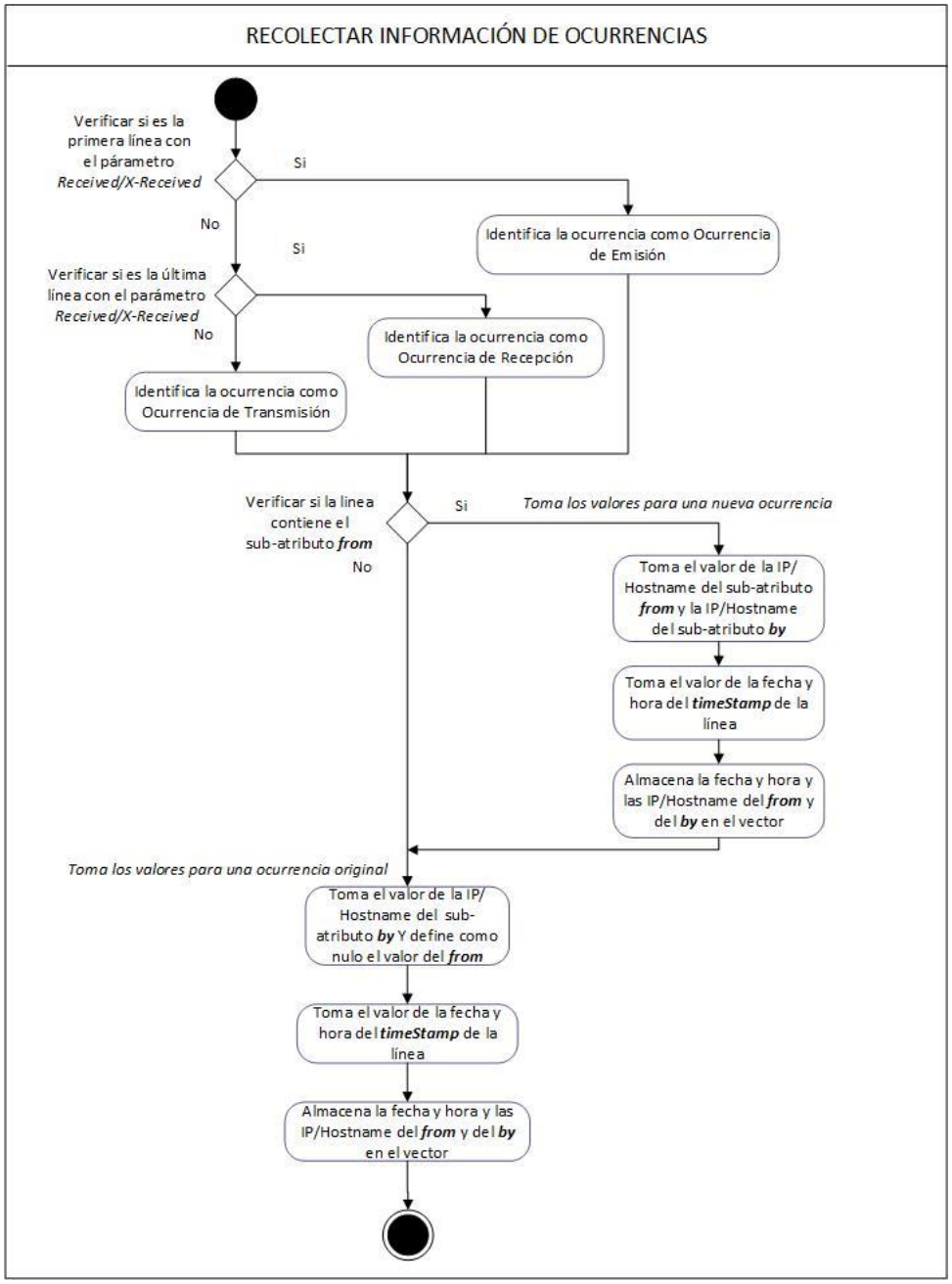


Figura 4-7: Diagrama de Actividades del Subproceso
RECOLECTAR INFORMACIÓN DE OCURRENCIAS

Teniendo presente lo dicho respecto a la necesidad de la presencia o no de los sub-atributos *from* y *by* en la línea del *Received/X-Received,* se analiza la cabecera considerando que:

- La cabecera se lee de abajo hacia arriba.
- Se toma la primera línea *Received/X-Received* que se encuentra, y se verifica si contiene el sub-atributo *from,* en ese caso, el valor de la dirección IP/Hostname

que contiene identifica el equipo emisor, de lo contrario la identificación se obtiene del valor contenido en el sub-atributo *by* de esa primera línea.

- Se toman los valores de los sub-atributos *from/by* de las siguientes líneas *Received/X-Received* para identificar los servidores de paso, salvo la última línea, que identificará el equipo receptor.
- De acuerdo a la norma RFC 822, la identificación de los equipos/servidores que participan del proceso de transmisión se encuentra como valor del sub-atributo *by*, de manera que si sólo se tomaran esos valores, se podría identificar a todos los equipos/servidores que participaron durante la transmisión del correo.
- La citada norma indica además que cuando el parámetro *from* está presente al inicio de la línea *Received/X-Received* es que está informando la dirección IP/Hostname de la línea anterior, pero de las pruebas realizadas, se ha encontrado que algunos gestores de correo incluyen direcciones IP al final de la sarta de caracteres que conforman el valor del sub-atributo *from,* enmascarando la dirección IP entre corchetes y paréntesis.
- Entonces, a fin de no descartar ninguna dirección IP, porque no se encuentra como valor del sub-atributo *by*, se propone generar una *nueva ocurrencia* que tome dicha dirección IP como valor del sub-atributo *by*. De este modo, es posible obtener dos ocurrencias a partir de los datos encontrados en una misma línea.

Para entender el algoritmo graficado en la Figura 4-7, tómese como ejemplo las líneas B y C de la cabecera del caso de estudio considerado en el Escenario 1, apartado 3.5.1 del capítulo 3, que se aunaron en la Figura 4-8.

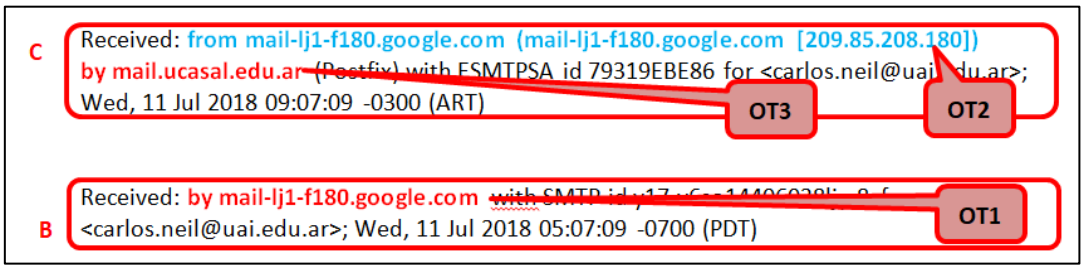


Figura 4-8: Análisis para generar una nueva ocurrencia

Allí se puede ver, que el *by* de la línea identificada como B y cuyo valor es *mail-lj1-f180.google.com*, coincide con el valor inicial del *from* de la línea C, salvo que ese valor incluye también la dirección IP "*209.85.208.180"*. Esto significa que ambos valores, el del *by* de la línea B y el del *from* de la línea C serían iguales salvo el valor *{[209.85.208.180]}*, el cual se descartaría del análisis pericial si se consideran los valores del sub-atributo *by* y se obviase el valor del sub-atributo *from.*

Para el ejemplo de la Figura 4-8, y siguiendo los pasos indicados en el algoritmo, la secuencia de acciones sería la siguiente:

1. Se lee la línea identificada como B en la Figura 4-8,
   1.1. No es la primera línea, ni es la última línea, entonces se identifica el tipo de la ocurrencia como ocurrencia de transmisión.
   1.2. La línea no contiene el sub-atributo *from* entonces se continúa con el siguiente paso.
   1.3. Se toma el valor del sub-atributo *by* y el vector almacena un elemento que contiene:
      - *by = mail-lj1-f180.google.com*
      - *from* = (en blanco)
      - *fecha y hora* = 11 Jul 2018 05:07:09-0700 (PDT)
2. Se retorna al módulo de TRANSFORMAR VALORES Y GENERAR INSTANCIAS y se lee la siguiente línea, es decir, se lee la línea C, y se verifica que:
   2.1. No es la primera línea, ni es la última línea, entonces se identifica el tipo de la ocurrencia como ocurrencia de transmisión.
   2.2. La línea contiene el sub-atributo *from* entonces el vector toma la dirección IP enmascarada al final del valor de ese sub-atributo y almacena un elemento en el vector que contiene los siguientes valores:
      - *by* = 209.85.208.180
      - *from* = (en blanco)
      - *fecha y hora* = 11 Jul 2018 09:07:09-0300 (ART)
   2.3. Se toma el valor del sub-atributo *by* y el vector almacena un elemento que contiene:
      - *by* = mail.ucasal.edu.ar
      - *from* = (en blanco)
      - *fecha y hora* = 11 Jul 2018 09:07:09-0300 (ART)
3. Se retorna al módulo de TRANSFORMAR VALORES Y GENERAR INSTANCIAS y se lee la siguiente línea.

Finalizado este algoritmo para cada línea de la cabecera en análisis, se cuenta con un *Vector de Ocurrencias* en memoria que registra tres datos: a) los valores de

las IP/Hostname de los equipos donde reside cada ocurrencia; b) los valores de fecha y hora de cada una de éstas; y c) el tipo de ocurrencia (de emisión, transmisión o recepción).

De este modo, el *Vector de Ocurrencias* contiene la información de todas las ocurrencias encontradas, ya sea las obtenidas a partir del valor del sub-atributo *by*, denominadas en el capítulo 3 como *ocurrencias originales*, o las obtenidas a partir del valor de una dirección IP de un sub-atributo *from*, que en el capítulo 3 se denominan *nuevas ocurrencias* u *ocurrencias generadas.*

Este Vector de Ocurrencias, que se obtiene luego de procesar todas las líneas de la cabecera será insumo para el siguiente subproceso del módulo denominado TRANSFORMAR VALORES Y GENERAR INSTANCIAS y que en la Figura 4-6 se identificó como *Crear Instancias de Ocurrencias e Identificación de Equipos*, el cual se describe en el siguiente apartado.

Crear Instancias de Ocurrencias e Identificación de Equipos

Este subproceso recorre el Vector de Ocurrencias que fue generado en el proceso previamente descripto y realiza dos acciones: renumera las ocurrencias considerando la inserción de las *nuevas ocurrencias* entre las *ocurrencias originales* y, crea las instancias para las clases *Ocurrencia* e *IdentificacionEquipo.* La Figura 4-9 muestra el diagrama de actividad de este subproceso.

Este subproceso toma como insumo el Vector de Ocurrencias y realiza la lectura de los elementos identificando si se trata de una ocurrencia de emisión (en el caso del primer elemento del vector), de una ocurrencia de transmisión (para los elementos intermedios del vector) o de una ocurrencia de recepción (cuando se trata del último elemento del vector). En este punto se crea una instancia de OcurrenciaDeEmision, OcurrenciaDeTransmisión u OcurrenciaDeRecepción, según el tipo de ocurrencia identificado. Esta tarea se representa en la Figura 4-9 con las actividades (*Nominar la ocurrencia como OE, Nominar la ocurrencia como OTi y Nominar la ocurrencia como OR, respectivamente).*

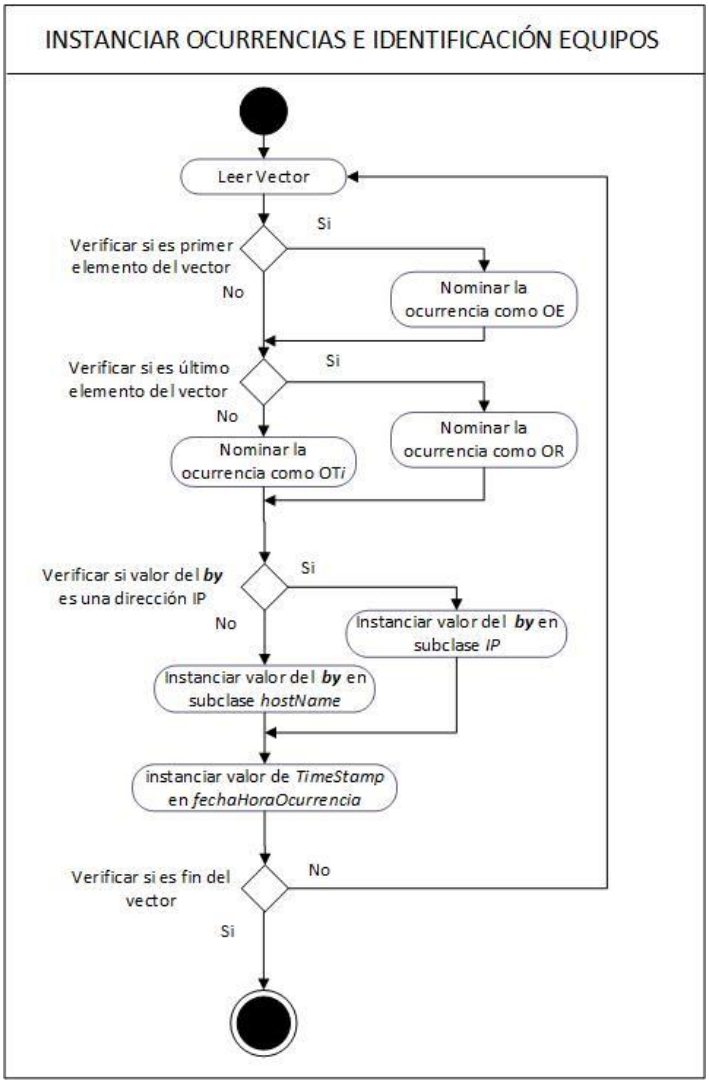


Figura 4-9: Diagrama de Actividad del Subproceso
INSTANCIAR OCURRENCIAS E IDENTIFICACIÓN DE EQUIPOS

Para la instancia creada se le asigna el *nombre* con el cual se mostrará en las pantallas de datos que visualiza el Perito, denominando *OE* si es la ocurrencia de emisión, *OR* si es la ocurrencia de recepción u *OT* si es una ocurrencia de transmisión, en este último caso, el algoritmo agrega un identificador numérico a fin de mostrar las distintas ocurrencias según el orden de aparición. Asimismo, durante este recorrido del vector, se toma los valores de fecha y hora, que son utilizados como valores del atributo *fechaHoraOcurrencia* de cada una de las instancias de las clases *OcurrenciaDeEmision, OcurrenciaDeTransmision u OcurrenciaDeRecepcion* según corresponda. De igual modo, se realiza la instanciación de las relaciones de vinculación de los distintos objetos, particularmente la relación *esAnteriorA* que se toma de la secuencia de orden con el que fueron almacenados los valores en el Vector de Ocurrencia.

Luego, se toman los restantes valores del elemento del vector para crear las instancias que representan el equipo en que se almacena la ocurrencia y la identificación del mismo. Para el equipo, se creará una instancia de las clases EquipoEmisor, Servidor o EquipoReceptor, según se trate de una ocurrencia de emisión, de transmisión o de recepción. En tanto, el valor del sub-atributo *by* del elemento del vector determinará si se crea una instancia de IP o de Hostname para identificar el equipo. El valor del *by* se almacena en el atributo *identificadorEquipo* de la instancia creada. Asimismo, se crean las instancias de las relaciones que vinculan estos objetos.

Aquí se debe aclarar que el valor del cual se genera la instancia para la clase *IdentificacionEquipo* puede encontrarse en diferentes formatos, por ejemplo:

- IPv4: 209.85.208.180, como en el caso de la línea L6 del correo ejemplo de la Figura 4-5
- IPv6: 2002:a2e:558c:: como en el caso de la línea L9 del mismo ejemplo, o
- Expresado en términos de un dominio, *mail.ucasal.edu.ar* por ejemplo.

En el mismo recorrido se toma los valores de fecha y hora para crear la instancia de *fechaHoraOcurrencia* de la clase *OcurrenciaDeEmision*, *OcurrenciaDeTransmision u OcurrenciaDeRecepcion* según corresponda.

Por último, se debe considerar que las Fechas de las Ocurrencias pueden estar presentes en distintos formatos, por ejemplo: Mon, 01 Jan 2018 10:00:00 (PDT); 2018-01-01 10:00:00 -0300; o 01/01/2018 10:00:00 (ART). Para poder identificar la fecha teniendo en cuenta todos los formatos, se hace uso de una librería de *PHP* llamada *Carbon*[54], la cual es capaz de reconocer todos los formatos y unificarlos. Además, dado el contexto del proyecto, todas las fechas son convertidas al huso UTC-3, o sea al Universal Time Coordinated de Argentina, por esa razón, aun cuando figuran horas diferentes en las ocurrencias, al momento de realizar el análisis forense, esa diferencia horaria se salva con la conversión de fechas que realiza ObE Forensics. Por ejemplo: si una ocurrencia contiene el valor “8 jul 2018 12:41:52 UTD-3” y la siguiente ocurrencia tiene el valor “8 jul 2018 08:41:58” con el huso horario PDT (Pacific Daylight Time), al momento de la conversión, todos los valores horarios se convierten al huso horario UTC-3, de modo que ambos valores se convierten a “8 jul 2018 12:41:52” y “8 jul 2018 12:41:58” respectivamente.

[54] https://carbon.nesbot.com/

Crear Instancia de Cuenta Receptor

Este es el último módulo del proceso *Transformar Valores y Generar Instancias* y consiste en identificar cuál es la cuenta que recibió el correo y crear la instancia de la clase *CuentaReceptor* que la represente así como las instancias de las relaciones propuestas por OntoFoCE que la vinculan con un correo, un equipo y un cliente de correo.

Según sea el caso deberá buscarse en distintos parámetros de la cabecera. Los ejemplos de la Figura 4-10 muestran dos posibles formas en que los sistemas gestores de correo electrónicos registran los datos de la cuenta receptora.

Ejemplo 1:
…
Received: from mail.ucasal.edu.ar (mail.ucasal.edu.ar [127.0.0.1])
by mail.ucasal.edu.ar (Postfix) with ESMTP id C9C0DEC923 for <***carlos.neil@uai.edu.ar***>;
Wed, 11 Jul 2018 09:07:14 -0300 (ART)
…

Ejemplo 2:
Delivered-To: ***luzbibianaclara@gmail.com***
Received: by 2002:a4a:d4d5:0:0:0:0:0 with…

Figura 4-10: Ejemplos de Variantes en el registro de la Cuenta Receptora

Normalmente la cuenta del receptor se encuentra en el parámetro *Delivered-To*, o también puede estar presente en el parámetro *X-Original-To*. Cuando no es posible localizar el receptor a partir de estos parámetros, se recorren las ocurrencias en busca de la primera que tenga un valor en el sub-parámetro *For* de la línea del *Received*.

Cabe mencionar que el valor que allí se encuentra, representa la cuenta de correo, no así el *alias* que el usuario puede haber configurado en su cliente de correo, el cual, de estar presente en la cabecera, figura como valor del parámetro *To.*

El proceso de búsqueda de la cuenta receptora se grafica en el diagrama de actividades de la Figura 4-11. En dicha figura se indica las búsquedas que se realiza para obtener el valor con el que luego se generará la instancia de la clase *CuentaReceptor*, y el valor para el atributo *aliasUsuario* de dicha nueva instancia.

En el ejemplo de la Figura 4-5 la instancia de la clase *CuentaReceptor* se crea a partir del parámetro *Received (L3)* y se observa que el parámetro *To*, que usualmente contiene la lista de cuentas receptoras con su alias, no se encuentra presente en esta cabecera.

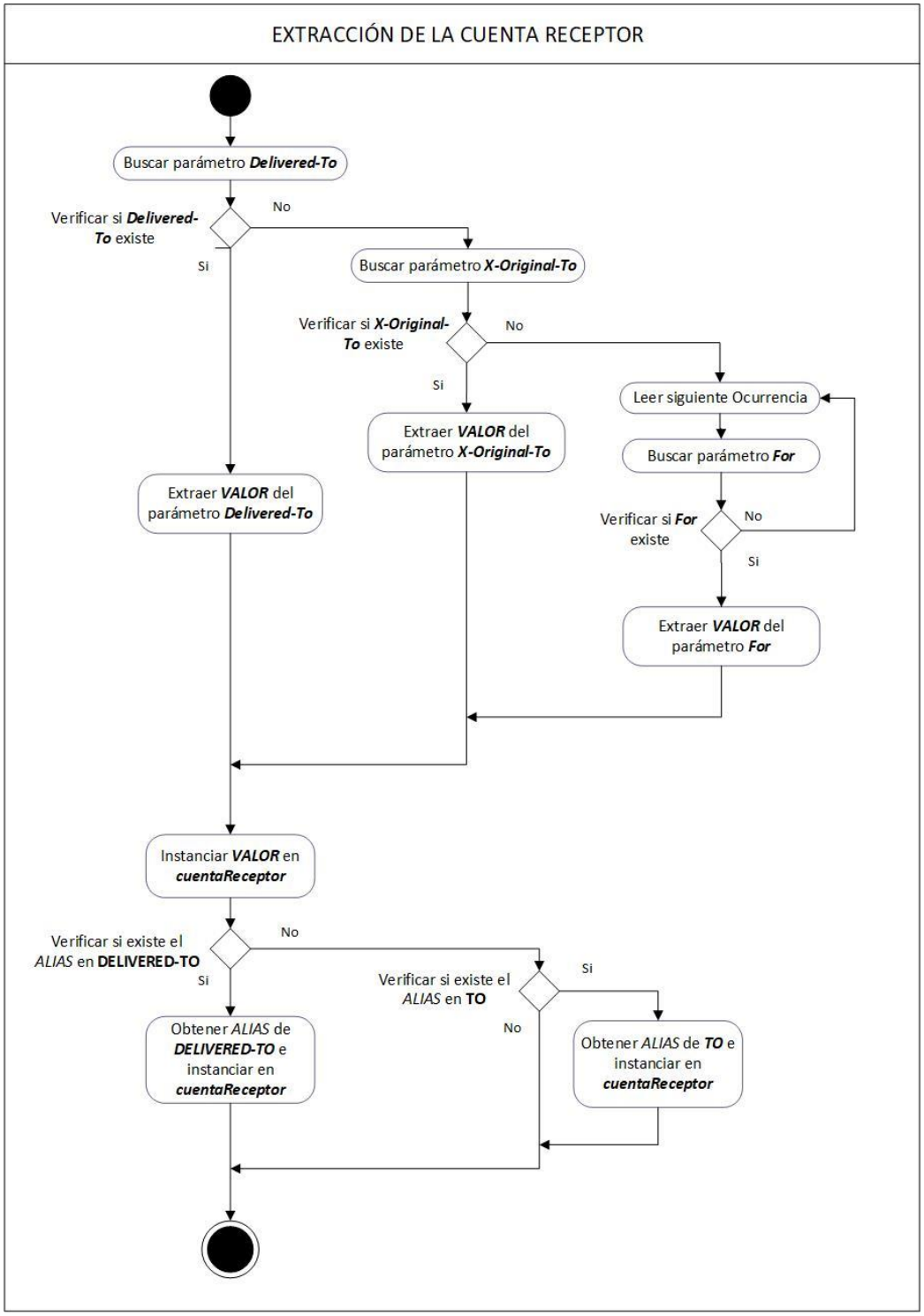


Figura 4-11: Diagrama de Actividad del Subproceso EXTRACCIÓN DE CUENTA RECEPTOR

Crear Instancia del Proceso de Transmisión

Este es el último módulo del proceso ***Identificador de Instancias***. Aquí, se generan las instancias de las clases ***Secuencia e Hilo*** que determinan la traza del proceso de transmisión a partir de la vinculación de las instancias ya creadas para las

clases ***OcurrenciaDeEmision, OcurrenciaDeTransmision*** y ***OcurrenciaDeRecepcion.***

Para ello, el proceso genera una instancia para la clase ***Secuencia*** a partir del siguiente valor autonumérico reservado para la variable de trabajo interna denominada *SecuenciaAutonum*, y genera además una instancia para la clase ***Hilo***, tomando el siguiente valor autonumérico reservado para la variable interna denominada *HiloAutonum*.

Cuando se identifica dos o más cabeceras con el mismo valor para el parámetro *Message-ID,* esto significa que las cabeceras corresponden a un mismo correo proveniente de la misma cuenta emisora, y esta consideración se tiene en cuenta para instanciar la relación *secuenciaTieneHil* y así vincular las instancias de Hilo que corresponden al mismo correo.

En este proceso también se crea la instancia de la relación *correoTieneSecuencia* que vincula un correo con su correspondiente secuencia.

4.3.2 Generación de Instancias a partir de Datos Complementarios

ObE Forensics también incluye la carga de los datos complementarios, es decir, información que se recaba al momento de realizar la pericia y que no se encuentra en la cabecera, y se requiere para contestar los puntos de pericia. La Figura 4-12 muestra el diagrama de actividades para este módulo.

Hay datos que conforman un único valor y hay otros, los referidos a los equipos y servidores, que requieren un ciclo iterativo de carga. Allí se indica el orden de verificación de datos cargados correspondientes a los valores de los atributos de las instancias oportunamente creadas de las clases ***Expediente, EquipoEmisor, EquipoReceptor*** y ***Servidor.***

Para la clase ***Expediente*** se carga un solo conjunto de valores, mientras que para las restantes la aplicación genera un ciclo de carga para tantos equipos como se hayan detectado.

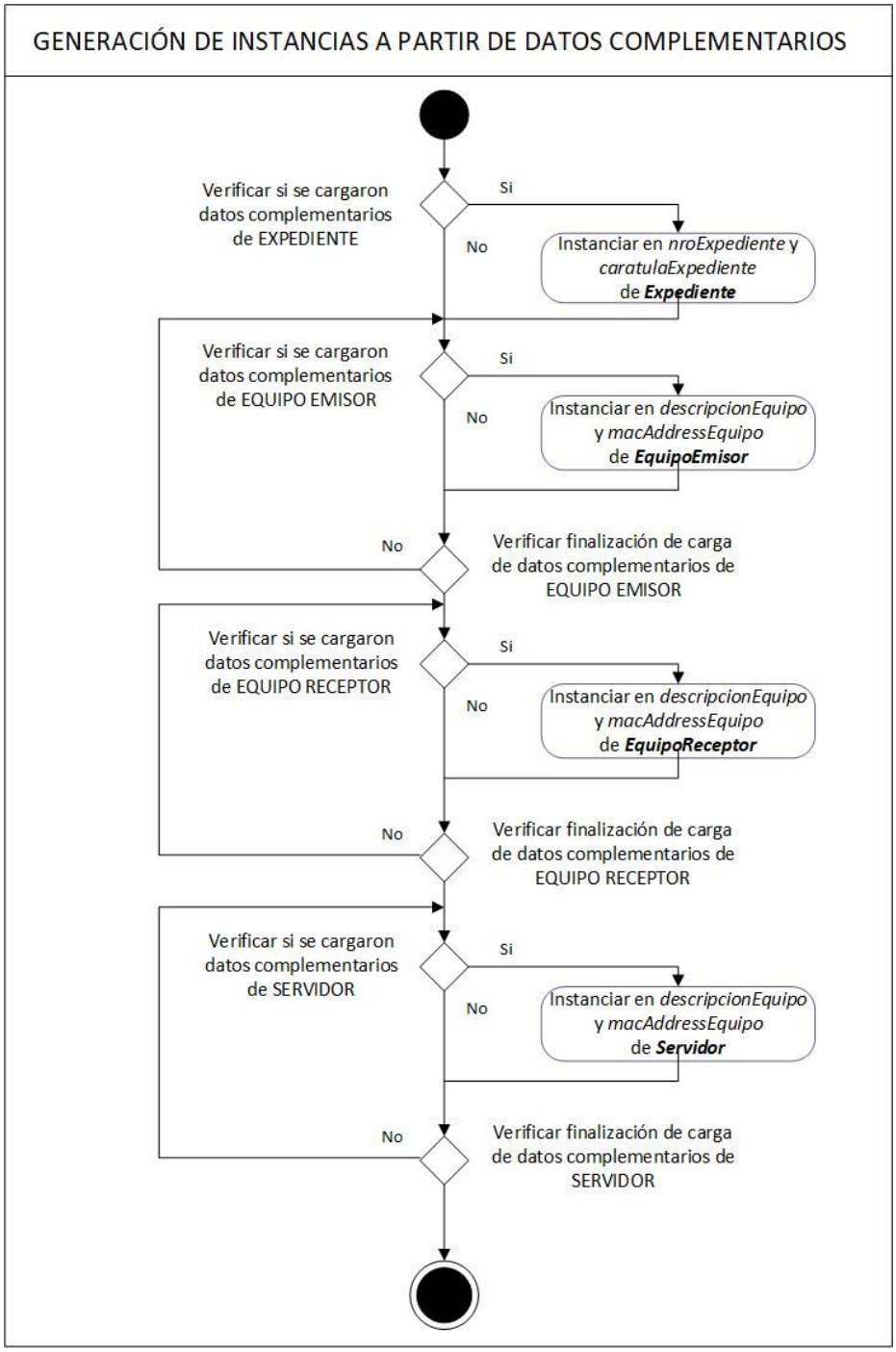


Figura 4-12: Diagrama de Actividades del Subproceso INSTANCIAR DATOS COMPLEMENTARIOS

4.4 Analizador de los Puntos de Pericia

Este componente de la herramienta de soporte para el análisis forense de los correos electrónicos tiene por objetivo mostrar los resultados a las preguntas de competencia que sean pertinentes al punto de pericia requerido.

El conjunto de preguntas de competencia a las que la ontología propuesta puede responder ha sido introducido en el Capítulo 3, junto con la formalización de las mismas mediante consultas SPARQL. El proceso que ejecuta ObE Forensics para poner a disposición del Perito las preguntas de competencia se muestra en el Diagrama de Actividades de la Figura 4-13.

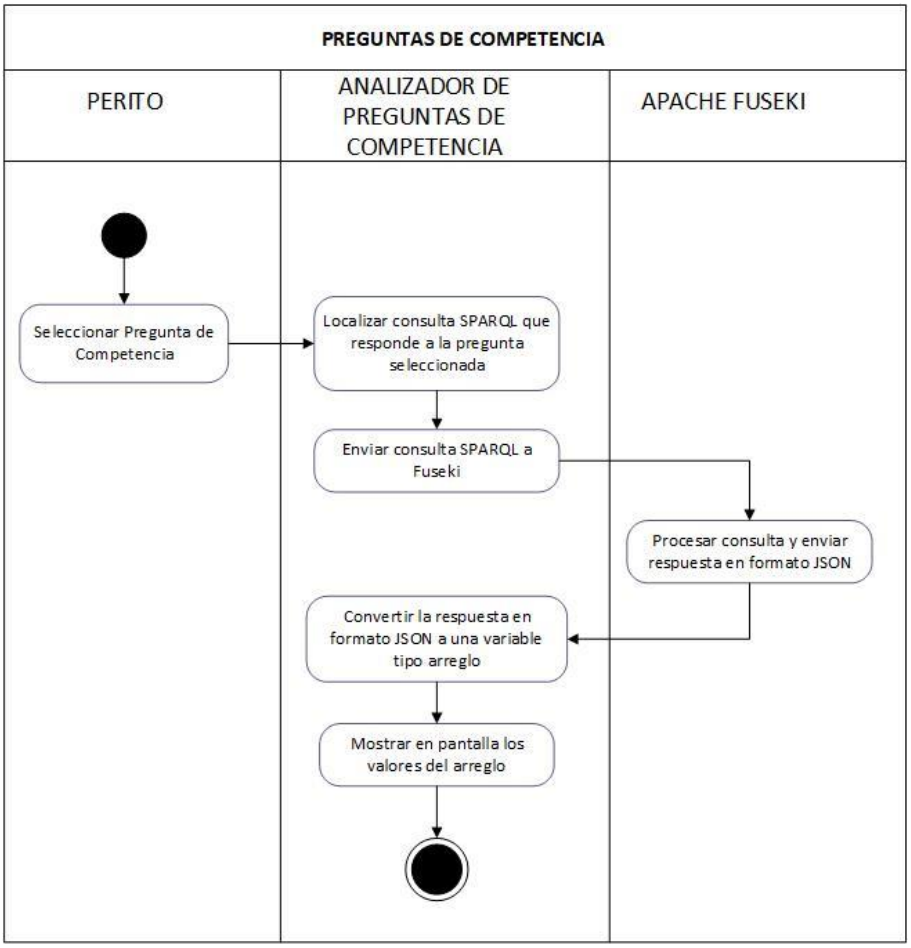


Figura 4-13: Diagrama de Actividad del Subproceso de
CONSULTA DE PREGUNTAS DE COMPETENCIA

Una vez que el perito ha ingresado la cabecera del correo electrónico a analizar (o un conjunto de cabeceras), así como los datos complementarios que pueda aportar, la aplicación realiza la creación de las instancias de las clases de OntoFoCE que representan dichas cabeceras, de acuerdo a lo indicado en la sección 4.3.1, y deja a disposición del usuario el menú de preguntas de competencia que puede seleccionar. Estas preguntas pueden estar referidas a un único correo o a un conjunto de correos.

El usuario selecciona la Pregunta de Competencia que le interesa, y con ello genera la búsqueda de la consulta SPARQL que responde a la pregunta seleccionada. Encontrada la misma, la aplicación envía dicha consulta al servidor Fuseki, que la

procesa, devolviendo la respuesta en un texto en formato JSON[55] (JavaScript Object Notation). La aplicación toma el texto y lo convierte en un arreglo para visualizarlo como respuesta en la pantalla correspondiente.

En la sección 4.6 se pueden observar las pantallas para la selección de las preguntas como para la visualización de las respuestas. Una vez mostrado los resultados de una pregunta, queda a disposición del usuario la selección de la siguiente pregunta o finalizar las consultas sobre preguntas de competencia.

4.5 Tecnologías Utilizadas Para Implementar ObE Forensics

El desarrollo de la aplicación se llevó a cabo utilizando una metodología ágil de software, fuertemente inspirada en *Scrum*[56], que consiste en realizar iteraciones incrementales, las cuales culminan con una pieza completamente funcional para el usuario final. Durante cada iteración se analizaron los requisitos, se planificó el Sprint (iteración), se hizo un diseño y se llevó a cabo la codificación para finalmente realizar las pruebas y documentación de la/s funcionalidad/es que proveía el Sprint.

Aun cuando las metodologías como SCRUM permiten llegar rápidamente a la aplicación ejecutable, no fue este el caso. ObE Forensics se desarrolló durante aproximadamente 2 años, durante los cuales se comenzó con una pequeña aplicación con algunas funcionalidades básicas, para ir construyendo las restantes sobre la misma base. En una serie de sucesivos sprint se fueron probando y concatenando los diferentes lenguajes y herramientas requeridos (Java, PHP, Protégé, Apache JENA, Apache Fuseki, etc.). Durante la última etapa se trabajó especialmente en mejorar la usabilidad y accesibilidad de la aplicación.

A continuación, se introducen el hardware y las tecnologías utilizadas para el desarrollo, almacenamiento y procesamiento de datos, así como para la implementación de las interfaces del usuario, y las cuestiones relacionadas con la seguridad de ObE Forensics. Las mismas se ilustran en la Figura 4-14 y se describen con las características correspondientes en los párrafos siguientes.

[55] https://www.json.org/
[56] https://hbr.org/1986/01/the-new-new-product-development-game

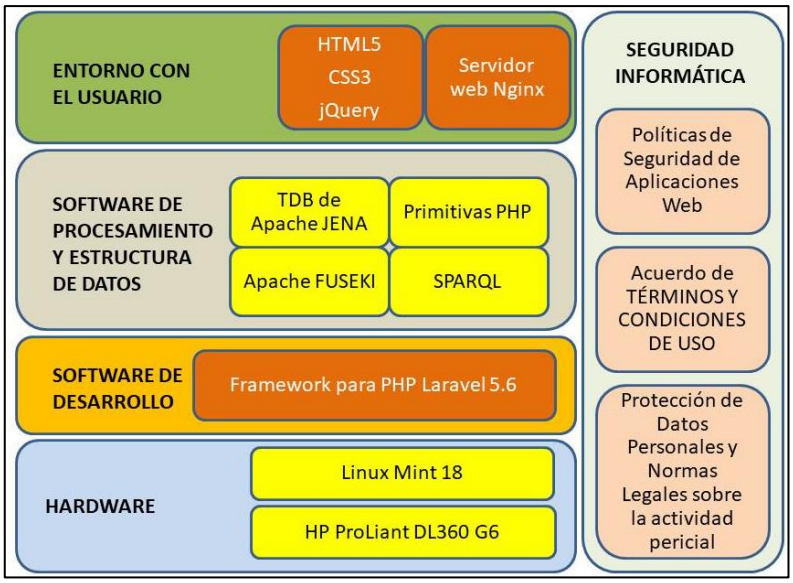


Figura 4-14: Arquitectura de Procesamiento de ObE Forensics

Las tecnologías utilizadas se describen atendiendo a los componentes básicos de una aplicación web: hardware, software de desarrollo de la aplicación, software para procesamiento de datos en la aplicación, estructura de datos utilizada, entorno de conectividad y políticas de seguridad informática.

Desde el punto de vista del ***hardware***, ObE Forensics ha sido desplegado en un servidor HP ProLiant DL360 G6 bajo Linux Mint 18, ubicado físicamente en el laboratorio Digilab de la Universidad Católica de Salta, el cual cuenta con un procesador Intel Quad Core Xeon E5504 y 4GB de RAM. Cabe destacar que el sistema completo puede ser perfectamente utilizado en, por ejemplo, una instancia t2.micro (capa gratuita) del servicio EC2 ofrecido por Amazon Web Services, t2.nano[57].

El conjunto de herramientas de ***software de desarrollo*** utilizado para construir la aplicación es variado. Se utilizó el framework para PHP[58] Laravel 5.6[59], el cual provee un entorno de trabajo del tipo Modelo-Vista-Controlador (MVC) que es la aplicación de tres componentes esenciales: el dominio de la aplicación (el modelo), la visualización del estado de la aplicación (la vista) y la interacción entre la vista y el modelo (el controlador). A través de las Vistas, el usuario interactúa con el sistema, proveyendo las entradas necesarias y pudiendo visualizar los procesamientos que el sistema realiza. Los Controladores se encargan de la lógica, el enrutamiento y la conexión con los Modelos, quienes proveen una interfaz para el

[57] https://aws.amazon.com/es/ec2/instance-types
[58] http://php.net/
[59] https://laravel.com/

acceso a los datos. Los Modelos no son más que clases de PHP que implementan lo necesario para interactuar con los datos, usualmente a través de un ORM (*Object Relational Mapping*), que permite una integración muy ajustada al lenguaje de programación. Sin embargo, dado el contexto de ObE Forensics, no ha sido posible utilizar Eloquent, el ORM proporcionado por Laravel, ya que está pensado para trabajar con base de datos relacionales y/o no relacionales.

Fue necesario construir clases que faciliten el armado de las consultas SPARQL para hacer posible la comunicación con Apache Fuseki, que provee una API para Apache Jena. Dicha comunicación entre el sistema y Apache Fuseki se realiza desde los Controladores a través de peticiones JSON utilizando la librería Guzzle[60], las cuales reciben una consulta SPARQL y devuelven una respuesta del tipo JSON.

La instanciación de la ontología es el eje principal de esta aplicación, por lo tanto, se trata de realizarla de la manera más conveniente. Cada proceso PHP que se encarga de poblar la ontología es capaz de ejecutar sólo un trabajo en un momento determinado. Es decir, para procesar más de un correo, un sólo proceso de PHP se irá ejecutando a la vez de manera secuencial, lo cual incide en la velocidad de procesamiento. Para solucionar este problema se recurrió al sistema de colas que provee *Laravel* en el que cada trabajo es agregado a una cola, donde eventualmente será procesado por el próximo proceso trabajador que esté desocupado. Así, es posible contar con varios procesos simultáneos de PHP disponibles para procesar los trabajos en cola.

La aplicación ObE Forensics se basa en ***estructuras de datos*** muy básicas, tales como texto plano, para leer los correos electrónicos, y las primitivas del lenguaje PHP (String, Integer, Arrays, etc.). Además, la herramienta propuesta utiliza TDB (Triple Database) provisto por Apache Jena, para almacenar las tripletas de la ontología OntoFoCE, de modo que luego se pueda realizar las consultas SPARQL mediante Apache Fuseki.

En cuanto al ***entorno de conectividad***, los usuarios pueden acceder a ObE Forensics a través de un navegador web, que se encuentra activo en la dirección https://digilab.ucasal.edu.ar, que facilita el acceso a la IP 200.10.181.131 a través del puerto 80. Esta conexión es provista por un servidor web Nginx. Internamente, la

[60] http://docs.guzzlephp.org/en/stable/

conexión entre ObE Forensics y Apache Fuseki se realiza en el mismo *host* a través del puerto 3030 y el protocolo HTTP.

La ***interface de comunicación con el usuario*** está estructurada en base a una aplicación web construida con HTML5[61], CSS3[62] y jQuery[63], que permite el desarrollo de las funcionalidades provista por ObE Forensics.

Respecto de la ***Seguridad Informática*** la aplicación contempla:

a) Políticas de Seguridad propias de aplicaciones web para garantizar el acceso solo a los usuarios autorizados, así como para el resguardo y recuperación de datos.
b) ObE Forensics requiere de un formulario de aceptación de condiciones de servicio por parte del usuario que ingresa por primera vez.
c) La aplicación requiere del cumplimiento de normas referidas a la protección de datos personales y otras normas pertinentes a la realización de pericias informática.

En respuesta al apartado a) se establecieron las políticas de acceso por parte de los usuarios expertos, restringiendo el ingreso a la aplicación mediante la entrega de un usuario y contraseña que debe solicitar vía correo electrónico. A cada usuario se le asigna un *dataset* de Apache Fuseki, lo cual le otorga un espacio privado y no compartido con otros usuarios.

Para dar respuesta al segundo apartado, y en atención a que la aplicación utiliza información reservada (evidencia digital) como insumo, se estableció el formulario de aceptación de las condiciones de servicio de la herramienta, establecidas en una página de "Términos de Uso", mediante la cual se informa al usuario acerca de las restricciones y condiciones de utilización de la aplicación, enfatizando los aspectos legales vinculados con la responsabilidad en el uso de datos que provienen de una evidencia digital.

En relación a las cuestiones de protección de datos personales, en caso de que ObE Forensics sea puesta en producción, es decir, disponible para su utilización en casos reales, será necesario cumplir con lo dispuesto por la Ley 25326, que en su artículo 3 trata sobre la licitud de los datos: "*La formación de archivos de datos será lícita cuando se encuentren debidamente inscriptos, observando en su operación los principios que establece la presente ley y las reglamentaciones que se dicten en su*

[61] https://developer.mozilla.org/es/docs/HTML/HTML5
[62] https://developer.mozilla.org/es/docs/Web/CSS
[63] https://jquery.com/

consecuencia“, así como cumplir con otros requerimientos técnicos y formales exigidos por esta norma respecto de la protección, uso y disposición de datos personales.

Por otra parte, ObE Forensics fue desarrollada atendiendo a las cuestiones de seguridad de aplicaciones tipo web. En este sentido el framework Laravel permite:

- Evitar la falsificación de petición en sitios cruzados (*Cross-Site Request Forgery*), mediante la implementación de un *token* generado específicamente para cada sesión de usuario.
- Evitar la inyección de scripts maliciosos XSS (*Cross-Site Scripting)* mediante la aplicación de reglas de validación sobre todos los datos ingresados por el usuario.

4.6 Uso de ObE Forensics

La Figura 4-1 señala los casos de usos de ObE Forensics, con las funciones que en cada caso se indicaron en la sección 4.3, es decir:

- Ingresar Archivo de Texto Plano
- Ingresar Datos Complementarios
- Instanciar los datos en OntoFoCE
- Responder las preguntas de competencia sobre un único correo
- Responder las preguntas de competencia sobre un conjunto de correo
- Visualizar e imprimir los datos procesados
- Limpiar el formulario de ingreso

Cada una de estas actividades se implementación en la aplicación web mediante una serie de interfaces de comunicación con el usuario experto, que a continuación se explican.

4.6.1 Registro de Usuarios

Previo a la utilización de ObE Forensics, el Perito debe registrarse en la aplicación. A continuación, se presenta la pantalla para ingresar los datos de registro del nuevo usuario (Figura 4-15).

Figura 4-15: Pantalla de Registro de Datos de Nuevo Usuario

Una vez que el usuario registra sus datos (nombre, correo electrónico, contraseña y confirmación de contraseña), debe aceptar los Términos y Condiciones de Uso, cliqueando en la casilla correspondiente. Los datos le llegan al administrador de la aplicación quien responde a la solicitud ingresando los datos del usuario en el Registro de Usuarios Habilitados y enviando un correo con la confirmación de la registración.

Para registrarse como usuario, es necesario ingresar a la aplicación por el link *Registrarse*, que habilita la pantalla de "Términos y Condiciones de Uso" (Figura 4-16) en la que se establece el marco legal al que debe ajustarse el usuario para utilizar la aplicación.

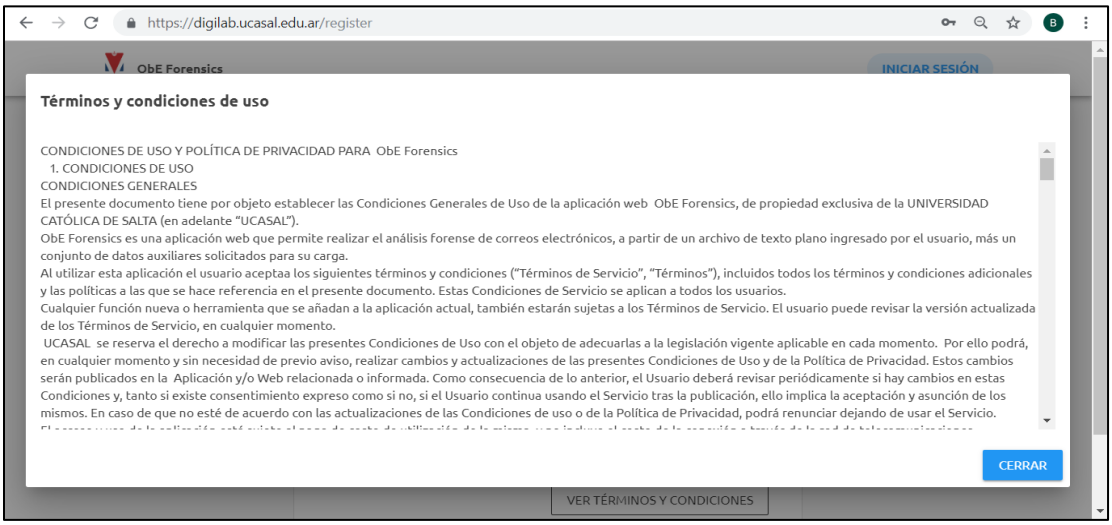


Figura 4-16: Pantalla sobre Términos y Condiciones de Uso de ObE Forensics

Una vez que el usuario recibe sus credenciales de acceso, puede utilizar ObE Forensics sucesivas veces, y en distintas ocasiones, con diferentes correos o conjuntos de correos.

Las credenciales quedarán habilitadas de manera permanente, de acuerdo a las condiciones de registro y uso de la aplicación. Por el momento, al tratarse de una versión Beta, la aplicación no tiene restricciones de uso más que las indicadas para tramitar el formulario de registro.

Cabe mencionar que en los datos que se trabajan en cada sesión se mantienen vigentes durante la misma y para mantener las condiciones de reserva y privacidad de los datos que componen la evidencia digital que se está analizando, todos los datos ingresados y procesados se borran cuando que el usuario cierra su cesión de trabajo.

La pantalla de la Figura 4-17 muestra los datos de ingreso a la aplicación para el usuario registrado.

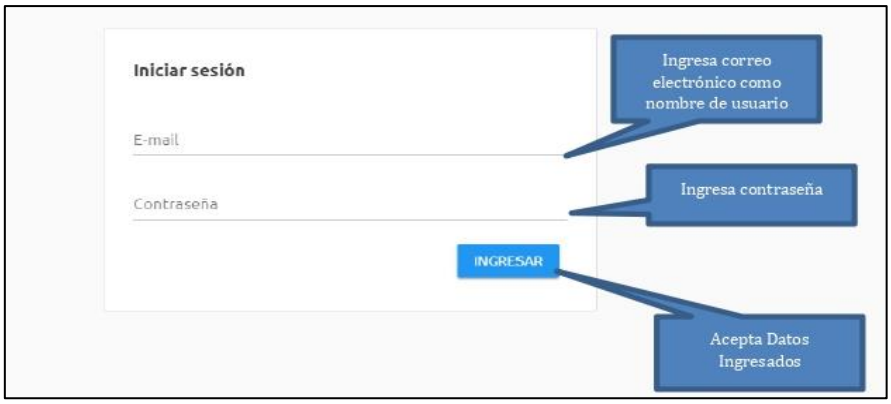


Figura 4-17: Pantalla de registro de datos de acceso para los usuarios registrados

4.6.2 Ingresar Archivo de Texto Plano

Cuando el usuario registrado ingresa, se presenta la pantalla inicial de la aplicación mediante la cual podrá introducir la cabecera o el conjunto de cabeceras en cuestión. La Figura 4-18 muestra dicha pantalla, en la cual, al activar el botón etiquetado como "***Ingresar Archivo de Texto Plano***" se presenta la pantalla para seleccionar la cabecera del correo a analizar (Figura 4-19).

Figura 4-18: Pantalla de Ingreso del Archivo de Texto Plano

Figura 4-19: Pantalla para seleccionar la cabecera a analizar

Haciendo click en el botón “Elegir Archivo” de la pantalla que se muestra en la Figura 4-19, el usuario puede acceder al directorio o dispositivo en el que tenga resguardado el archivo de texto plano de la cabecera. Es en esta opción en donde el usuario puede seleccionar una o varias cabeceras de correo que participen del análisis forense. Es importante destacar que no hay límite a la cantidad de archivos a subir. Seleccionado el archivo, y una vez que se cliquea en el botón “Aceptar” de esa pantalla (Figura 4-19), la aplicación, siguiendo el proceso explicado en la sección 4.3.1, separa la cabecera y cuerpo del correo electrónico, verifica la validez de la cabecera para realizar el análisis forense, separa y crea las instancias en las clases y relaciones de OntoFoCE.

Cuando culmina dicho procesamiento, se muestra la pantalla de Figura 4-20 para que el usuario interactúe según sea lo que necesita. Se presentan las opciones que puede trabajar el perito, organizadas en dos módulos denominados *Datos* y *Puntos de Pericia*, los cuales se describen a continuación.

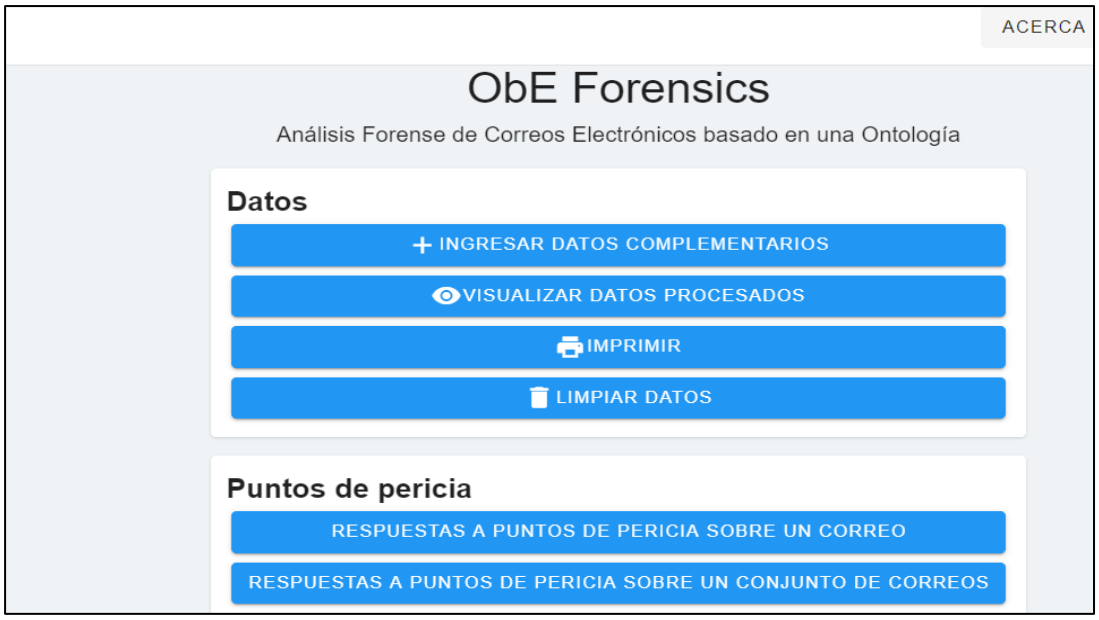


Figura 4-20: Pantalla de Selección de Opciones de Análisis Forense del Correo Electrónico

4.6.3 Módulo de DATOS

El módulo *Datos* permite procesar diferentes acciones referidas al conjunto de datos que se está analizando, las cuales se resumen a continuación:

- El botón "***Ingresar Datos Complementarios***" permite que el usuario registre los datos relevados durante la fase de adquisición de la evidencia y que son de interés para el análisis forense pero no se encuentran en la cabecera del correo electrónico, tal es el caso de las direcciones MAC Address de los equipos o el número de expediente de la causa, entre otros.
- El botón "***Visualizar Datos Procesados***" muestra toda la información de la cabecera analizada y las ocurrencias identificadas en la trazabilidad realizada.
- El botón "***Imprimir***" presenta una visualización de todos los datos ingresados y las ocurrencias identificadas, en formato PDF para su impresión.
- El botón "***Limpiar Datos***" permite borrar los datos del archivo de texto plano procesado y volver a cargar otro archivo para realizar el análisis.

En los párrafos siguientes, se describen las sucesivas pantallas que describen cada una de estas actividades.

Ingresar Datos Complementarios

Cuando el usuario selecciona el botón de "***Ingresar Datos Complementarios***", se presenta la pantalla de la Figura 4-21 que le permite registrar los datos del Expediente, Equipo Emisor, Equipos Receptores y Servidores Intermedios, a la que se accede seleccionando la opción Expediente en la pantalla que se muestra en la Figura 4-21.

Figura 4-21: Pantalla para Ingresar Datos Complementarios

La carga de estos datos no es condicionante para la realización del análisis pericial, se puede obviar en caso de que el perito no cuente con la información necesaria, y se registra a fin de contar con más detalles al momento de responder las preguntas de competencia.

Los datos del Expediente Judicial, con el detalle de N° de Expediente y Causa se registran según la pantalla que se muestra en la Figura 4-22. En el ejemplo que se muestra en dicha pantalla, se cargó el valor 12345/19 como número de expediente, y la carátula de la causa es JUAN vs PEDRO POR COBRO DE COSTAS JUDICIALES.

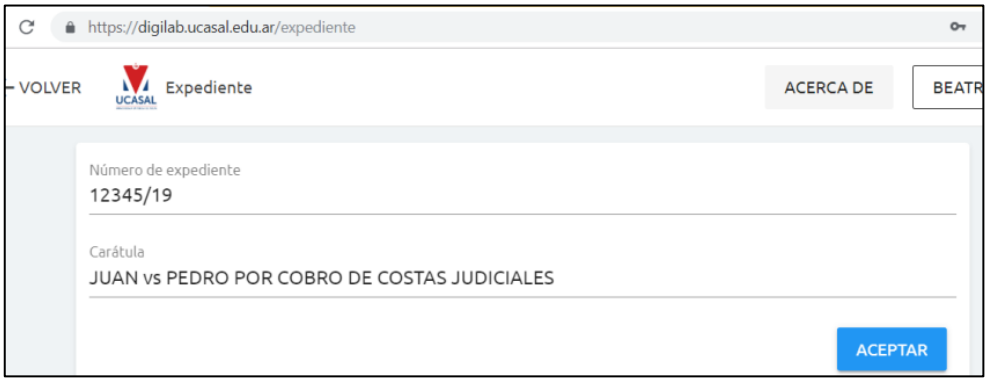


Figura 4-22: Pantalla de Carga de Datos Complementarios sobre EXPEDIENTE

La información que el Perito recaba respecto del o los equipos emisores se carga en la pantalla de la Figura 4-23. Allí se puede informar, para cada cuenta que ObE Forensics identificó como cuenta emisora, los datos de la identificación de MAC Address del equipo, y el tipo de cliente de correo utilizado, seleccionando además si se trata de un cliente local o remoto. En particular en la figura mencionada se puede observar la carga de datos del equipo emisor del correo que se presentó en la Figura 4-5. Este correo ha sido emitido por la cuenta bgallo@ucasal.edu.ar, desde un equipo cuya MAC Address es A5-F2-DA-B1-C3, utilizando Outlook como cliente local.

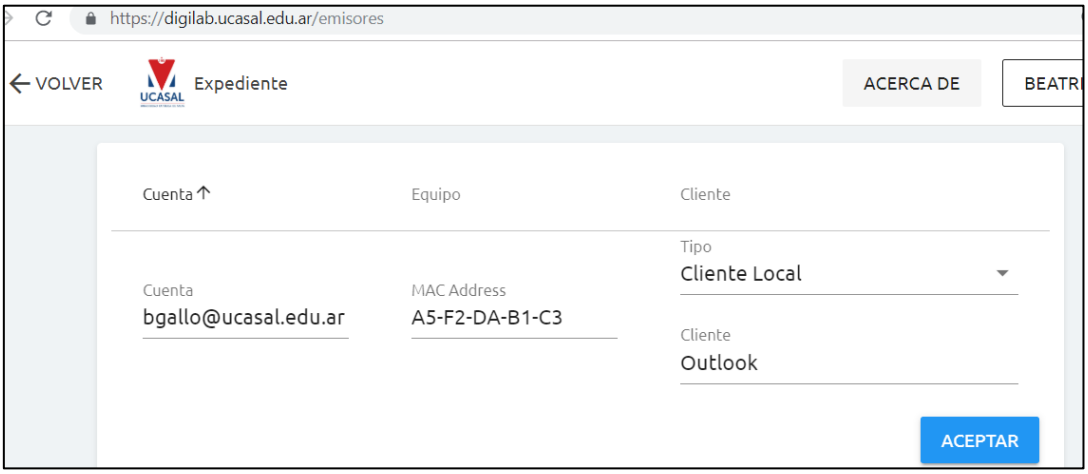


Figura 4-23: Pantalla de Carga de Datos Complementarios sobre EQUIPO EMISOR

Por último, en la pantalla que se muestra en la Figura 4-24 el Perito puede registrar la información de cada servidor que interviene en la transmisión del correo. ObE Forensics despliega el total de servidores identificados en las ocurrencias por su dirección IP/Hostname, y para cada uno, se puede cargar la Identificación de MAC Address del equipo y una breve descripción del mismo. Para los datos de identificación de los servidores que se muestran en la Figura 4-24 no fue posible obtener la información de la MAC Address ni la descripción particular de cada equipo, por ello, solo se indican con un nombre genérico a los fines referenciar más fácilmente los servidores.

Figura 4-24: Pantalla de Carga de Datos Complementarios sobre SERVIDORES

Visualizar Datos Procesados

Continuando con las opciones principales del Módulo de Datos enunciadas en la Figura 4-20, corresponde describir ahora la opción *Visualizar Datos Procesados.*

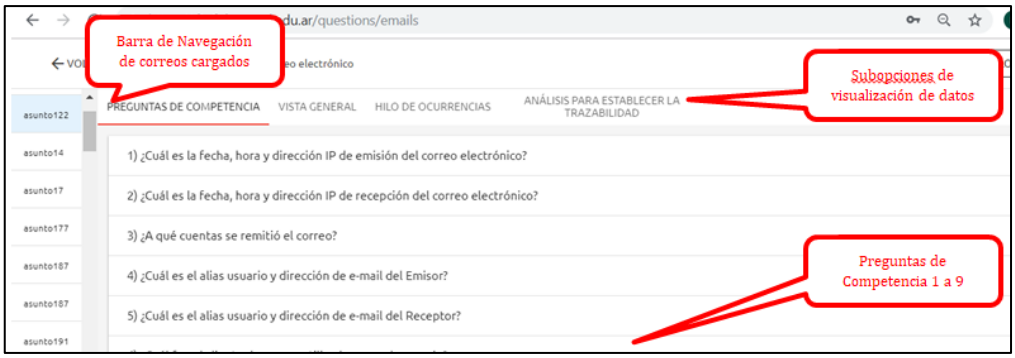


Figura 4-25: Pantalla de VISUALIZAR DATOS PROCESADOS
y PREGUNTAS DE COMPETENCIA

Cuanto el usuario selecciona esta opción en la pantalla de la Figura 4-20, se muestra la pantalla que se introduce en la Figura 4-25, en la que se distinguen tres áreas: a) sobre el margen izquierdo hay una *lista de navegación* de los correos ingresados identificados por el *Asunto*; b) en la parte superior está la *barra horizontal de sub-opciones* de visualización de datos sobre el correo seleccionado; y c) en la parte principal de la pantalla se muestran la pantalla correspondiente a cada elección de la barra de sub-opciones y los *resultados* obtenidos para el correo que se haya seleccionado de la lista de navegación. Por defecto, esta área muestra la pantalla que permite seleccionar las preguntas de competencia P01 a P09 para el correo seleccionado.

Este menú de sub-opciones de visualización de datos presenta cuatro alternativas:

- PREGUNTAS DE COMPETENCIA,
- VISTA GENERAL,
- HILO DE OCURRENCIAS y
- ANÁLISIS PARA ESTABLECER LA TRAZABILIDAD.

A continuación se explica cada una de estas alternativas, mostrando un ejemplo de los resultados obtenidos para cada caso, siempre considerando que previamente debe seleccionarse una cabecera de la lista de navegación.

La sub-opción PREGUNTAS DE COMPETENCIA permite consultar las preguntas PC01 a PC09 para la cabecera seleccionada en la lista de navegación, mostrando los resultados más usualmente requeridos en las pericias de correos electrónicos. Una muestra de esa pantalla se puede observar en la Figura 4-25.

La opción de VISTA GENERAL, que se muestra en la Figura 4-26, muestra los pares {*Parametro*}:{*Valor*} que se instanciaron para el correo seleccionado de la barra de navegación izquierda.

El uso de colores para identificar los parámetros y sub-parámetros de la cabecera ayuda al perito a analizar con más facilidad los datos de la cabecera.

Figura 4-26: Pantalla de Visualización de Sub-opción VISTA GENERAL

Si se selecciona la opción HILO DE OCURRENCIAS la aplicación presenta todas las ocurrencias correspondientes al hilo del correo seleccionado, en el orden en que se fueron generando.

La Figura 4-27 muestra una vista parcial de esa pantalla, con el hilo de ocurrencias del ejemplo que se está procesando.

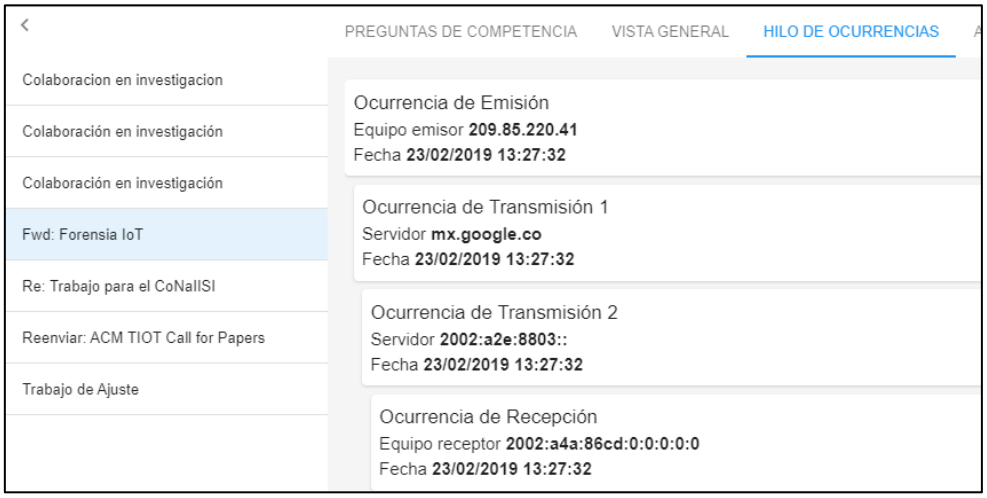


Figura 4-27: Pantalla de Visualización de sub-opción HILO DE OCURRENCIAS

La selección de la opción ANÁLISIS PARA ESTABLECER LA TRAZABILIDAD muestra los tres momentos de la identificación de ocurrencias y equipos:

- las instancias de las ocurrencias que solo se identifican a partir del valor del sub-atributo by,
- luego muestran en orden, las nuevas ocurrencias que se van generando a partir del from, y por último,
- ordena y renumera las instancias de las ocurrencias para presentar la trazabilidad de la transmisión.

Esta funcionalidad de la aplicación se explicará en detalle al momento de presentar el caso de estudio en las secciones siguientes de este mismo capítulo.

Imprimir

Mediante esta opción, presente en el menú que se muestra en la Figura 4-20, la aplicación muestra dos paneles de navegación:

- el primero con el listado de correos cargados, para seleccionar desde allí el o los que se quiera imprimir, y
- el segundo panel de navegación con las preguntas de competencia, con la posibilidad de que el usuario seleccione una o varias.

Con los datos seleccionados, se genera el archivo PDF que incluye toda la información procesada para la cabecera seleccionada: datos de identificación de las cuentas, ocurrencias y respuestas a las preguntas de competencia, con el objetivo de que este documento se imprima y el perito pueda adjuntarlo como anexo técnico al Informe de Pericia.

La Figura 4-28 muestra este anexo técnico para el ejemplo en cuestión, y en las secciones siguientes de este capítulo se ejemplifica el informe a obtener para los escenarios de prueba.

La pantalla de la figura citada muestra 3 sectores:

- el de la izquierda es una barra de navegación para seleccionar uno o varios correos electrónicos que se quiera incluir en el Informe,
- el sector del centro permite marcar las preguntas de competencia cuya respuesta se quiere agregar en el Informe Técnico, y
- el sector de la derecha presenta una vista preliminar del informe.

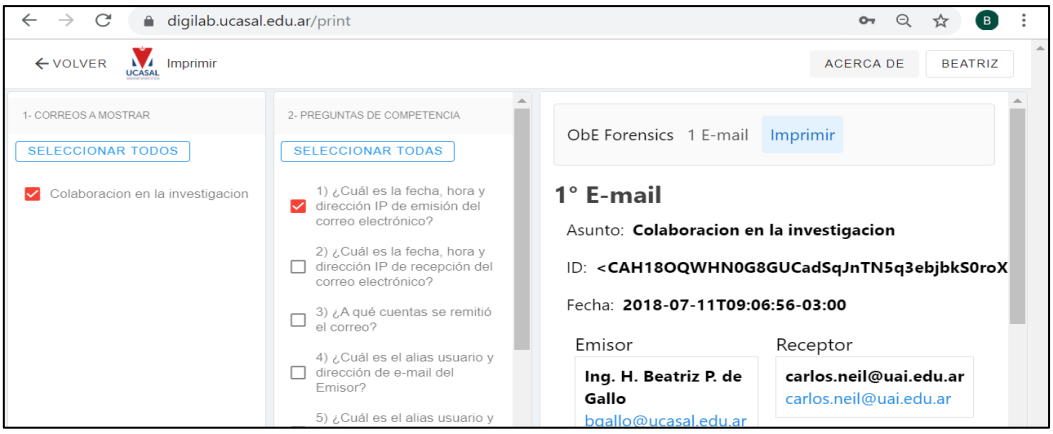


Figura 4-28: Pantalla de VISTA PREVIA

Limpiar Datos

Con esta opción, la última del menú disponible en la pantalla principal de ObE Forensics (Figura 4-20), es posible borrar los datos cargados y procesados, para volver a ingresar nuevas cabeceras. Esta opción tiene una pregunta de confirmación de eliminación de los datos, que se muestra en la Figura 4-29.

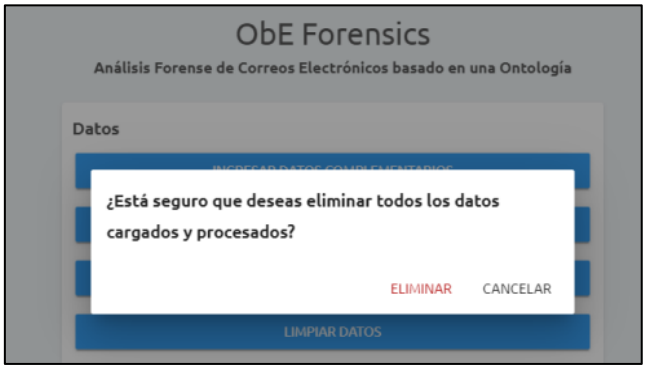


Figura 4-29: Pantalla de LIMPIAR DATOS

4.6.4 Módulo de Puntos de Pericia

El módulo *Puntos de Pericia* permite acceder a las consultas sobre las preguntas de competencia sobre OntoFoCE y presenta dos opciones: "***Respuestas a Puntos de Pericia sobre un correo***" y "***Respuestas a Puntos de Pericia sobre un conjunto de correos***".

El primer botón muestra los resultados de las preguntas de competencia relacionadas a un único correo (preguntas de competencia P01 a P09).

En tanto, el segundo muestra los resultados de las restantes preguntas de competencia (P10 a P21). A continuación se describen ambas opciones.

Respuestas a Preguntas de Competencia para un único correo

Luego que el perito cargó las cabeceras de los correos electrónicos a analizar, puede utilizar esta opción para encontrar las respuestas a las nueve preguntas de competencia que se pueden realizar sobre un único correo electrónico.

Para ello, se accede a la opción correspondiente desde la pantalla principal de ObE Forensics (Figura 4-20).

Al clickear el botón correspondiente, accede a una nueva pantalla (Figura 4-30) que señala:

- en el panel de navegación de la izquierda, el listado de correos disponibles para análisis, y
- en el centro de la pantalla, se visualizan las 9 preguntas que se pueden responder.

Al seleccionar alguna o todas ellas con el botón *Expandir Todas*, se despliega la respuesta correspondiente para la cabecera seleccionada en el panel de navegación de la izquierda y para la pregunta de competencia seleccionada en el panel central.

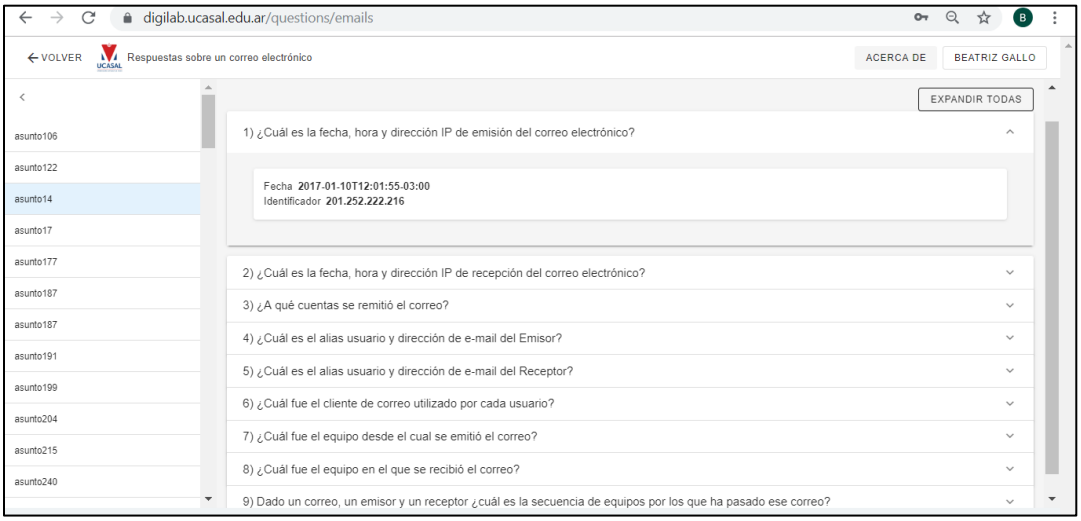


Figura 4-30: Pantalla de RESPUESTAS A PREGUNTAS DE COMPETENCIA SOBRE UN ÚNICO CORREO

Respuestas a Preguntas de Competencia para un Conjunto de Correos

Esta opción solo es válida cuando se ha cargado un grupo de correos, ya que las preguntas de competencia involucran el cruce de datos sobre dicho conjunto.

La Figura 4-31 muestra la pantalla que se presenta, en la que figuran las preguntas P10 a P21 para que el Perito seleccione la que desee.

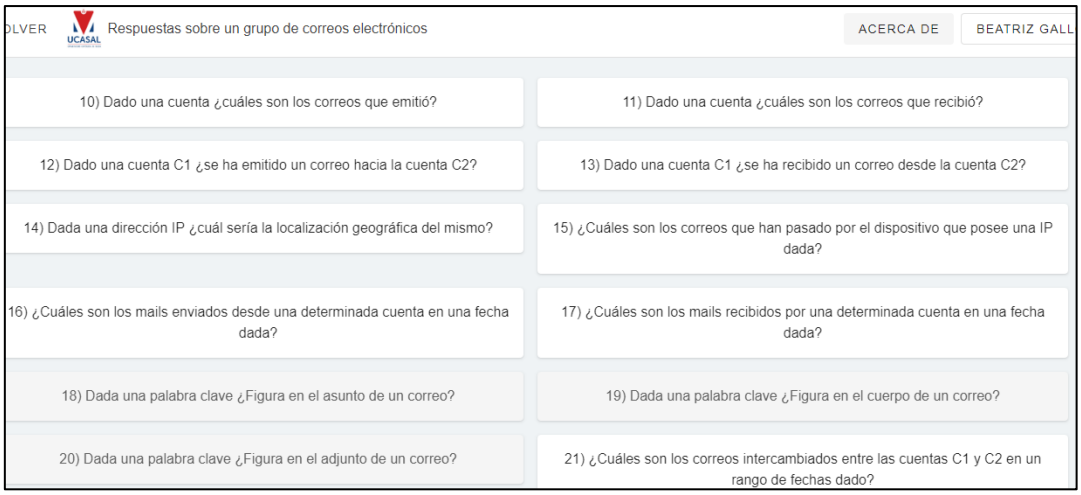


Figura 4-31: Pantalla de SELECCIÓN DE PREGUNTAS DE COMPETENCIA SOBRE UN GRUPO DE CORREOS ELECTRÓNICOS

Cabe mencionar que, aunque OntoFoCE contempla la representación de la búsqueda por palabras claves en el contenido del ASUNTO, CUERPO y ADJUNTO del correo en análisis, tales funcionalidades no se han implementado aún en el prototipo vigente de ObE Forensics, razón por la cual las preguntas P17, P18 y P19 no están disponibles para utilizarse en la aplicación web.

El perito seleccionará la pregunta necesaria presionando el botón correspondiente, y según sea el caso, se pedirán los datos de filtro necesarios para procesar los correos. A modo de ejemplo se muestran las pantallas de interface de comunicación de la pregunta de competencia PC10, las restantes preguntas –PC11 a PC21- cuentan con una estructura similar para el ingreso de los datos a filtrar.

PC 10: Dado una cuenta C ¿cuáles son los correos que emitió?

El proceso de la Pregunta de Competencia P10 es el siguiente:

- Paso 1: se requiere de la elección de una cuenta sobre la cual se quiere consultar cuales son los correos emitidos (Figura 4-32).

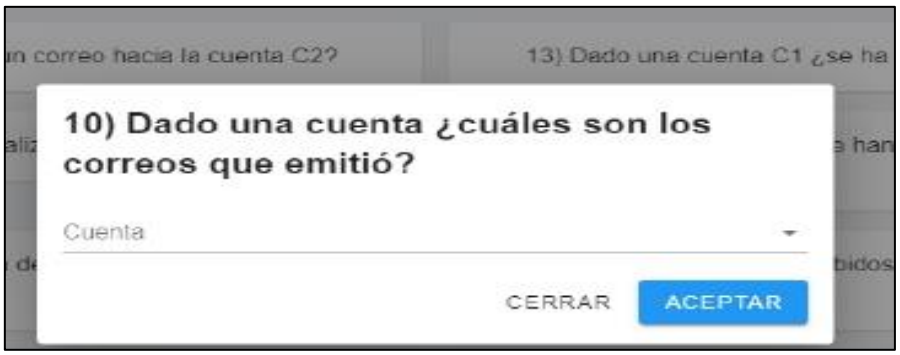


Figura 4-32: Pantalla 1 de Selección de Parámetros de la Pregunta de Competencia PC10

- Si el Perito así lo desea puede desplegar el listado de todas las cuentas que se han identificado en el total de cabeceras ingresadas (Figura 4-33).

Figura 4-33: Pantalla 2 de Selección de Parámetros de la Pregunta de Competencia PC10

- Una vez seleccionada la cuenta se acepta la opción y se presenta la pantalla de la Figura 4-34. Allí se observa, en la barra de navegación de la izquierda, todos los correos que cumplen la condición de la pregunta de competencia, es decir, que hayan sido emitidos desde la cuenta seleccionada.

Figura 4-34: Pantalla 1 de Visualización de Resultados de la Pregunta de Competencia PC10

- Se selecciona el correo de interés, y se muestran cuatro opciones: “Preguntas de Competencia” específicas para el correo, “Vista General”, “Hilo de Ocurrencias” y “Análisis para establecer la Trazabilidad” (ver Figura 4-35). Todas estas opciones se condicen con lo descripto para cada una en la sección 4.6.3 de este mismo capítulo.

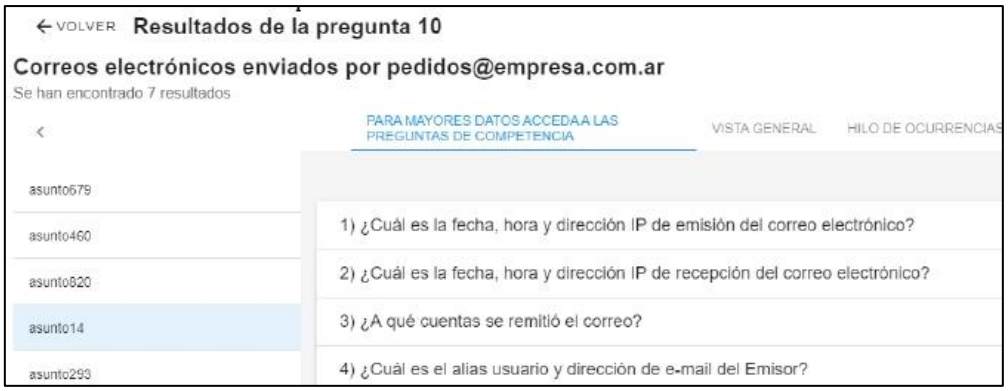


Figura 4-35: Pantalla 2 de Visualización de Resultados de la Pregunta de Competencia PC10

Para la pregunta P10 se solicitan los datos referidos a una cuenta en particular, mientras que la pregunta P15 pide una dirección IP y la P21 requiere que se ingresen dos cuentas de correo y un rango de fechas. El resto de las preguntas requieren de parámetros propios, que se indican en oportunidad de ejemplificar las mismas con los casos de estudio.

Hasta aquí se describió la funcionalidad de ObE Forensics, en términos de las pantallas de interface de comunicación con el usuario. A fin de mostrar el uso de esta herramienta en el análisis forense de correos electrónicos, en la siguiente sección de este capítulo se presentan cuatro casos de estudio.

4.7 Casos de Estudio

Para mostrar la funcionalidad de ObE Forensics, desde un enfoque integral, en el que se pueda destacar la vinculación entre OntoFoCE y ObE Forensics, se considerarán los tres escenarios de prueba ya considerados en el capítulo anterior: a) Análisis Forense de un Correo Electrónico a un único receptor, b) Análisis Forense de un correo enviado a varios receptores, c) Análisis Forense de un conjunto de correos y se agrega un cuarto caso: d) Análisis Forense de una cuenta de correo. En las secciones siguientes se describe cada caso.

Para garantizar que la actividad cumple con lo dispuesto por las normas técnicas y legales correspondientes, el perito debe realizar el análisis forense siguiendo el procedimiento indicado en el capítulo 2 sección 2.5, y que se resume en los siguientes párrafos

En la *Fase de Relevamiento* toma conocimiento del caso y de acuerdo a los puntos de pericia solicitados, identificará los equipos y cuentas a la que deberá

acceder también observará si debe tramitar ante el Juez la autorización de acceso al equipo emisor o a los servidores de correo si fuera el caso.

En las *Fases de Recolección y Adquisición* el perito accede al correo cuya copia impresa se adjunta en el expediente de la causa judicial y obtiene la cabecera del mismo, en ese momento recaba también los datos complementarios necesarios (datos del equipo receptor y del cliente de correo utilizado).

Luego sigue la *Fase de Preparación* en la que el Perito se registra como usuario de ObE Forensics obteniendo las credenciales para la utilización de la herramienta, y tramita ante el Juez la autorización para utilizar su propia computadora o la disponible en el Laboratorio Forense que utilizará.

A continuación, ya en la *Fase de Extracción y Análisis*, se accede a la herramienta para el análisis forense de correos electrónicos, según los pasos que se describen más adelante.

Por último, en la *Fase de Presentación de Resultados*, el Perito utiliza el Informe Técnico Pericial obtenido de ObE Forensics y lo entrega al Juez.

Durante la *Fase de Recolección y Adquisición* el Perito obtuvo lo siguiente:

- Cabecera del correo electrónico en formato de texto plano.
- Datos de identificación (nombre, MAC Address y cliente de correo) del equipo emisor, del equipo receptor y de los servidores, siempre que estos datos resulten necesarios para responder a los puntos de pericia.
- Datos de identificación del expediente (número y carátula de la causa).

Se considerará que dicho procedimiento se ha realizado ajustado a regla y a continuación se profundiza las fases de recolección de la evidencia, análisis forense con ObE Forensics e informe técnico pericial. Considerando que ya cuenta con la evidencia digital y los datos complementarios, el Perito accede a la aplicación, se registra como usuario, los ingresa y comienza a utilizar la herramienta. En la siguiente sección se explica el uso de ObE Forensics según los tres escenarios de estudio ya detallados en el capítulo 3.

4.7.1 Análisis Forense de un Único Correo

Retomemos el correo electrónico de ejemplo de la sección 3.5.1 del capítulo 3.

Si bien en el presente apartado se describe ObE Forensics, a través de la ejecución de la aplicación con el caso ejemplo del Escenario 1, no en todos los casos es posible mostrar las pantallas completas de los resultados que se obtienen, o explicar con sencillez el funcionamiento de la aplicación.

Por esa razón el lector puede ingresar a la siguiente URL https://digilab.ucasal.edu.ar, e ingresar y probar la aplicación a partir del caso ejemplo del Escenario 3 que ya se encuentra previamente ingresado (ver ANEXO VIII). Las credenciales de acceso son:

- Usuario: prueba@digilab.ucasal.edu.ar
- Contraseña: prueba

A fin de recordar el caso al lector, se reitera el correo en la Figura 4-36.

Colaboración en investigación
1 mensaje

Ing. H. Beatriz P. de Gallo <bgallo@ucasal.edu.ar> 8 de julio de 2018, 12:41
Para: Carlos Neil <carlos.neil@uai.edu.ar>

Estimados:

He seleccionado algunas personas de mis contactos para que reciban este correo, y me ayuden con mi investigación.
Lo único que tienen que hacer son DOS cosas:

1) Responder mi correo contestando estas preguntas:
- ¿Que dispositivo utilizó para abrir y responder el correo? (pc, celular, tablet, etc.)
- ¿Cuál es el cliente de correo que utilizó para abrir y responder el correo? (outlook, web mail de gmail o de hotmail, etc.)

2) Extraer la cabecera del correo y enviármela como archivo adjunto.

Con la información que Uds. me envíen armaré una batería de ejemplos para probar la ontología para análisis forense de mails en la que estoy trabajando.
Los molesto además para que -en caso de responder- otorguen su permiso para utilizar esos datos en esta instancia experimental de la investigación.

Muchísimas gracias por su ayuda!!

Sds
Beatriz

Figura 4-36: Correo ejemplo C1 tomado del Escenario 1 (sección 3.5.1 del Capítulo 3)

Se realiza la carga del archivo de texto plano, según lo indicado en la sección anterior de este mismo capítulo (ver Figura 4-37).

Figura 4-37: Carga de la cabecera del correo ejemplo C1

Consideremos además que el Perito consigue los datos complementarios referidos al EXPEDIENTE y procede a ingresarlos en la aplicación (Figura 4-38).

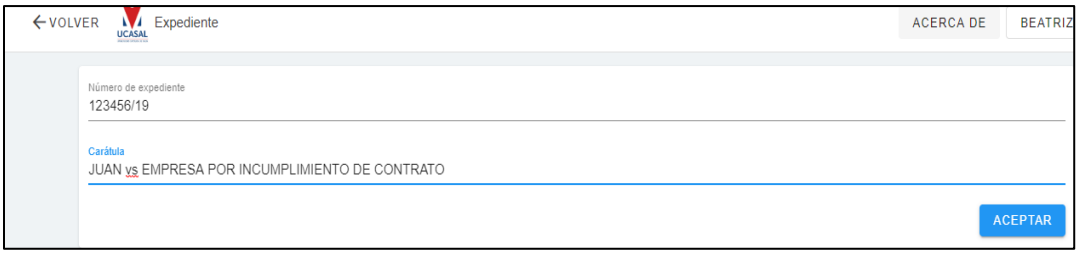


Figura 4-38: Carga de Datos Complementarios del EXPEDIENTE

Si bien en el ejemplo del Escenario 1 del capítulo 3, no se mostró la instanciación de estos datos, cabe mencionar que resulta de utilidad esta información para emitir luego el Informe Técnico de Pericia, mediante la opción de *Imprimir* que presenta ObE Forensics. Además el Perito Accede a la opción correspondiente para cargar los datos de identificación del equipo receptor (Figura 4-39).

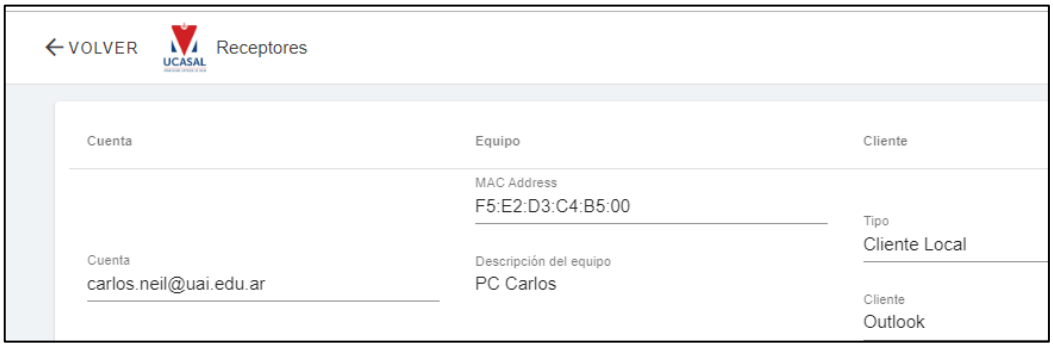


Figura 4-39: Datos Complementarios del EQUIPO RECEPTOR

Una vez registrados los datos complementarios, y con el archivo de texto plano ya procesado por ObE Forensics, el Perito puede visualizar los datos de la cabecera, accediendo a la opción de *Visualizar Datos Procesados* del menú principal de la aplicación, según lo que se muestra en la pantalla que se indica en la Figura 4-40.

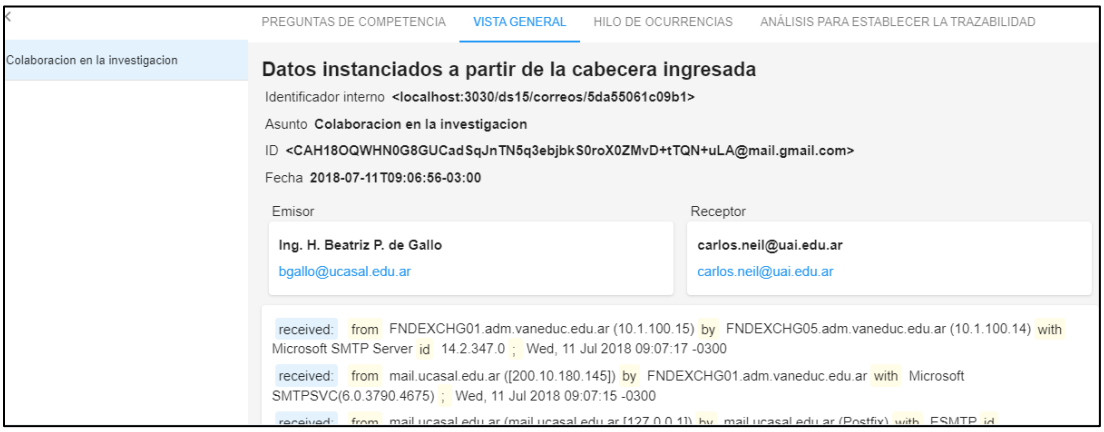


Figura 4-40: Visualización de Datos Procesados para la cabecera ejemplo C1

Allí se puede visualizar, entre otros datos, los siguientes, para el correo seleccionado según su *Asunto* en el panel de navegación de la izquierda:

- Datos de Identificación del correo: además del Asunto y Fecha de Emisión, se muestra el ID del correo electrónico.
- Datos de la Cuenta Emisora y de la Cuenta Receptora, con indicación de la cuenta y el alias de usuario correspondiente.
- Cabecera completa del correo en análisis. Por razones de espacio la Figura 4-40 muestra solo la parte superior de la cabecera. Obsérvese que se resalta con color el *parámetro* del par *{Parámetro}:{Valor}* de cada línea de la cabecera.

También puede acceder a la opción *Hilo de Ocurrencias*, cuyo resultado se muestra en la Figura 4-41.

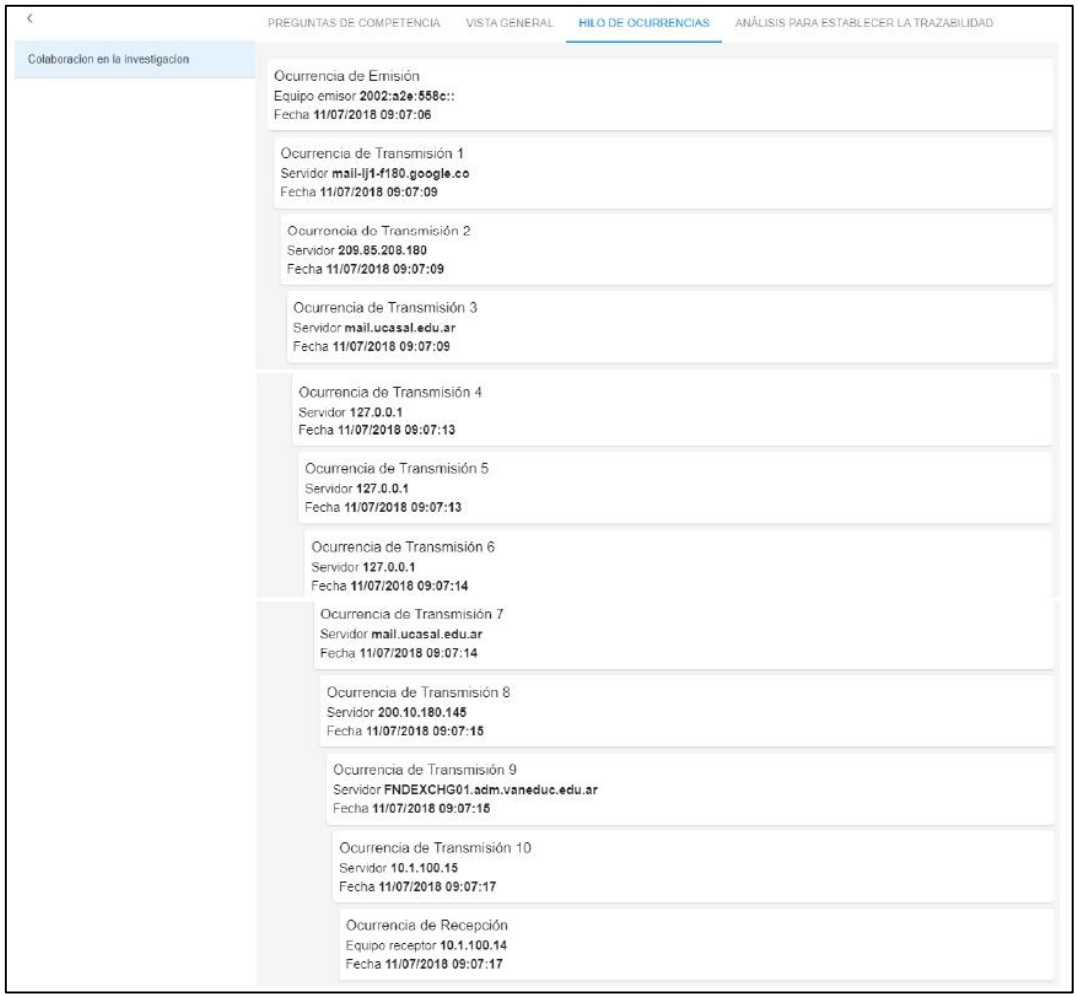


Figura 4-41: Visualización de Hilo de Ocurrencias para el correo ejemplo C1

De igual modo, cuando el Perito accede a la opción de *Análisis Para Establecer la Trazabilidad*, visualiza una pantalla que tiene 3 apartados: Ocurrencias Originales Identificadas (Figura 4-42), Ocurrencias Intermedias (Figura 4-43) y Ocurrencias Definitivas (Figura 4-44).

Aquí es donde ObE Forensics muestra el procesamiento realizado por el módulo denominado *Identificador de Instancias* descripto en la sección 4.3.1 de este mismo capítulo.

En dicho apartado, bajo el subtítulo de *Crear Instancias de Ocurrencias*, se explica el modo en que se consideran los valores de los parámetros *from* y *by*, para vincular las ocurrencias del proceso de transmisión e inferir de ello la trazabilidad del envío.

Si bien se recurre a OntoFoCE para instanciar los valores en cada clase correspondiente, desde ObE Forensic se realiza el proceso de ordenamiento de las ocurrencias descripto en el proceso indicado.

En particular, se muestra este procesamiento con el ejemplo de la cabecera C1 considerado. La Figura 4-42 muestra lo siguiente:

Figura 4-42: Visualización de Ocurrencias Originales Identificadas para la cabecera ejemplo CR1

Luego de mostrar las ocurrencias originales, aquellas obtenidas de los *By* en la primera recorrida de la cabecera, la aplicación muestra a continuación el detalle de las ocurrencias generadas, con una breve explicación del análisis sobre cada una de éstas ocurrencias. Esto se muestra en la pantalla de la Figura 4-43.

Figura 4-43: Visualización de Nuevas Ocurrencias para la cabecera ejemplo CR1

	127.0.0.1	2018-07-11T12:07:13.000000Z	Entre la ocurrencia OT2 y OT3 se ha agregado esta ocurrencia con la identificación 127.0.0.1 ya que el valor del sub-atributo FROM de OT3 (127.0.0.1) no coincide con el del sub-atributo BY de OT2 (mail.ucasal.edu.ar)
OT3	127.0.0.1	2018-07-11T12:07:13.000000Z	Ocurrencia original
	127.0.0.1	2018-07-11T12:07:14.000000Z	Entre la ocurrencia OT3 y OT4 se ha agregado esta ocurrencia con la identificación 127.0.0.1 ya que el valor del sub-atributo FROM de OT4 (127.0.0.1) no coincide con el del sub-atributo BY de OT3 (127.0.0.1)
OT4	mail.ucasal.edu.ar	2018-07-11T12:07:14.000000Z	Ocurrencia original
	200.10.180.145	2018-07-11T12:07:15.000000Z	Entre la ocurrencia OT4 y OT5 se ha agregado esta ocurrencia con la identificación 200.10.180.145 ya que el valor del sub-atributo FROM de OT5 (200.10.180.145) no coincide con el del sub-atributo BY de OT4 (mail.ucasal.edu.ar)
OT5	FNDEXCHG01.adm.vaneduc.edu.ar	2018-07-11T12:07:15.000000Z	Ocurrencia original
	10.1.100.15	2018-07-11T12:07:17.000000Z	Entre la ocurrencia OT5 y OR se ha agregado esta ocurrencia con la identificación 10.1.100.15 ya que el valor del sub-atributo FROM de OR (10.1.100.15) no coincide con el del sub-atributo BY de OT5 (FNDEXCHG01.adm.vaneduc.edu.ar)
OR	10.1.100.14	2018-07-11T12:07:17.000000Z	Ocurrencia original

Figura 4-43 (Cont): Visualización de Ocurrencias Generadas para la cabecera ejemplo CR1

Por último, en la Figura 4-44 se muestra la pantalla que presenta las ocurrencias renumeradas y organizadas, indicando las que fueron inicialmente identificadas, de aquellas que se generaron para no perder valores de identificación de equipos.

Ocurrencias definitivas

Ocurrencia	IDENTIFICACIÓN DEL EQUIPO	TIMESTAMP	
OE	2002:a2e:558c::	2018-07-11T12:07:06.000000Z	Ocurrencia original
OT1	mail-lj1-f180.google.co	2018-07-11T12:07:09.000000Z	Ocurrencia original
OT2	209.85.208.180	2018-07-11T12:07:09.000000Z	Ocurrencia generada
OT3	mail.ucasal.edu.ar	2018-07-11T12:07:09.000000Z	Ocurrencia original
OT4	127.0.0.1	2018-07-11T12:07:13.000000Z	Ocurrencia generada
OT5	127.0.0.1	2018-07-11T12:07:13.000000Z	Ocurrencia original
OT6	mail.ucasal.edu.ar	2018-07-11T12:07:14.000000Z	Ocurrencia original
OT7	200.10.180.145	2018-07-11T12:07:15.000000Z	Ocurrencia generada
OT8	FNDEXCHG01.adm.vaneduc.edu.ar	2018-07-11T12:07:15.000000Z	Ocurrencia original
OT9	10.1.100.15	2018-07-11T12:07:17.000000Z	Ocurrencia generada
OR	10.1.100.14	2018-07-11T12:07:17.000000Z	Ocurrencia original

Figura 4-44: Visualización de Ocurrencias Definitivas para la cabecera ejemplo CR1

Obsérvese que en las figuras que muestran las sucesivas pantallas de resultados hasta aquí mostradas coinciden con el contenido de la cabecera CR1 descripta en la sección 3.5.1 del capítulo 3, y con la identificación que allí se realiza de todas las ocurrencias del proceso de transmisión del correo ejemplo.

Una vez que el correo ha sido procesado, el Perito procede a identificar las preguntas de competencia que le permitirán responder los puntos periciales. Así, para cada uno de éstos, se procesan las preguntas de competencia necesarias.

Considerando el orden seguido en el capítulo 3 al verificar los resultados de las consultas SPARQL de cada pregunta de competencia, se muestra para cada una la pantalla del visor SPARQL que muestra las instancias generadas en ***OntoFoCE*** y parte de la pantalla de ***ObE Forensics*** donde se visualiza el resultado obtenido por esta herramienta para el correo C1 bajo análisis.

- **PC 01: ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?**

Para esta pregunta de competencia, se muestra ahora los resultados obtenidos luego de procesar la cabecera C1 en OntoFoCE –primeramente- y luego en ObE Forensics. Las Figuras 4-45.a y 4-45.b muestran los datos obtenidos en cada caso, que por una cuestión de espacio, las capturas de pantallas se recortan para mostrar la pregunta de competencia y los resultados obtenidos.

SPARQL query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?fechaEmision ?direccionIP
        WHERE { ?correo oc:correoTieneSecuencia ?s.
                ?s oc:secuenciaTieneHilo ?h.
                ?h oc:hiloTieneOcurrencia ?o.
                ?o rdf:type oc:OcurrenciaDeEmision.
                ?o oc:fechaHoraOcurrencia ?fechaEmision.
                ?o oc:ocurrenciaResideEnEquipo ?q.
                ?q oc:equipoTieneId?ip.
                ?ip oc:identificadorEquipo ?direccionIP.
                FILTER (?correo=oc:C1).
}
```

fechaEmision	direccionIP
"2018-07-11T12:07:06"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"2002:a2e:658d::"

Figura 4-45.a: Resultados obtenidos por OntoFoCE para la pregunta PC01

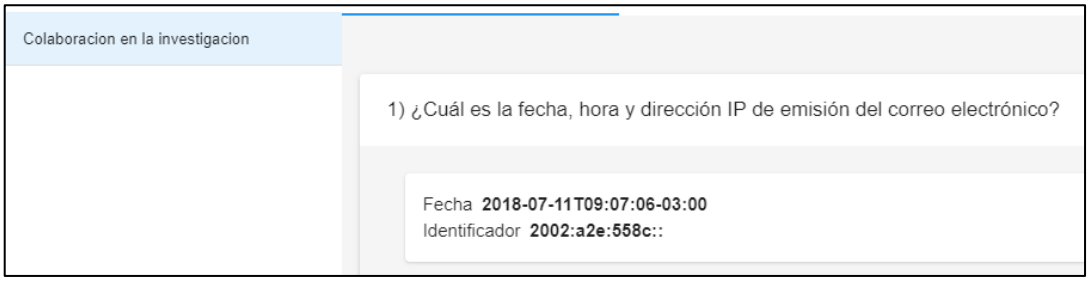


Figura 4-45.b: Resultados obtenidos por ObE Forensics para la pregunta PC01

Obsérvese que no hay coincidencia exacta en la *hora* de cada ocurrencia, debido a que ObE Forensics aplica la conversión de huso horario descripta y todas las fechas se indican para el huso horario de Argentina.

- **PC02: ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?**

En esta pregunta de competencia, los resultados obtenidos en OntoFoCE y en ObE Forensics se indican en las Figuras 4-46.a y 4-46.b que muestran los datos obtenidos en cada caso.

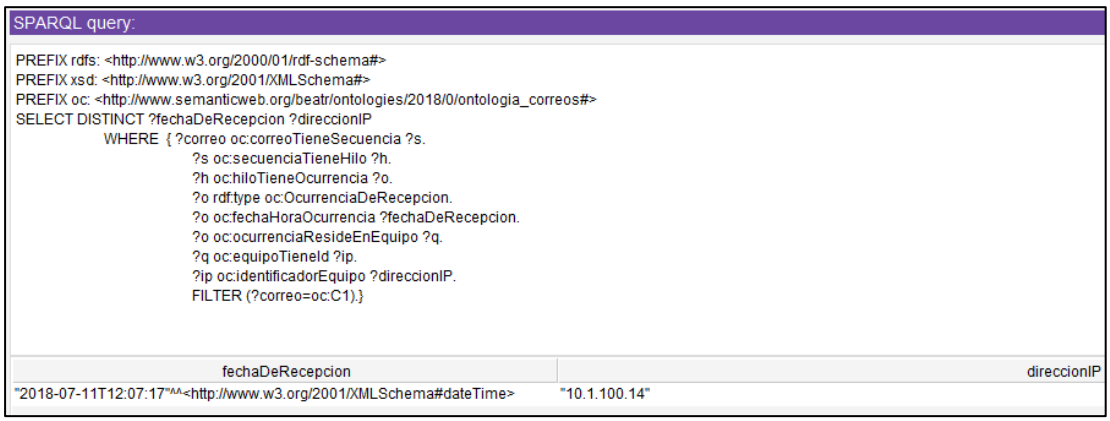


Figura 4-46.a: Resultados obtenidos por OntoFoCE para la pregunta PC02

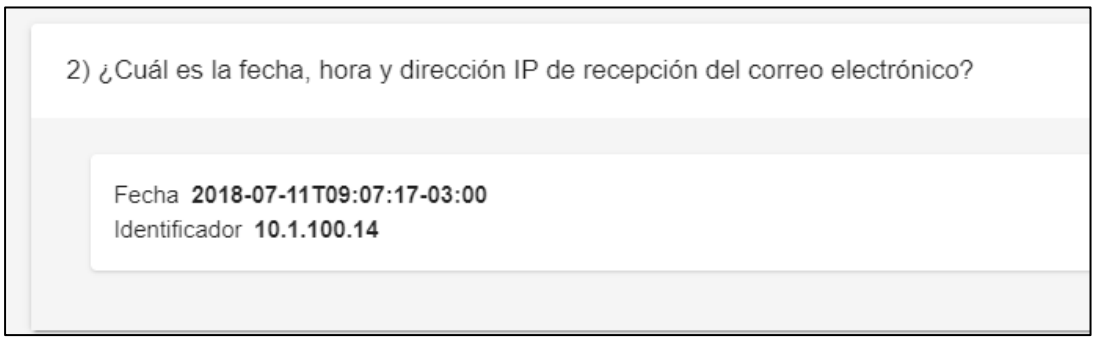


Figura 4-46.b: Resultados obtenidos por ObE Forensics para la pregunta PC02

- **PC03: Dado un correo CE ¿A qué cuentas se remitió?**

Para esta pregunta de competencia, en las Figuras 4-47.a y 4-47.b se muestran los resultados obtenidos por OntoFoCE y por ObE Forensics luego del procesamiento de datos del correo C1.

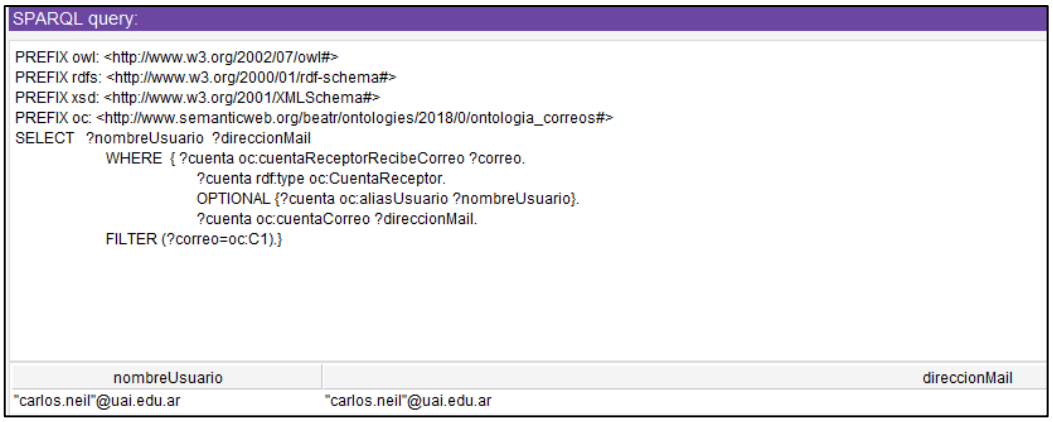


Figura 4-47.a: Resultados obtenidos por OntoFoCE para la pregunta PC03

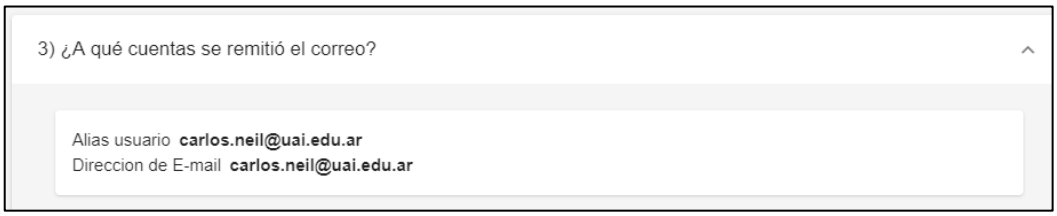


Figura 4-47.b: Resultados obtenidos por ObE Forensics para la pregunta PC03

- **PC04: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del emisor?**

En este caso, los resultados de OntoFoCE y ObE Forensics, también coincidentes, se muestran en las Figuras 4-48.a y 4-48.b:

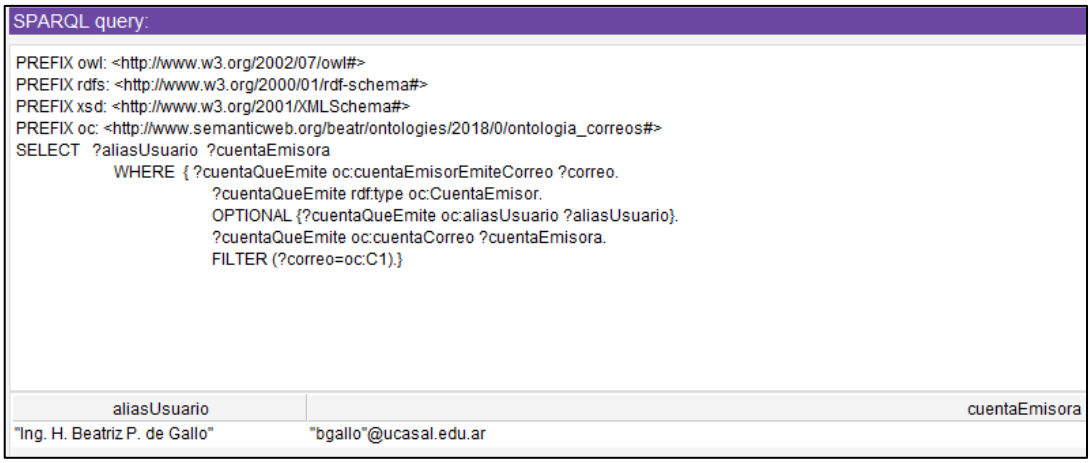


Figura 4-48.a: Resultados obtenidos por OntoFoCE para la pregunta PC04

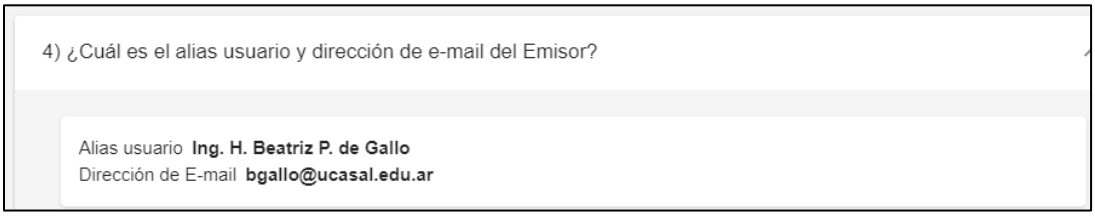


Figura 4-48.b: Resultados obtenidos por ObE Forensics para la pregunta PC04

- **PC05: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del Receptor?**

Para la pregunta de competencia PC05 los datos del receptor obtenidos con OntoFoCE y con ObE Forensics se muestran en la Figura 4-49.a y 4-49.b.

SPARQL query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT  ?aliasUsuario  ?cuentaReceptora
        WHERE  { ?cuentaQueRecibe oc:cuentaReceptorRecibeCorreo ?correo.
                 ?cuentaQueRecibe rdf:type oc:CuentaReceptor.
                 OPTIONAL {?cuentaQueRecibe oc:aliasUsuario ?aliasUsuario}.
                 ?cuentaQueRecibe oc:cuentaCorreo ?cuentaReceptora.
                 FILTER (?correo=oc:C1).}
```

aliasUsuario	cuentaReceptora
"carlos.neil"@uai.edu.ar	"carlos.neil"@uai.edu.ar

Figura 4-49.a: Resultados obtenidos por OntoFoCE para la pregunta PC05

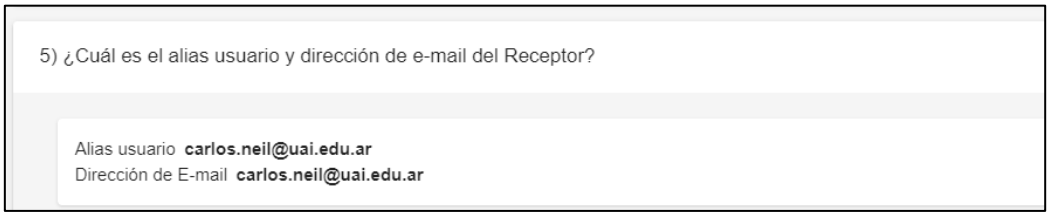


Figura 4-49.b: Resultados obtenidos por ObE Forensics para la pregunta PC05

- **PC06: Dado un correo CE ¿Cuál fue el cliente de correo utilizado por cada usuario?**

Las Figuras 4-50.a y 4-50.b muestran los resultados de esta pregunta de competencia obtenidos en ambos espacios.

SPARQL query:

```
SELECT DISTINCT  ?cuenta ?usuario ?cliente
        WHERE  {{ ?cuenta_receptor oc:cuentaReceptorRecibeCorreo ?correo.
                  ?cuenta_receptor oc:cuentaCorreo ?cuenta.
                  ?cuenta_receptor oc:recibeEn ?cli.
                  ?cli oc:nombreClienteCorreo ?cliente.
                  ?cuenta_receptor rdf:type oc:CuentaReceptor.
                  OPTIONAL {?cuenta_receptor oc:aliasUsuario ?usuario}.}
        UNION
                   { ?cuenta_emisor oc:cuentaEmisorEmiteCorreo ?correo.
                  ?cuenta_emisor oc:cuentaCorreo ?cuenta.
                  ?cuenta_emisor oc:emiteDesde ?cli.
                  ?cli oc:nombreClienteCorreo ?cliente.
                  ?cuenta_emisor rdf:type oc:CuentaEmisor.
                  OPTIONAL {?cuenta_emisor oc:aliasUsuario ?usuario}.}
        FILTER (?correo=oc:C1).}
```

cuenta	usuario	cliente
"carlos.neil"@uai.edu.ar	"carlos.neil"@uai.edu.ar	"Outlook"
"bgallo"@ucasal.edu.ar	"Ing. H. Beatriz P. de Gallo"	"ThunderBird"

Figura 4-50.a: Resultados obtenidos por OntoFoCE para la pregunta PC06

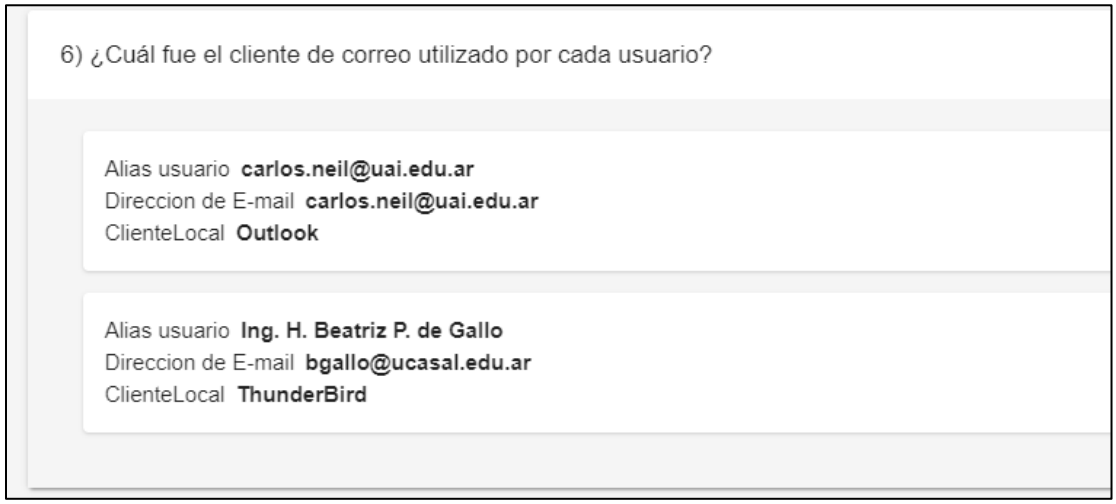


Figura 4-50.b: Resultados obtenidos por ObE Forensics para la pregunta PC06

- **PC07: Dado un correo CE ¿Cuál fue el equipo desde el cual se emitió el correo?**

Los resultados de la pregunta de competencia PC07, que muestra los datos del equipo emisor, se muestran en las Figuras 4-51.a y 4-51.b para OntoFoCE y ObE Forensics respectivamente.

Figura 4-51.a: Resultados obtenidos por OntoFoCE para la pregunta PC07

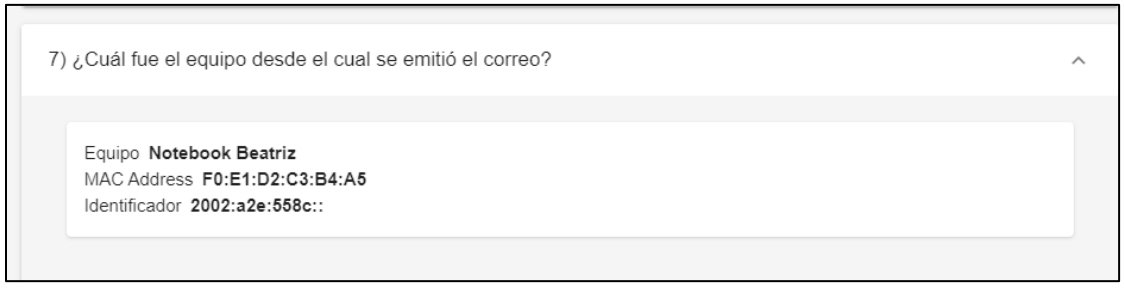


Figura 4-51.b: Resultados obtenidos por ObE Forensics para la pregunta PC07

- **PC08: Dado un correo CE ¿Cuál fue el equipo en el que se recibió el correo?**

Esta pregunta se responde desde OntoFoCE y ObE Forensics, según los resultados que muestran en las Figuras 4-52.a y 4-52.b.

SPARQL query:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?descripcionEquipo ?macAddress ?direccion
        WHERE { ?correo oc:correoTieneSecuencia ?s.
                ?s oc:secuenciaTieneHilo ?h.
                ?h oc:hiloTieneOcurrencia ?o.
                ?o rdf:type oc:OcurrenciaDeRecepcion.
                ?o oc:ocurrenciaResideEnEquipo ?equipo.
                ?equipo oc:macAddressEquipo ?macAddress.
                ?equipo oc:equipoTieneId ?IP.
                ?IP oc:identificadorEquipo ?direccion.
                ?equipo oc:descripcionEquipo ?descripcionEquipo.
                FILTER (?correo=oc:C1).}
```

descripcionEquipo	macAddress	direccion
"PC Carlos"	"F5:E2:D3:C4:B5:00"	"10.1.100.14"

Figura 4-52.a: Resultados obtenidos por OntoFoCE para la pregunta PC08

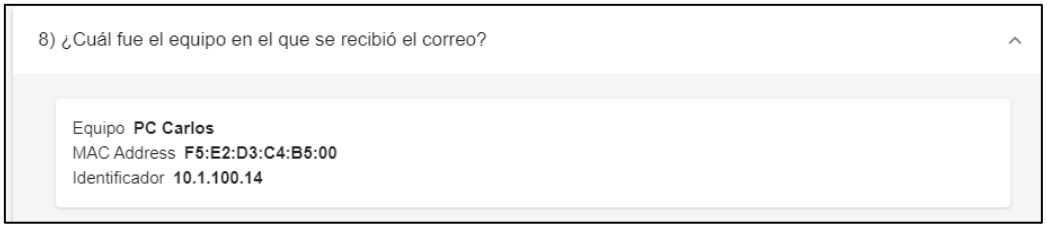


Figura 4-52.b: Resultados obtenidos por ObE Forensics para la pregunta PC08

- **PC09: Dado un correo, un emisor y un receptor ¿cuál es la secuencia de equipos por los que pasó ese correo?**

Para esta pregunta de competencia, los resultados obtenidos en OntoFoCE y ObE Forensics se muestran en las Figuras 4-53.a y 4-53.b.

SPARQL query:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?ocurrencias ?fecha  ?direccionIP  ?descripcionEquipo
WHERE {     ?emisor oc:cuentaEmisorEmiteCorreo ?correo.
            ?emisor oc:cuentaEUtilizaEquipoE ?equipoE.
            ?receptor oc:cuentaReceptorRecibeCorreo ?correo.
            ?receptor oc:cuentaRUtilizaEquipoR ?equipo.
            ?o oc:ocurrenciaResideEnEquipo ?equipo.
            ?correo oc:correoTieneSecuencia ?s1.
            ?s1 oc:secuenciaTieneHilo ?hilo1.
            ?hilo1 oc:hiloTieneOcurrencia ?ocurrencias.
            ?ocurrencias oc:ocurrenciaResideEnEquipo ?equipo2.
            ?equipo2 oc:equipoTieneId ?IP.
            ?equipo2 oc:descripcionEquipo ?descripcionEquipo.
            ?IP oc:identificadorEquipo ?direccionIP.
```

ocurrencias	fecha	direccionIP	descripcionEquipo
OE	"2018-07-11T12:07:06"^^<http://www.w3.org/2001/XMLSchema#dateTi	"2002:a2e:658d::"	"Notebook Beatriz"
OT1	"2018-07-11T12:07:09"^^<http://www.w3.org/2001/XMLSchema#dateTi	"mail-lj1-f180.google.com"	"servidor 1"
OT3	"2018-07-11T12:07:09"^^<http://www.w3.org/2001/XMLSchema#dateTi	"mail.ucasal.edu.ar"	"servidor 3"
OT2	"2018-07-11T12:07:09"^^<http://www.w3.org/2001/XMLSchema#dateTi	"209.85.208.180"	"servidor 2"
OT5	"2018-07-11T12:07:13"^^<http://www.w3.org/2001/XMLSchema#dateTi	"127.0.0.1"	"servidor 4"
OT4	"2018-07-11T12:07:13"^^<http://www.w3.org/2001/XMLSchema#dateTi	"127.0.0.1"	"servidor 4"
OT7	"2018-07-11T12:07:14"^^<http://www.w3.org/2001/XMLSchema#dateTi	"mail.ucasal.edu.ar"	"servidor 3"
OT6	"2018-07-11T12:07:14"^^<http://www.w3.org/2001/XMLSchema#dateTi	"127.0.0.1"	"servidor 4"
OT9	"2018-07-11T12:07:15"^^<http://www.w3.org/2001/XMLSchema#dateTi	"FNDEXCHG01.adm.vaneduc.edu.ar"	"servidor6"
OT8	"2018-07-11T12:07:15"^^<http://www.w3.org/2001/XMLSchema#dateTi	"200.10.180.145"	"servidor 5"
OR	"2018-07-11T12:07:17"^^<http://www.w3.org/2001/XMLSchema#dateTi	"10.1.100.14"	"PC Carlos"
OT10	"2018-07-11T12:07:17"^^<http://www.w3.org/2001/XMLSchema#dateTi	"10.1.100.15"	"servidor7"

Figura 4-53.a: Resultados obtenidos por OntoFoCE para la pregunta PC09

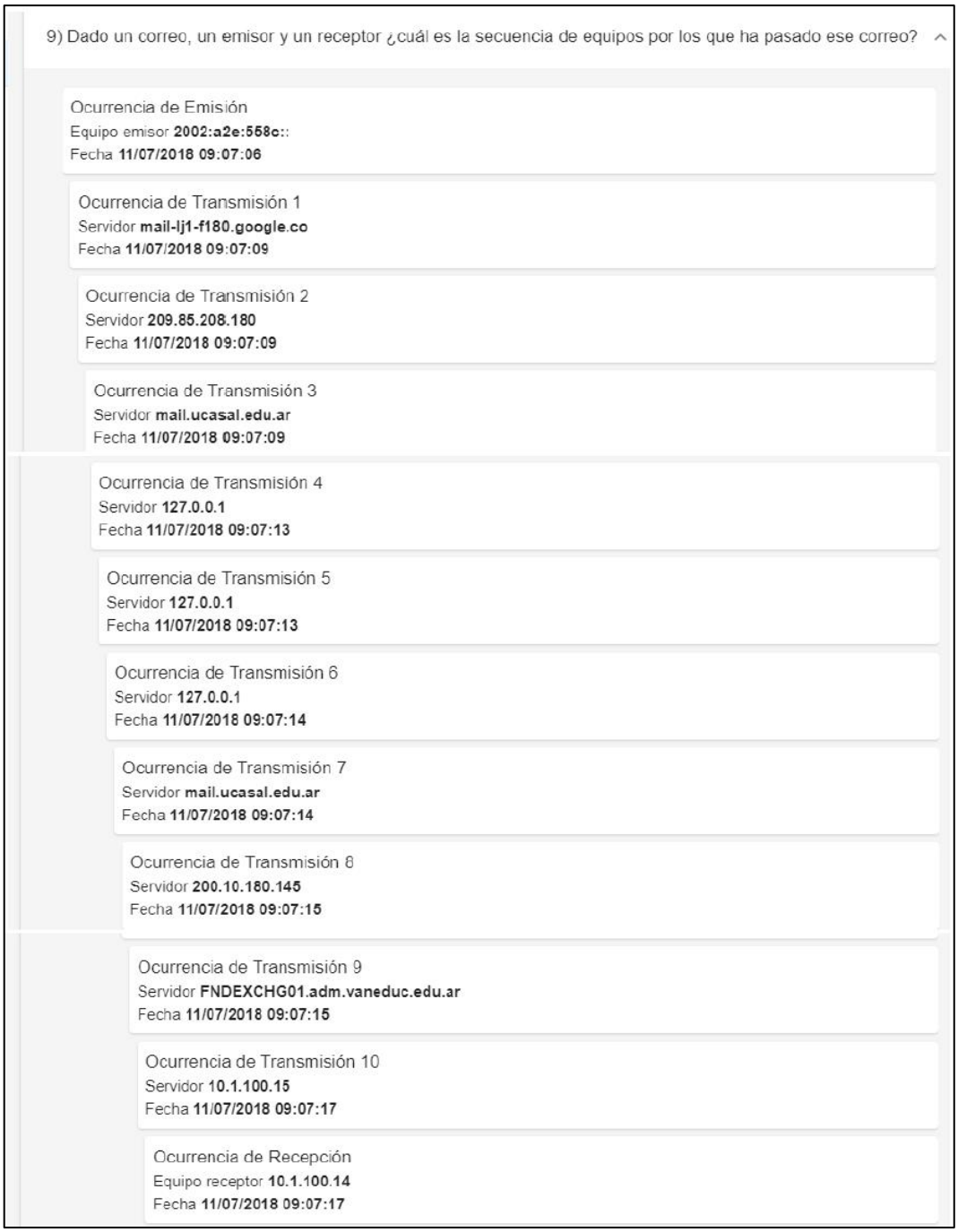


Figura 4-53.b: Resultados obtenidos por ObE Forensics para la pregunta PC09

Cabe mencionar que a fin de que los datos se visualicen en estricto orden cronológico, desde la aplicación se realiza un ordenamiento interno en base a la fecha de ocurrencia, cosa que no es posible realizar de manera directa en OntoFoCE.

Por último, es posible emitir el INFORME TÉCNICO PERICIAL, haciendo uso de la opción correspondiente de la pantalla inicial de ObE Forensics (ver Figura 4-20), mediante la cual accede a la opción de *Imprimir* según muestra la pantalla de la Figura 4-54. Así para el ejemplo que se está analizando, se selecciona el correo C1

en el panel de navegación de la izquierda, se selecciona las preguntas de competencia PC01 a PC09 en el panel de navegación del centro, y la aplicación muestra la vista preliminar del informe obtenido, el cual se adjunta en el ANEXO IV – INFORME TÉCNICO PERICIAL PARA ESCENARIO 1.

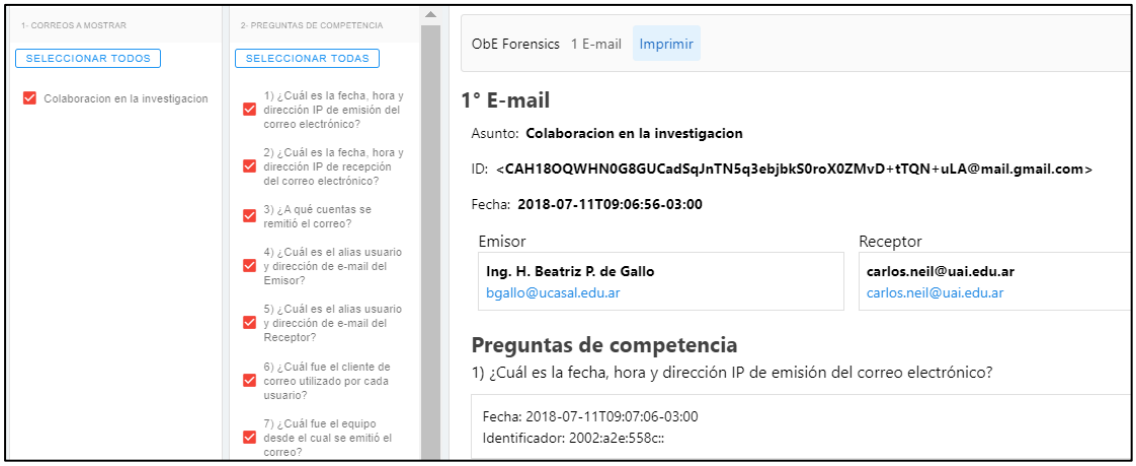


Figura 4-54: Pantalla de Menú de Opción IMPRIMIR para el Escenario 1

4.7.2 Análisis Forense de un Correo Electrónico enviado a Varios Receptores

En la sección anterior se mostró un caso de estudio en el que se realizó la pericia sobre un único correo, verificando los resultados obtenidos para las preguntas de competencia P01 a P09. Corresponde ahora ejemplificar la funcionalidad de ObE Forensics para las preguntas PC10, PC14 y PC15, para seguir la explicación del Escenario 2 que se incluye en la sección 3.5.2 del capítulo 3. En este caso se cargan los archivos de texto plano de los correos ya identificados en ese escenario como C1, el correo enviado a tres receptores, de quienes se obtienen las cabeceras CABECERA_R1, CABECERA_R2 y CABECERA_R3, y cuyo análisis desde OntoFoCE se indica en la antedicha sección del capítulo 3.Supuesto que el Perito procedió a cargar las cabeceras correspondientes y que accede al menú de las preguntas P10, P14 y P15, se muestran los resultados obtenidos por las instanciaciones en OntoFoCE y el análisis forense en ObE Forensic del correo C1 y las cabeceras indicadas.

- **PC10: Dado una cuenta C ¿cuáles son los correos que emitió?**

Para el ejemplo del escenario 2, se tomó esta pregunta puntual “Dada la cuenta bgallo@ucasal.edu.ar ¿cuáles con los correos que emitió?”.

Las respuestas de OntoFoCE y de ObE Forensics se muestran en las Figuras 4-55.a y 4-55.b.

SPARQL query:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo ?asunto ?fechaEmision ?nombreCuentaReceptor
        WHERE  { ?cuenta oc:cuentaEmisorEmiteCorreo ?correo.
                 ?cuentaRep oc:cuentaReceptorRecibeCorreo ?correo.
                 ?cuentaRep oc:cuentaCorreo ?nombreCuentaReceptor.
                 ?correo oc:correoTieneSecuencia ?s.
                 ?s oc:secuenciaTieneHilo ?h.
                 ?h oc:hiloTieneOcurrencia ?o.
                 ?o oc:ocurrenciaTieneAsunto ?a.
                 ?a oc:contenidoAsunto ?asunto.
                 ?o rdf:type oc:OcurrenciaDeEmision.
                 ?o oc:fechaHoraOcurrencia ?fechaEmision.
                 FILTER (?cuenta=oc:CUENTA_E1)}
```

correo	asunto	fechaEmision	nombreCuentaReceptor
C1	"Colaboración en investigación"	"2018-07-08T15:41:46"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"luzbibianaclara"@gmail.com
C1	"Colaboración en investigación"	"2018-07-08T15:41:46"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"Enzo Notario"
C1	"Colaboración en investigación"	"2018-07-08T15:41:46"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"erivetti83"@gmail.com

Figura 4-55.a: Resultados obtenidos por OntoFoCE para la pregunta PC10

En ObE Forensics, la pregunta de competencia PC10 requiere que se definan los parámetros del filtro, así, para el ejemplo en cuestión, la Figura 4-56.b muestra las dos pantallas: la primera en la que se indican los parámetros ingresados y la segunda en la que se muestran los resultados obtenidos.

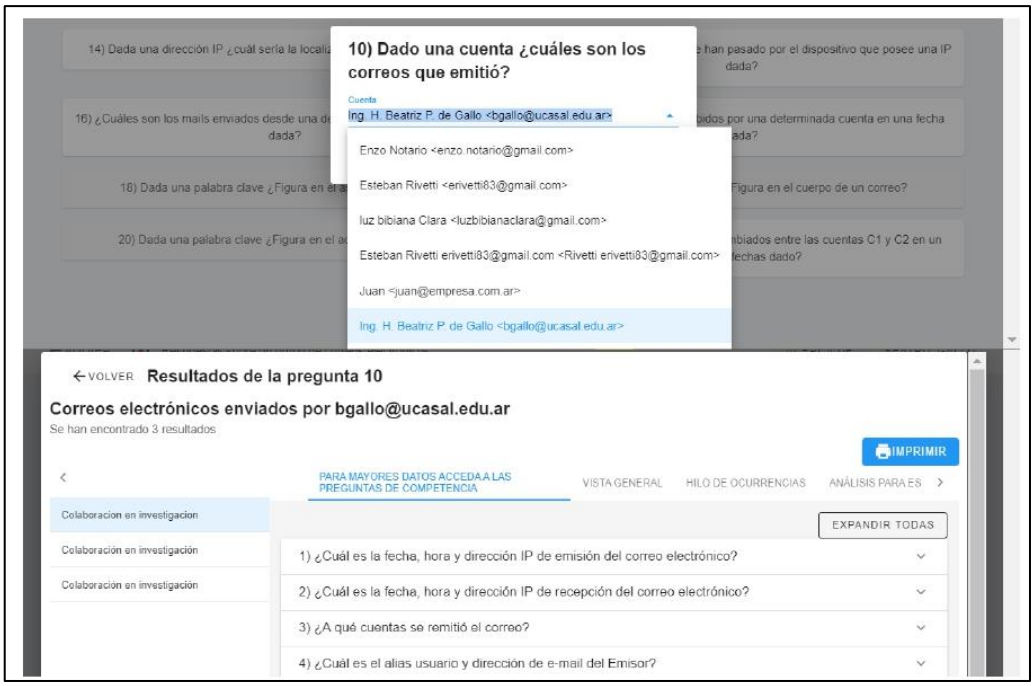


Figura 4-55.b: Resultados obtenidos por ObE Forensics para la pregunta PC10

Para el caso de ObE Forensics, los resultados de la pregunta de competencia muestran tres correos que cumplen la condición de la consulta, y para cada uno de ellos, la Figura 4-55.b muestra el detalle de las cuentas que actúan como emisor/receptor. Obsérvese que en estas preguntas, que pueden filtrar varios correos como resultado de la consulta, ObE Forensics le permite al Perito ingresar a cada correo y consultar sus datos desde las cuatro opciones ya señaladas para el análisis de un único correo: *Preguntas de Competencia*, *Vista General*, *Hilo de Ocurrencias* y *Análisis de la Trazabilidad.*

- **PC14: Dada una dirección IP ¿cuál sería la localización geográfica del equipo asociado?**

Para esta pregunta, desde OntoFoCE la Figura 4-56.a muestra los resultados, mientras que para ObE Forensics, se muestran en la Figura 4-56.b.

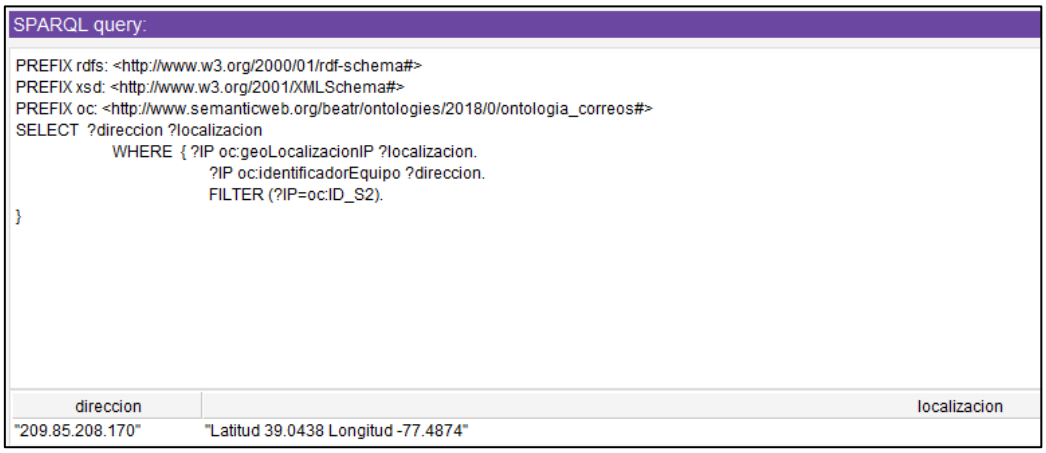


Figura 4-56.a: Resultados obtenidos por OntoFoCE para la pregunta PC14

Desde ObE Forensics, la aplicación se conecta al servicio web ofrecido desde el sitio de http://ip-api.com/docs/api:json, obteniendo más información que la que se instanció en OntoFoCE para la dirección IP 209.85.208.170.

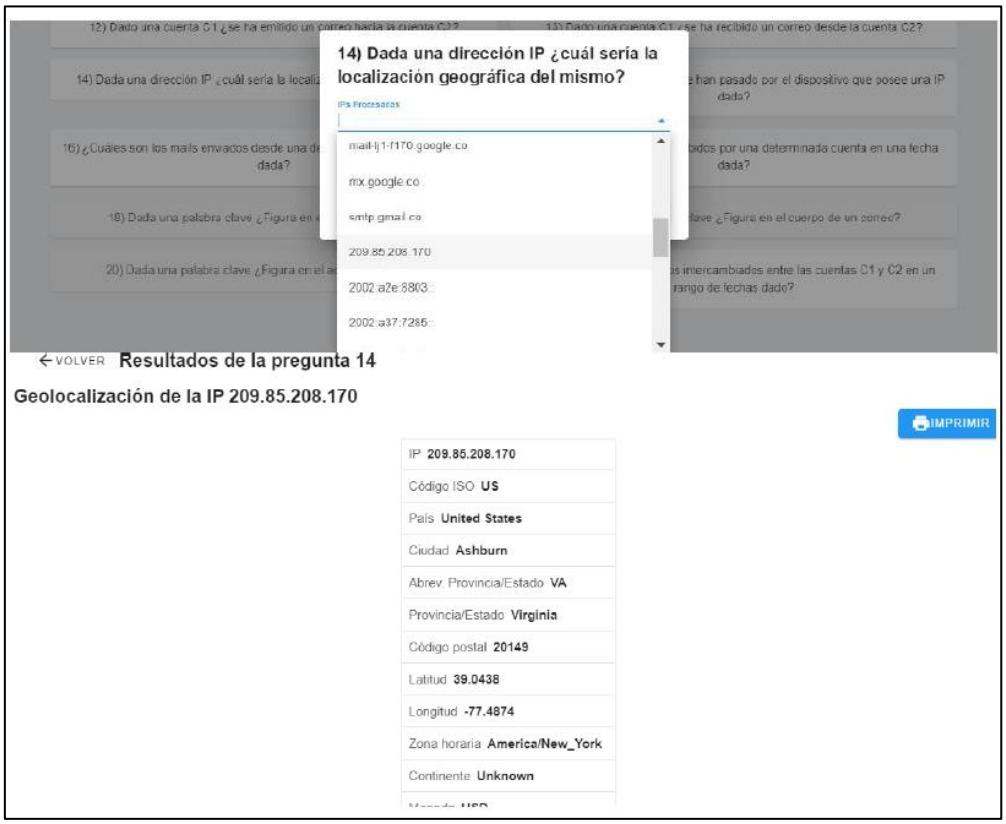


Figura 4-56.b: Resultados obtenidos por ObE Forensics para la pregunta PC14

- **PC15: ¿Cuáles son los correos que han pasado por el dispositivo que posee una IP dada?**

En este caso, la pregunta sería "¿Cuáles son los correos que han pasado por el dispositivo que posee la dirección IP 209.85.208.170?", y la respuesta desde OntoFoCE se muestra en la pantalla de la Figura 4-57.a, mientras que los resultados de procesar esta pregunta de competencia con los mismos datos en ObE Forensics, se muestran en la pantalla de la Figura 4-57.b.

SPARQL query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT ?equipo ?Cabecera
        WHERE {         ?equipo oc:equipoTieneId ?IP.
                ?ocurrencia oc:ocurrenciaResideEnEquipo ?equipo.
                ?hilo oc:hiloTieneOcurrencia ?ocurrencia.
                ?s oc:secuenciaTieneHilo ?hilo.
                ?correo oc:correoTieneSecuencia ?s.
                ?e oc:cuentaEmisorEmiteCorreo ?correo.
                ?IP oc:identificadorEquipo ?direccion.
                ?ocurrencia oc:ocurrenciaTieneCabecera ?Cabecera.
                FILTER regex(?direccion,"209.85.208.170").}
```

equipo	Cabecera
SERVIDOR_2	CABECERA_R3
SERVIDOR_2	CABECERA_R1
SERVIDOR_2	CABECERA_R2

Figura 4-57.a: Resultados obtenidos por OntoFoCE para la pregunta PC15

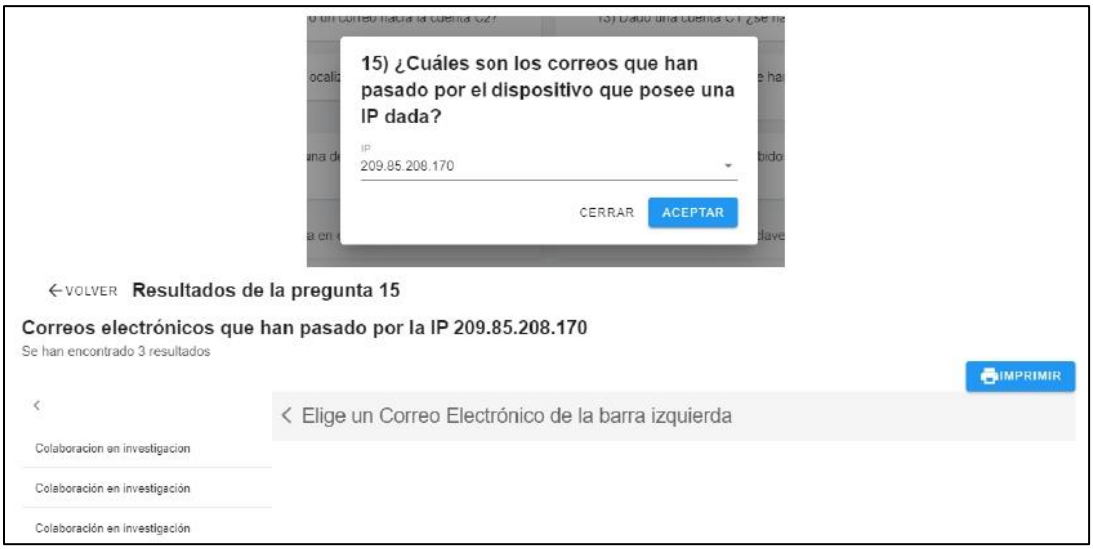


Figura 4-57.b: Resultados obtenidos por ObE Forensics para la pregunta PC15

Ahora, el Perito puede ingresar a cada pregunta de competencia y emitir el informe impreso correspondiente a cada una, estos documentos para los ejemplos del

Escenario 2 se muestran en el ANEXO V – INFORME TÉCNICO PERICIAL ANÁLISIS ESCENARIO 2.

4.7.3 Análisis Forense de un Conjunto de Correos

Continuando con el desarrollo de presentar las preguntas de competencias en términos de escenarios periciales, corresponde abordar el Escenario 3, ya descripto en el capítulo 3, en donde se analizan un conjunto de correos intercambiados entre varias cuentas. Considerando el ejemplo de dos cuentas emisores y tres cuentas receptoras, que se tomó como ejemplo en el citado caso de estudio, Escenario 3, se describirán los puntos de pericia, describiendo los resultados brindados mediante OntoFoCE con los que devuelve la aplicación ObE Forensics.

Resumiendo, el Escenario 3 plantea una pericia en la que se intercambian siete correos entre cuatro cuentas distintas. Una de esas cuentas actúa como cuenta emisora y receptora en el mismo escenario, y el resto como cuentas emisoras y/o receptoras. Supuesto que el perito realizó la extracción de las siete cabeceras de correos electrónicos, y las procesará con ObE Forensics, la Figura 4.58 muestra la carga de los siete correos considerados y sobre los cuales se realizarán la verificación de las preguntas de competencia.

Figura 4-58: Pantalla de Carga de las Siete Cabeceras del Escenario 3

A continuación, se comparan los resultados de las preguntas de competencia obtenidos desde la instanciación en OntoFoCE de los casos de ejemplo, con los resultados emitidos por ObE Forensics a partir de la carga de las cabeceras correspondiente al Escenario 3.

Debe considerarse que estas preguntas cuentan con una primera pantalla de selección de parámetros de filtrado de datos.

- **PC11: Dado una cuenta C ¿cuáles son los correos que recibió?**

Esta pregunta de competencia toma como ejemplo, los correos recibidos por la cuenta *enzo.notario@gmail.com*. Las Figuras 4-59.a y 4-59.b muestran los resultados de la consulta en OntoFoCE y en ObE Forensics. La segunda Figura también muestra una vista parcial de la pantalla de resultados, pero se observa que se seleccionan los dos correos, al igual que la consulta realizada en OntoFoCE.

SPARQL query:

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo ?asunto ?fechaEmision  ?cuentaEmisora
                WHERE  { ?cuenta oc:cuentaReceptorRecibeCorreo ?correo.
                         ?correo oc:correoTieneSecuencia ?s.
                         ?correo oc:correoEsEmitidoPorCuentaEmisor ?ctaEmisor.
                         ?ctaEmisor oc:cuentaCorreo ?cuentaEmisora.
                         ?s oc:secuenciaTieneHilo ?h.
                         ?h oc:hiloTieneOcurrencia ?o.
                         ?o oc:ocurrenciaTieneAsunto ?a.
                         ?a oc:contenidoAsunto ?asunto.
                         ?o rdf:type oc:OcurrenciaDeRecepcion.
                         ?o oc:fechaHoraOcurrencia ?fechaEmision.
                         FILTER (?cuenta=oc:CUENTA_ENZO).}
```

correo	asunto	fechaEmision	cuentaEmisora
C2	"Re: Trabajo para el CoNaIISI"	"2018-10-02T03:01:22"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"juan"@empresa.com.ar
C1	"Colaboración en investigación"	"2018-07-08T15:41:59"^^<http://www.w3.org/2001/XMLSchema#dateTime>	"bgallo"@ucasal.edu.ar

Figura 4-59.a: Resultados obtenidos por OntoFoCE para la pregunta P11

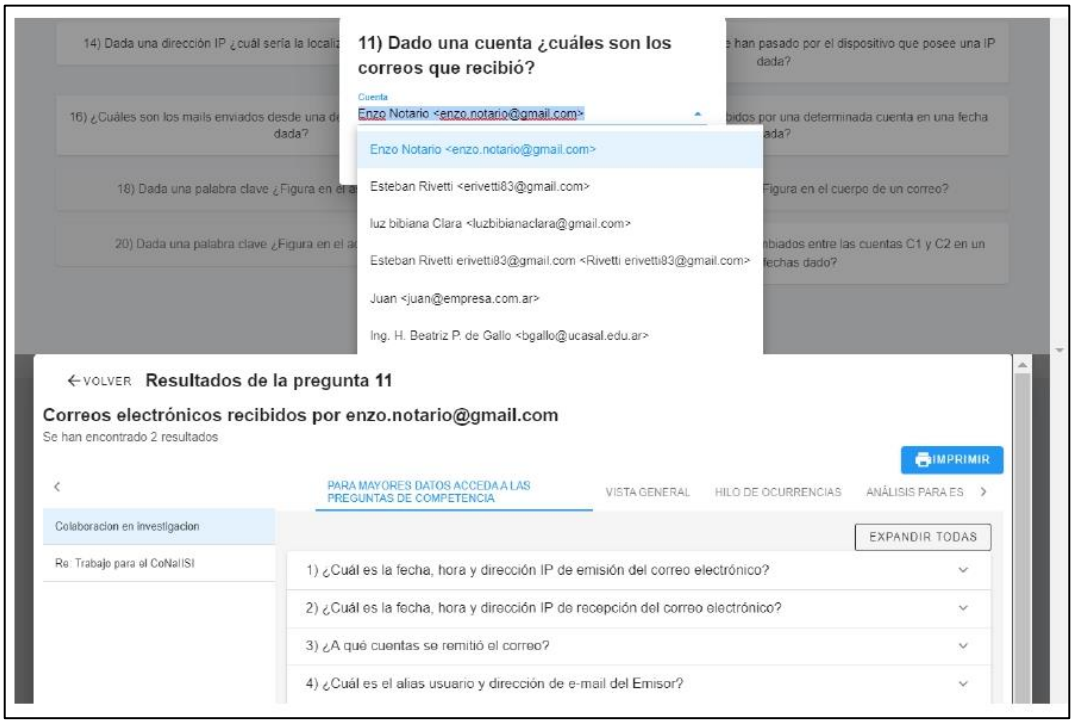


Figura 4-59.b: Resultados obtenidos por ObE Forensics para la pregunta P11

- **PC12: Dado una cuenta C1 ¿se ha emitido un correo hacia la cuenta C2?**

Esta pregunta de competencia se trabajó en OntoFoCE tomando como ejemplo si la cuenta *juan@empresa.com.ar* se emitió un único correo a la cuenta *erivetti83@gmail.com*.

El mismo ejemplo se probó en ObE Forensics, y los resultados se muestran en las pantallas de las Figuras 4-60.a y 4-60.b.

SPARQL query:

```
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo ?asunto ?fecha
        WHERE  { ?cuentaC1 oc:cuentaEmisorEmiteCorreo ?correo.
                 ?cuentaC2 oc:cuentaReceptorRecibeCorreo ?correo.
                 ?cuentaC1 oc:cuentaCorreo ?cuentaEmisora.
                 ?cuentaC2 oc:cuentaCorreo ?cuentaReceptor.
                 ?correo oc:correoTieneSecuencia ?s.
                 ?s oc:secuenciaTieneHilo ?h.
                 ?h oc:hiloTieneOcurrencia ?o.
                 ?o rdf:type oc:OcurrenciaDeEmision.
                 ?o oc:ocurrenciaTieneAsunto ?a.
                 ?o oc:fechaHoraOcurrencia ?fecha.
                 ?a oc:contenidoAsunto ?asunto.
        FILTER (?cuentaC1=oc:CUENTA_JUAN &&  ?cuentaC2=oc:CUENTA_ESTEBAN).}
```

correo	asunto	fecha
C3	"Fwd: Forensia IoT"	"2019-02-23T16:27:32"^^<http://www.w3.org/2001/XMLSchema#dateTime>

Figura 4-60.a: Resultados obtenidos por OntoFoCE para la pregunta P12

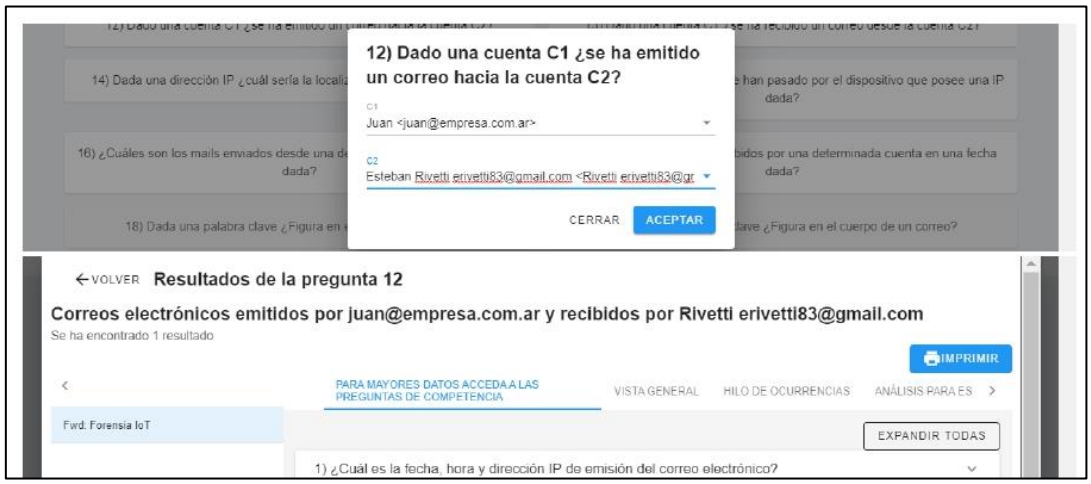


Figura 4-60.b: Resultados obtenidos por ObE Forensics para la pregunta P12

- **PC13: Dado una cuenta C1 ¿se ha recibido un correo desde la cuenta C2?**

En OntoFoCE esta pregunta de competencia indagó si en la cuenta *luzbibianaclara@gmail.com* recibió un correo enviado desde la cuenta *juan@empresa.com.ar*.

Ahora, se realiza la misma consulta en ObE Forensics, y se muestran los resultados de ambas consultas en las Figuras 4-61.a y 4-61.b.

SPARQL query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT DISTINCT  ?correo ?asunto ?fecha
        WHERE  { ?cuentaC1 oc:cuentaReceptorRecibeCorreo ?correo.
                 ?cuentaC2 oc:cuentaEmisorEmiteCorreo ?correo.
                 ?correo oc:correoTieneSecuencia ?s.
                 ?s oc:secuenciaTieneHilo ?h.
                 ?h oc:hiloTieneOcurrencia ?o.
                 ?o rdf:type oc:OcurrenciaDeRecepcion.
                 ?o oc:fechaHoraOcurrencia ?fecha.
                 ?o oc:ocurrenciaTieneAsunto ?a.
                 ?a oc:contenidoAsunto ?asunto.
        FILTER (?cuentaC1=oc:CUENTA_LUZ &&  ?cuentaC2=oc:CUENTA_JUAN).}
```

co...	asunto	fecha
C4	"Reenviar: ACM TIOT Call for Papers"	"2018-06-05T23:57:02"^^<http://www.w3.org/2001/XMLSchema#dateTime>

Figura 4-61.a: Resultados obtenidos por OntoFoCE para la pregunta P13

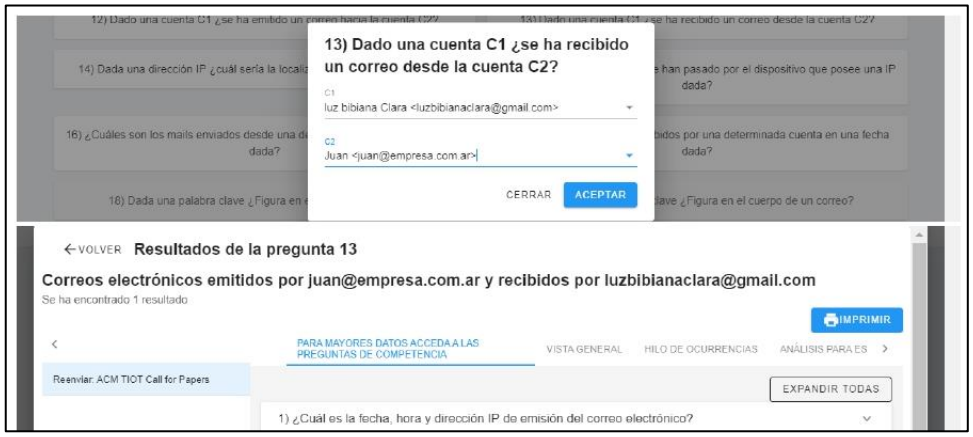


Figura 4-61.b: Resultados obtenidos por ObE Forensics para la pregunta P13

- **PC16: ¿Cuáles son los mails enviados desde una determinada cuenta en una fecha dada?**

Esta pregunta de competencia filtra los correos enviados desde una cuenta según una fecha determinada, en OntoFoCE se tomó como ejemplo la cuenta *juan@empresa.com.ar* en una fecha dada (02/10/2018), el mismo ejemplo se presenta en ObE Forensics, y los resultados se muestran en las Figuras 4-62.a y 4-62.b.

SPARQL query:

```
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT  DISTINCT ?correo  ?cuentaReceptora ?asunto ?fecha
        WHERE {    ?cuenta rdf:type oc:CuentaEmisor.
                   ?cuenta oc:cuentaEmisorEmiteCorreo ?correo.
                   ?correo oc:correoTieneSecuencia ?s.
                   ?s oc:secuenciaTieneHilo ?h.
                   ?Ocurrencia oc:ocurrenciaCorrespondeAHilo ?h.
                   ?Ocurrencia oc:ocurrenciaTieneAsunto ?a.
                   ?a oc:contenidoAsunto ?asunto.
                   ?Ocurrencia rdf:type oc:OcurrenciaDeEmision.
                   ?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
                   ?cuentaR oc:cuentaReceptorRecibeCorreo ?correo.
                   ?cuentaR oc:cuentaCorreo ?cuentaReceptora.
                   FILTER(("2018-10-02T03:01:21"^^xsd:dateTime=?fecha) &&
(?cuenta=oc:CUENTA_JUAN)).}
```

correo	cuentaReceptora	asunto	fecha
C2	"Enzo Notario"	"Re: Trabajo para el CoNallSI"	"2018-10-02T03:01:21"^^<http://www.w3.org/2001/XMLSchema#dateTime>

Figura 4-62.a: Resultados obtenidos por OntoFoCE para la pregunta P16

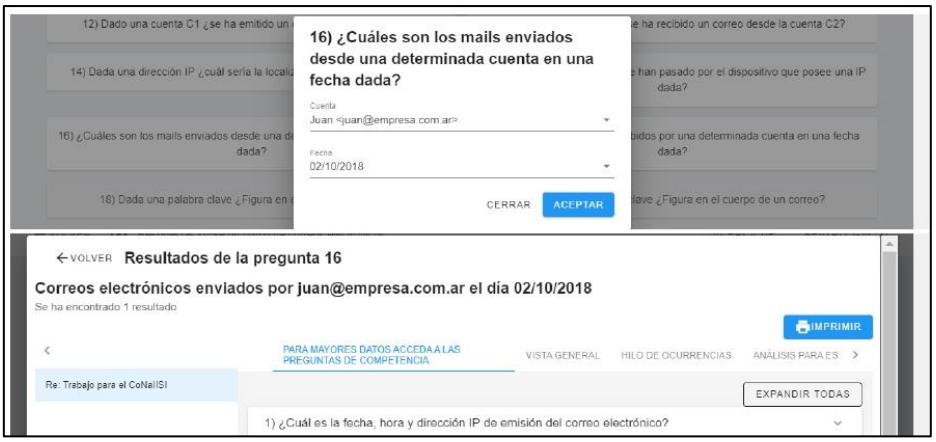


Figura 4-62.b: Resultados obtenidos por ObE Forensics para la pregunta P16

- **PC17: ¿Cuáles son los mails recibidos por una determinada cuenta en una fecha dada?**

Para esta pregunta de competencia el parámetro de fecha del ejemplo es *8 Jul 2018 15:41:46* y la cuenta es *erivetti83@gmail.com*. Las Figuras 4-63.a y 4-63.b muestran los resultados para OntoFoCE y para ObE Forensics.

SPARQL query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX oc: <http://www.semanticweb.org/beatr/ontologies/2018/0/ontologia_correos#>
SELECT  DISTINCT  ?correo ?cuentaEmisora  ?cuentaReceptora ?asunto ?fecha
          WHERE {    ?cuenta rdf:type oc:CuentaEmisor.
                     ?cuenta oc:cuentaCorreo ?cuentaEmisora.
                     ?cuenta oc:cuentaEmisorEmiteCorreo ?correo.
                     ?correo oc:correoTieneSecuencia ?s.
                     ?s oc:secuenciaTieneHilo ?h.
                     ?Ocurrencia oc:ocurrenciaCorrespondeAHilo ?h.
                     ?Ocurrencia oc:ocurrenciaTieneAsunto ?a.
                     ?a oc:contenidoAsunto ?asunto.
                     ?Ocurrencia rdf:type oc:OcurrenciaDeEmision.
                     ?Ocurrencia oc:fechaHoraOcurrencia ?fecha.
                     ?cuentaR oc:cuentaReceptorRecibeCorreo ?correo.
                     ?cuentaR oc:cuentaCorreo ?cuentaReceptora
```

correo	cuentaEmisora	cuentaReceptora	asunto	fecha
C1	"bgallo"@ucasal.edu.ar	"erivetti83"@gmail.com	"Colaboración en investigación"	"2018-07-08T15:41:46"^^<http://www.w3.org/2001/XMLSchema#dateTime>

Figura 4-63.a: Resultados obtenidos por OntoFoCE para la pregunta P17

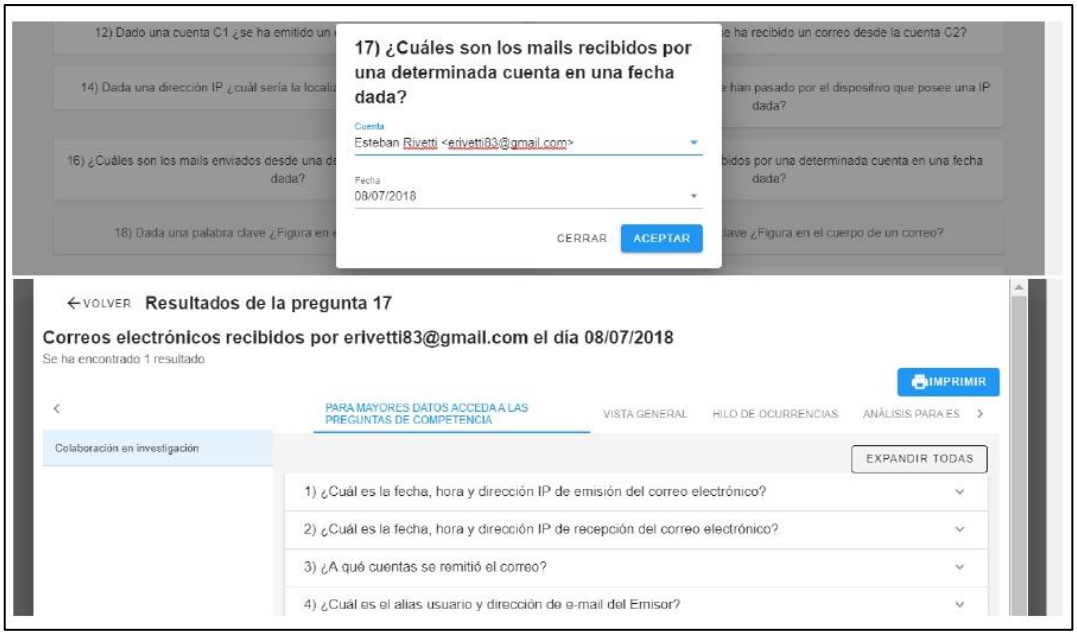


Figura 4-63.b: Resultados obtenidos por ObE Forensics para la pregunta P17

Tal como se indicó en el apartado 3.5. del capítulo 3, las preguntas referidas a la búsqueda de palabras claves, PC18, PC19 y PC20, requieren del desarrollo del algoritmo de búsqueda de dichos términos en las instancias generadas para el *Cuerpo* de cada correo, que usualmente se encripta, además de resolver la problemática de analizar un archivo adjunto que puede estar en distintos formatos en los que no se puede identificar palabras o textos (imagen, audio, entre otros). Por esta razón, estas preguntas no se han implementado todavía en ObE Forensics.

- **PC21: ¿Cuáles son los correos intercambiados entre las cuentas C1 y C2 en un rango de fechas dado?**

Para esta pregunta de competencia mantengamos el ejemplo del capítulo 3, con las cuentas *juan@empresa.com.ar* y *enzo.notario@gmail.com*, entre el período del 05/08/2018 al 23/02/2019.

Los resultados se muestran en las pantallas de las Figuras 4-64.a y 4-64.b.

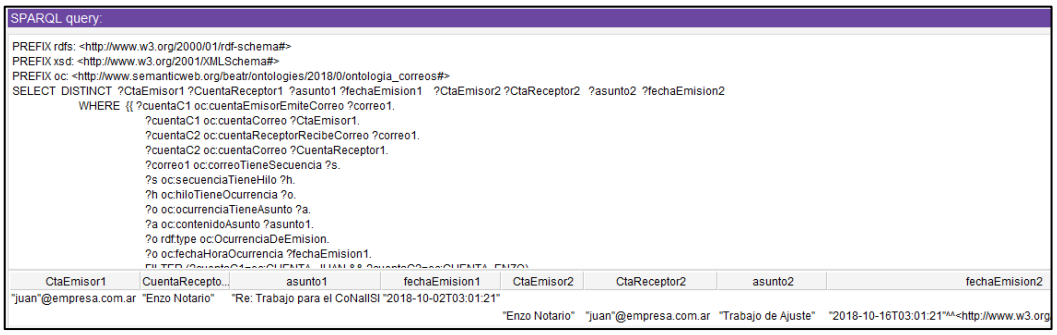


Figura 4-64.a: Resultados obtenidos por OntoFoCE para la pregunta P21

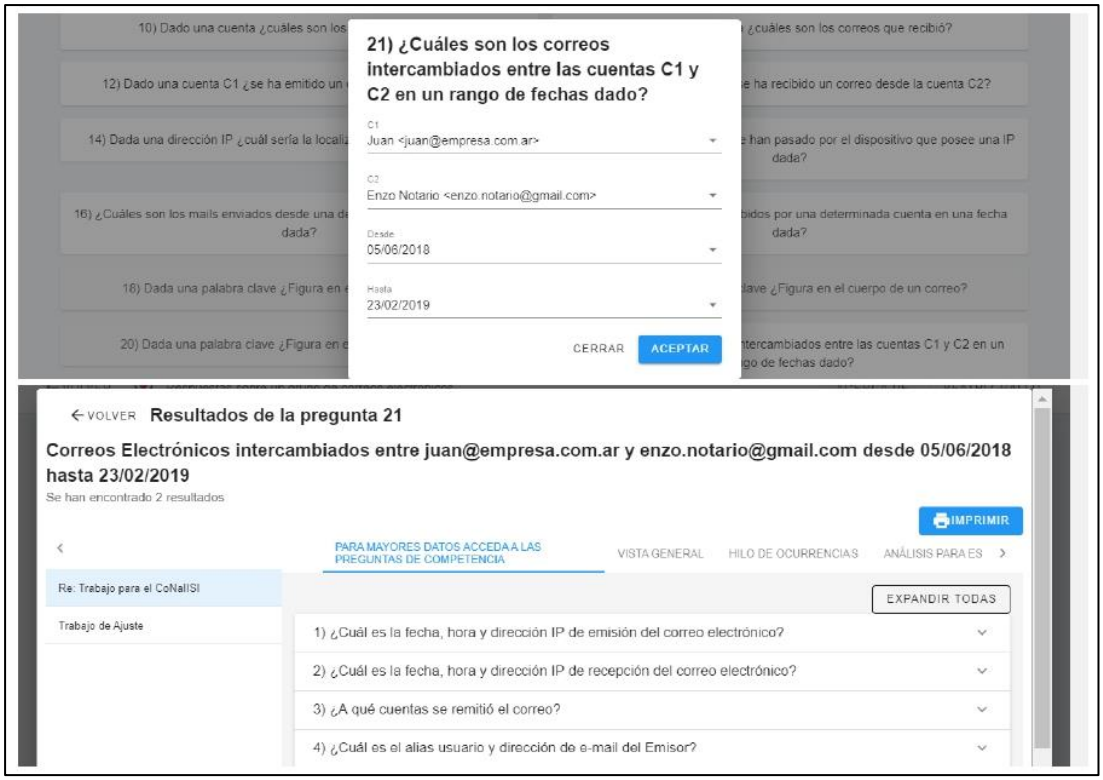


Figura 4-64.b: Resultados obtenidos por ObE Forensics para la pregunta P21

Por último, el Perito emite el informe impreso correspondiente a cada pregunta de competencia. Estos documentos para los ejemplos del ESCENARIO 3 se muestran en el ANEXO VI – INFORME TÉCNICO PERICIAL ANÁLISIS ESCENARIO 3.

4.7.4 Análisis Forense de Varias Cuentas de Correo

En las secciones anteriores se mostraron casos de estudio en el que se realizó la pericia sobre una o varias cuentas, detallando el funcionamiento de la aplicación para realizar las consultas de las preguntas de competencia y/o indicando los filtros que permiten buscar y seleccionar aquellos correos que cumplen con la condición de filtrado. Pero en los ejemplos señalados se trabajó con pocos correos, ¿cómo se comporta la aplicación cuando se debe analizar una cantidad importante de correos?

Para estos casos se presenta un Escenario 4 en el que supongamos se debe analizar un conjunto de correos pertenecientes todos a la misma cuenta. El Perito realiza el procedimiento indicado para la realización de la pericia siguiendo todas las

fases indicadas hasta obtenerla cuenta de correo ejemplo, la cual contiene 576 correos electrónicos, todos ellos pertenecientes a una persona a la que denominaremos JUAN.

Cabe señalar que los correos se obtuvieron de una cuenta real, y a fin de mantener la reserva de la identidad y de los datos consignados en los mismos, se procedió a anonimizar las cuentas, asuntos y demás datos de las cabeceras, reemplazando el valor real del Asunto en la frase genérica *asuntoXXX* donde XXX es un número secuencial de 1 a 576, y las cuentas se reemplazaron por el valor *usuarioXXX@empresa.com.ar*, en donde también XXX representa una secuencia numérica única para cada cuenta de emision/recepción encontrada en los 576 correos, con excepción de dos cuentas en particular a las que se identificó como *juan@empresa.com.ar* y *pedidos@empresa.com.ar* y que se utilizarán para ejemplificar el funcionamiento de ObE Forensics en este escenario. La pantalla que se indica en la Figura 4-65 muestra la carga de los 576 correos en la aplicación.

Figura 4-65: Pantalla de carga de los 576 correos del Escenario 4

Sobre este contexto, las preguntas de competencia PC10 a PC21 toman otro valor, ya que permiten realizar el filtrado de los correos con prontitud, y es posible luego profundizar un correo en particular aplicando sobre el mismo las preguntas de competencia P01 a P09. Aquí es donde también se observa la versatilidad de la aplicación para seleccionar un conjunto determinado de correos, aplicar las consultas que se requieran e incluir los resultados en el informe pericial correspondiente. Valgan como ejemplo los siguientes casos.

Supongamos el siguiente punto de pericia: *Analice la computadora portátil aportada como prueba y verifique que desde allí se estableció correspondencia electrónica entre las cuentas juan@empresa.com.ar y pedidos@empresa.com.ar, considerando el conjunto de todos los correos existentes en la cuenta durante el período 01/01/2017 y 28/02/2017 y obtenga información acerca del intercambio de correos realizados y toda otra información de interés para la causa que pudiera recabar.*

Se indica un punto de pericia lo suficientemente amplio como para mostrar la funcionalidad de las preguntas de competencia que se realizan sobre un conjunto de correos. Esto permitiría que el Perito consulte las preguntas P10 a P21 y presente al Juez el INFORME TÉCNICO DE PERICIA con los resultados obtenidos. A continuación se muestran los resultados de estas preguntas, para el conjunto de los 576 correos.

- **PC 10: Dado una cuenta ¿Cuáles son los correos que emitió?**

En este caso, la cuenta a analizar es la de pedidos@empresa.com.ar, de modo que se selecciona ese valor en la pantalla de la pregunta P10 y se obtienen los resultados señalados en la Figura 4-66. Allí se observa que desde la cuenta indicada se enviaron 123 correos.

Figura 4-66: Pantalla de resultados para la PREGUNTA DE COMPETENCIA P10 del Escenario 4

- **PC 11: Dado una cuenta ¿cuáles son los correos que recibió?**

El proceso de la pregunta 11 es similar al de la pregunta 10 solo que se procesan los correos recibidos en la cuenta del usuario JUAN, observándose que se han recibido 307 correos electrónicos (Figura 4-67).

Figura 4-67: Pantalla de resultados para la PREGUNTA DE COMPETENCIA P11 del Escenario 4

- **PC 16: ¿Cuáles son los mails enviados desde una determinada cuenta en una fecha dada?**

Para procesar la pregunta de competencia P16 se solicita la selección de dos valores: una cuenta de correo y la fecha determinada, para la cual se requiere averiguar cuáles son los correos enviados. Suponiendo que interese averiguar los correos enviados desde pedidos@empresa.com.ar el día 07/02/17, las respuesta de esta pregunta, que se muestran en la Figura 4-68, indican que hubo siete correos enviados en esa fecha.

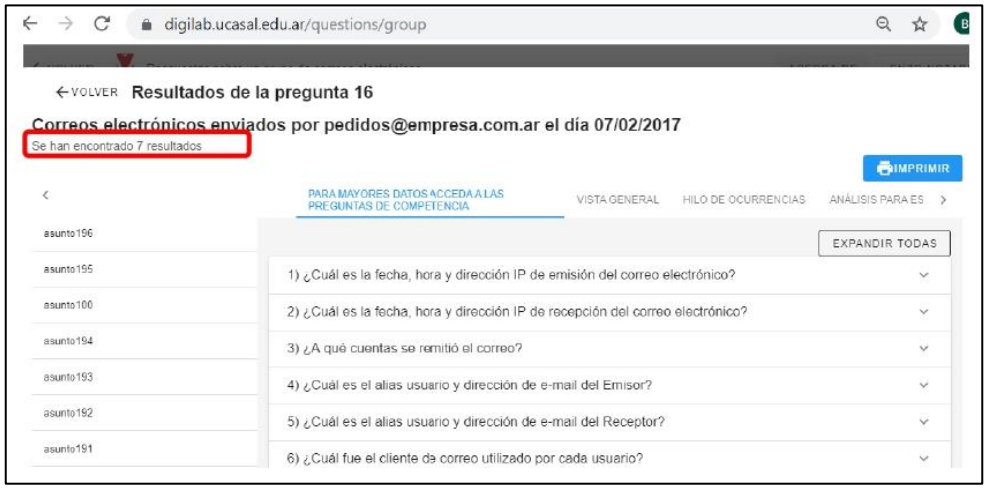


Figura 4-68: Pantalla de resultados para la PREGUNTA DE COMPETENCIA P16 del Escenario 4

- **PC 17: ¿Cuáles son los mails recibidos por una determinada cuenta en una fecha dada?**

De manera similar al proceso de la pregunta de competencia P16, aquí se solicita la selección de dos valores: una cuenta de correo y la fecha determinada, para la cual se requiere averiguar cuáles son los correos recibidos. Supongamos que se pregunte si el usuario JUAN recibió correos el 7/02/17, entonces la Figura 4-69 muestra que recibió 3 correos; de éstos, los valores *asunto100, asunto192* y *asunto191* que identifican a los correos distintos según el proceso de anonimización realizados, coinciden con los mismos datos que figura en la pantalla anterior (Figura 4-68), es decir, de los 7 correos enviados desde *pedidos@empresa.com.ar* se observa que al menos 3 fueron recibidos en la cuenta *juan@empresa.com.ar*.

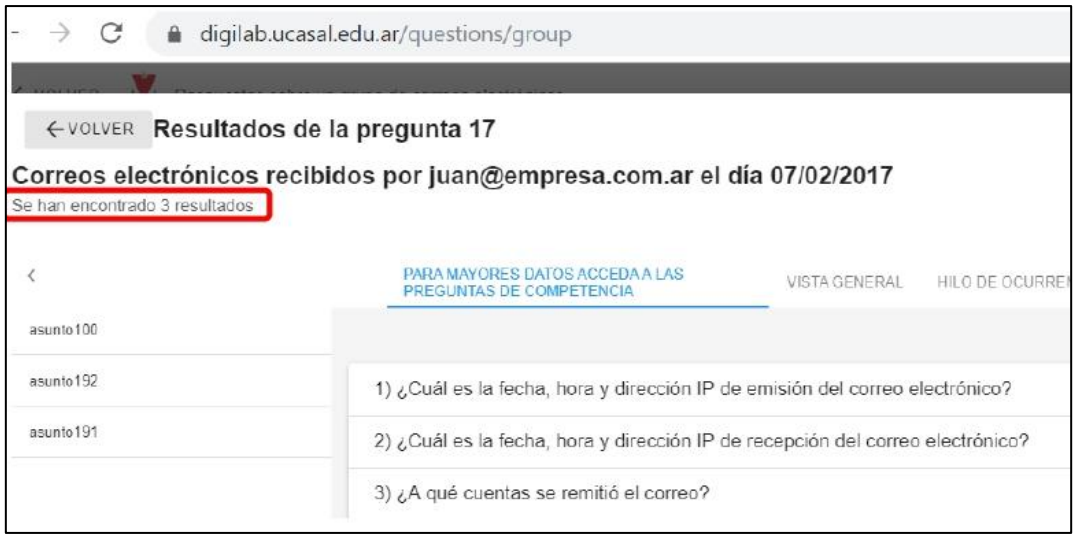


Figura 4-69: Pantalla de resultados para la PREGUNTA DE COMPETENCIA P17 del Escenario 4

PC 21: ¿Cuáles son los correos intercambiados entre las cuentas C1 y C2 en un rango de fechas dado?

Para la pregunta de competencia 21, los datos que se requieren son cuatro: las cuentas *pedidos@empresa.com.ar* y *juan@empresas.com.ar*, y dos fechas para definir un rango (01 al 08 de febrero de 2017). Con estos datos, se informan los 15 correos intercambiados por ambas cuentas en las fechas dadas considerando el conjunto total de correos analizados. La Figura 4-70 muestra los resultados obtenidos.

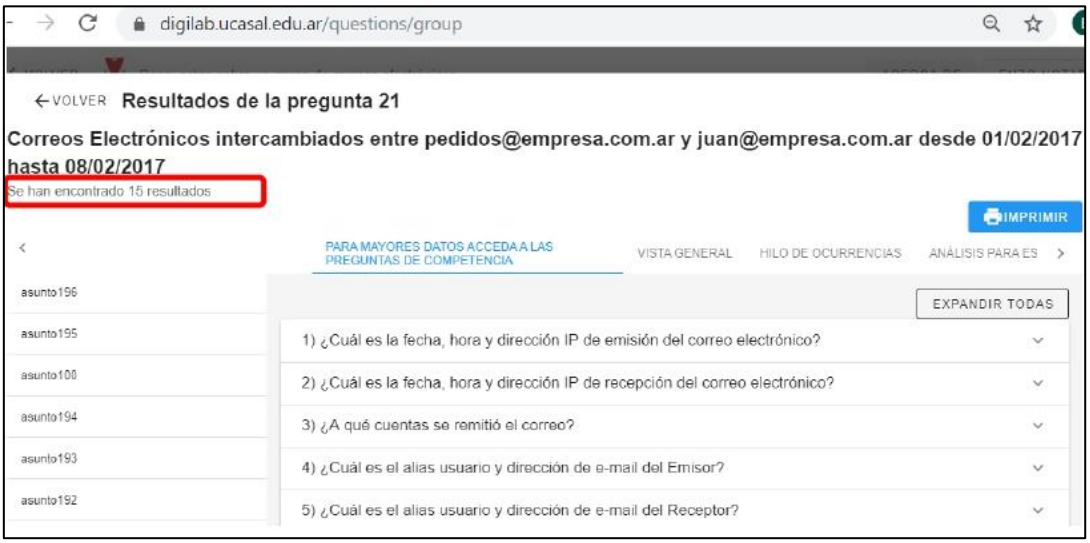


Figura 4-70: Pantalla de resultados para la PREGUNTA DE COMPETENCIA P21 del Escenario 4

Obsérvese que los 15 correos que se muestran en la pantalla de la Figura 4-70, intercambiados en el período señalado, incluyen los 3 correos identificados como *asunto100*, *asunto192* y *asunto191* que se enviaron desde *pedidos@empresa.com.ar*

y que se recibieron en la cuenta *juan@empresa.com.ar*. Es decir: la pregunta PC16 informa que se enviaron desde la cuenta *pedidos@empresa.com.ar*, la pregunta PC17 informa que se recibieron en la cuenta *juan@empresa.com.ar* y la pregunta PC21 efectivamente confirma que en esa fecha se intercambiaron esos correos. Los 12 correos restantes que señala la pantalla de la Figura 4-70 corresponden al intercambio realizado entre ambas cuentas durante el período señalado pero que no se intercambiaron exactamente el día 07 de febrero de 2017, sino en otras fechas dentro del rango señalado del 01 al 08 de febrero de 2017.

Por último, el Perito puede emitir el INFORME TÉCNICO DE PERICIA, en el que figuran los datos de los correos analizados, así como las respuestas a las distintas preguntas de competencia, obteniendo informes impresos similares a los indicados en los ANEXOS IV,V y VI.

4.8 Conclusiones del Capítulo

En este capítulo se explicó el uso de ObE Forensics como herramienta para el análisis forense de correos electrónicos, explicando en cuales instancias del procedimiento pericial se utiliza la herramienta.

Se describe la funcionalidad de la misma, y se ejemplifica mediante escenarios distintos para realizar pericias de correos electrónicos, en los que se muestran la carga de las cabeceras, el análisis de datos que permite ObE Forensics, los resultados de las preguntas de competencia y los Informes Técnicos de Pericia, que luego el Perito emita para entregar al Juez como resultado de su trabajo.

Las características de estos casos de estudio permitieron mostrar las ventajas de ObE Forensics en cuanto a herramienta de ayuda para las actividades del Perito, señalando la versatilidad de la herramienta cuando se debe considerar una cantidad importante de correos electrónicos, y también cuando en el mismo marco de la pericia se deben considerar varias cuentas de correo para las cuales se obtiene la evidencia digital correspondiente.

Es importante destacar que al realizar la extracción de las cabeceras en archivos de texto plano, no es impedimento para ObE Forensics considerar que el conjunto de cabeceras a cargar provenga de *distintas cuentas.* Es decir, si se generara un quinto

escenario, en el que se toman correos de cuentas independientes, incluso si las mismas corren en distintos clientes de correo, ObE Forensics analizará las cabeceras y brindará las respuestas requeridas según los filtros que se establezcan para las preguntas de competencia.

Como se indicó en el Capítulo 2 existen herramientas para el análisis forense de correos electrónicos, tales como *Aid4Mail*, *EmailTrackerPro, MailNavigator, OSForensics, E-mail Examiner, MailXaminer, IEFAF* y otras más robustas como *EnCase Forensic* que trabajan con diferentes formatos de evidencia digital, o las orientadas específicamente a teléfonos celulares como *TULP2G*, *MOBILedit FORENSIC* y *Elcomsoft Phone Breaker*. En relación a todas estas herramientas, ObE Forensics presenta diversas ventajas:

- Es una aplicación web, con todos los beneficios que éstas tienen respecto de libre disponibilidad e independencia de requerimientos de hardware y software específicos, cuando se accede directamente desde un navegador.
- ObE Forensics parte de la cabecera en formato de archivo de texto plano, lo que permite procesar correos provenientes de diversos gestores de correo, sin restricciones acerca del formato y/o estructura interna del correo.
- Identifica y señala la trazabilidad del proceso de transmisión, con toda claridad mediante el señalamiento del orden y secuencialidad de las ocurrencias del correo electrónico. Detalla particularmente la identificación de un equipo o servidor a partir de la dirección IP o el Hostname asociado.
- El proceso de generación de nuevas ocurrencias a partir de la identificación de direcciones IP en el valor del parámetro *from* vinculado al valor del parámetro *by* de la línea siguiente, asegura la detección de todas las direcciones IP/Hostnames que figuren en la cabecera.
- Además de brindar respuestas rápidas a los puntos de pericia más habituales como fecha de emisión/recepción y cuentas intervinientes, ObE Forensics está diseñada según criterios de navegabilidad que le da libertad al Perito para analizar la cabecera y los datos desde diferentes puntos de vista.
- Emite un Informe Técnico de Pericia que contiene toda la información de soporte del análisis forense, con posibilidades de que el Perito seleccione el contenido de dicho informe.

- La herramienta es lo suficientemente versátil como para responder de manera eficiente al procesamiento de datos en escenarios complejos, en los que intervienen varias cuentas de correo en conjunto.

CAPÍTULO 5. VALIDACIÓN DE OntoFoCE Y DE ObE Forensics

5.1 Introducción

Este capítulo presenta las técnicas y herramientas utilizadas para validar la ontología construida, buscando identificar y corregir errores en la conceptualización y diseño de OntoFoCE, así como en la aplicación web ObE Forensic que utiliza dicha ontología para el análisis forense de correos electrónicos. Además, se muestran los resultados de estas validaciones que permiten observar que la propuesta cumple con los objetivos propuestos.

El capítulo está organizado en tres secciones: en la sección 5.2 se describe el proceso realizado para validar OntoFoCE, en la sección 5.3 se explica la validación de ObE Forensics realizada por usuarios expertos, en la sección 5.4 se presentan consideraciones acerca de la completitud de los resultados que brinda la herramienta y en la sección 5.5 se señalan las conclusiones arribadas en este capítulo.

5.2 Validación de OntoFoCE

La bibliografía sobre el tema muestra diversos métodos para validar la calidad de una ontología, ya sea que se definan criterios generales de medición como criterios que midan determinados aspectos de la ontología. En los inicios del tema (Gomez-Perez, 1995) propuso criterios de evaluación basados en el uso de métricas, por su parte (J. Yu, 2008) plantea la evaluación de ontologías con criterios enfocados a la expresión, precisión, diseño y semántica del modelo. Otros autores se orientan a evaluar la ontología en función de criterios de calidad, como (Barchini & Álvarez, 2010) en el que definen el concepto de *calidad de una ontología* basado en cuatro dimensiones (descriptiva, estructural, funcional y operacional), y proponen la medición de la misma en base a un enfoque integral que toma –además de los componentes estructurales de clases y propiedades- la participación de quienes desarrollan y trabajan con la ontología (usuarios expertos y usuarios finales).

Sin embargo, no se ha encontrado un método formal que considere todos los criterios que pueden aplicarse para evaluar una ontología y encarar la validación de

OntoFoCE con una sola de las metodologías mencionadas, no permitiría una validación completa. Es por eso, que se consideró validar la ontología propuesta en esta tesis a partir de una metodología integrada propuesta por (Ramos et al., 2009) que utiliza los siguientes criterios: uso correcto del lenguaje (para evaluar la codificación de la ontología en base a las reglas y características del lenguaje utilizado); exactitud de la estructura taxonómica (se debe analizar la taxonomía revisando la consistencia, completitud y no redundancia de los conceptos y relaciones); validez del vocabulario (analizar el significado de los términos y conceptos a partir del conocimiento de expertos, compilaciones de textos o cualquier otra fuente de conocimiento disponible sobre el dominio) y adecuación a requerimientos (en esta fase se propone revisar si la ontología se ajusta a los requerimientos preestablecidos y si responde a las preguntas de competencia). El detalle de este proceso de validación se explica en la sección 5.2 de este capítulo.

Durante la primera etapa de construcción de OntoFoCE, en el que se realizó la implementación en Protégé, se realizó una validación de uso del lenguaje, en cuanto a la suficiencia de la misma para responder a las preguntas de competencia y, en consecuencia, a los puntos de pericia. Para ello se conformó un banco de pruebas, que consideró diferentes escenarios de forensia de correos electrónicos: análisis de un correo enviado a varios receptores y análisis de un conjunto de cuentas de correo con cierta cantidad de correos. También se utilizó OOPS! y OWL Validator[65] para validar OntoFoCE de manera semiautomática e identificar errores constructivos de la ontología.

Además de aplicar la metodología de validación integral señalada, se recurrió a la validación de la aplicación web desarrollada para OntoFoCE, invitando a usuarios expertos (peritos informáticos) a probar el desempeño de la aplicación con casos reales de pericias de correos electrónicos en los que estos expertos hayan participado. Esta actividad, que se describe en la sección 5.3 del presente capítulo, resultó sumamente provechosa para ajustar la aplicación web según las observaciones y sugerencias aportadas por los expertos. Como aspecto complementario, se trabajó también las cuestiones relativas a la seguridad informática de la aplicación, incorporando tanto las restricciones de estilo para las

[65] http://visualdataweb.de/validator/

aplicaciones web, como las cuestiones legales y técnicas relacionadas al procesamiento de datos reservados.

Se buscó una metodología que aborde una mirada integral para la validación de la ontología. En el caso de OntoFoCE, se recurrió a un modelo de validación integral, propuesto por (Ramos et al., 2009) que contiene cuatro criterios de validación de la ontología, propuestos a partir de un análisis comparativo de seis estudios sobre la temática (Brewster, Alani, Dasmahapatra, & Wilks, n.d.); (Burton-Jones, Storey, Sugumaran, & Ahluwalia, 2005); (Obrst, Ceusters, Mani, Ray, & Smith, 2007); (Porzel & Malaka, 2004); (Brank, Grobelnik, & Mladenic, 2005) y (Lozano Tello, 2002).

El análisis comparativo realizado por (Ramos et al., 2009) considera los métodos de evaluación de ontologías que cada uno de estos seis autores propone, considerando el siguiente conjunto de criterios: Taxonomía, Lenguaje, Aplicación, Vocabulario, Requerimientos de Arquitectura, Aceptación Social, Razonamiento Automático y Software. Este análisis comparativo se resume en la tabla de la Figura 5-1.

Criterios de evaluación de ontologías

Criterio / Autor	Taxonomía	Lenguaje	Aplicación	Vocabulario	Arquitectura Requerimientos	Aceptación Social	Razonamiento Automático	Software
(1)	√	-	√	√	-	-	-	-
(2)	-	√	√	√	-	-	√	-
(3)	√	-	√	√	-	-	-	-
(4)	-	√	√	√	-	√	-	-
(5)	√	√	√	√	√	-	-	-
(6)	√	√	√	√	√	-	-	√

(1) Brewster y cols, 2004 (2) Obrst y cols, 2007 (3) Porzel y Malaka, 2004
(4) Burton-Jones y cols, 2005 (5) Brank y cols, 2005 (6) Lozano-Tello, 2002 √ Considera - No considera

Figura 5-1: Cuadro Comparativo de Criterios de Evaluación de Ontologías [Fte (Ramos et al., 2009)]

Allí se observa que si se tiene en cuenta el grado de impacto de cada criterio en las metodologías analizadas se observa que todas incluyen la *validación* de la aplicación y del *vocabulario*, al que le siguen como criterios más utilizados la verificación de la *taxonomía* y del *lenguaje*. Los restantes criterios (Requerimientos de Arquitectura, Aceptación Social, Razonamiento Automático y Software) son los menos utilizados. Como resultado de este análisis comparativo de los criterios de

evaluación, (Ramos et al., 2009) proponen un método para evaluar ontologías basado en un conjunto selectivo de estos criterios.

La propuesta de (Ramos et al., 2009) incluye cuatro actividades: 1) validación del uso correcto del lenguaje, 2) exactitud de la estructura taxonómica, 3) validez del vocabulario y 4) adecuación a requerimientos.

En la primera actividad, ***validación del uso correcto del lenguaje***, se realizó la verificación de la ontología en cuanto a la capacidad de representatividad del modelo ontológico desarrollado atendiendo a características del lenguaje utilizado. A partir de la implementación de OntoFoCE en Protégé, y utilizando las herramientas auxiliares de este entorno, se realizaron las primeras verificaciones de consistencia del modelo OWL construido, utilizando para ello el banco de pruebas que se instanció manualmente para conformar los ejemplos de los escenarios 1, 2 y 3 desarrollados en el capítulo 3. Utilizando uno de los razonadores provisto por la herramienta, denominado Hermit 1.3.8.413[66], fue posible verificar la consistencia de las definiciones del modelo OWL construido.

En esta actividad también se validó su construcción utilizando herramientas semiautomáticas que analizaron el código de OntoFoCE identificando errores y malas prácticas que se fueron ajustando debidamente.

Para la identificación de errores, se validó la sintaxis con OWL Validator, esta herramienta permite verificar ontologías escritas en RDF/XML, OWL/XML, Sintaxis funcional OWL, Sintaxis OWL de Manchester, Sintaxis OBO o Sintaxis KRSS, e incluso puede comparar con los perfiles OWL 2.

Para descartar el uso de malas prácticas se recurrió a OOPS![67] (OntOlogy Pitfall Scanner), esta herramienta realiza un análisis y diagnóstico del código OWL, en base a un catálogo de 41 fallas habituales en la construcción de las ontologías e informa al desarrollador un conjunto de posibles causas de error, provenientes muchas veces de las malas prácticas en el modelado de la ontología. OOPS! clasifica los errores en *dimensiones* y dentro de cada una, los identifica por *criterios*, ayudando al desarrollador a entender por qué se producen las fallas.

Para la actividad 2, ***exactitud de la estructura taxonómica***, se recurrió a la colaboración de usuarios expertos (peritos informáticos) que a través de actividades de Focus Group y Talleres, analizaron y discutieron la representatividad de

[66] http://www.hermit-reasoner.com/
[67] http://oops.linkeddata.es/response.jsp

OntoFoCE, teniendo presente su conocimiento del dominio sobre Forensia Digital. Las propuestas de ajustes que se lograron a partir de la participación de estos usuarios expertos permitieron mejorar la representatividad de OntoFoCE.

En la tarea 3 abocada a la ***validez del vocabulario***, se realiza una verificación de los términos incluidos en la ontología versus un corpus sobre correos electrónicos a partir de una fuente de conocimiento independiente. Se tomó como tal el trabajo elaborado por (Banday, 2011) en el que se describe la arquitectura de un correo electrónico, definiendo los distintos componente y los procesos técnicos necesarios para describir cómo se realiza el proceso de transmisión. Considerando entonces los términos de OntoFoCE y los del corpus de referencia, se analizaron dos métricas: *Precisión* y *Recall*, la primera de ellas mide el grado de coincidencia entre los términos de la ontología y el corpus, y la segunda indica la relación inversa, es decir, cuantos términos del corpus se representan en la ontología. Analizadas en conjunto ambas métricas, permiten identificar el grado de validez de los términos que conforman el vocabulario de OntoFoCE.

La última actividad, referida a la ***adecuación a requerimientos,*** consiste en verificar si los requerimientos exigidos a la ontología se cumplen, considerando los objetivos de la misma, así como las preguntas de competencia.

A continuación se describe el desarrollo de estas actividades para validar OntoFoCE de acuerdo a los criterios formulados por (Ramos et al., 2009).

5.2.1 Validación del Uso Correcto del Lenguaje

Esencialmente esta fase consiste en verificar que el lenguaje utilizado para construir la ontología sea el adecuado y que la escritura de la ontología no contenga errores o defectos que pudieran impactar en el funcionamiento de la misma o en su correcta implementación.

En esta fase se analizan los lenguajes utilizados: OWL, RDF, SWRL, etc. para lo cual es conveniente implementar la ontología en un framework de desarrollo que posibiliten validar estos lenguajes mientras se está construyendo la ontología.

En el caso de OntoFoCE, la implementación se realizó en Protégé, debido a las bondades de este entorno para validar continuamente la ontología que se está

construyendo a través de las funciones que propone esta herramienta, permitiendo corregir inconsistencias sintácticas y construir un código libre de errores.

Asimismo, como primera instancia de validación, se ejecutó el razonador con un conjunto de valores instanciados a partir de los ejemplos trabajados en los escenarios del capítulo 3, y se observó que el modelo lógico de OntoFoCE realiza correctamente las inferencias correspondientes a ese conjunto de datos de ejemplo.

Durante la construcción de la ontología, una vez que el modelo se probó con instancias de ejemplo en Protégé, se consideró oportuno recurrir a OOPS! para revisar los aspectos constructivos de OntoFoCE.

Dentro del conjunto de diversas herramientas semiautomáticas para validar ontologías, (Poveda Villalón, 2016) proponen "OOPS!" (Ontology Pitfall Scanner), una aplicación web que permite identificar errores en la construcción de la ontología OWL, que habitualmente cometen los desarrolladores que no están familiarizados con lenguajes de implementación de ontologías.

Para llevar adelante esta identificación, OOPS! posee un catálogo compuesto por 41 malas prácticas habituales en la construcción de la ontología.

Mediante un proceso de análisis del código OWL, la herramienta on line OOPS! realiza un diagnóstico en base a un catálogo de 41 fallas o pitfalls, sobre las cuales entrega al diseñador de la ontología un conjunto de posibles causas provenientes muchas veces de las malas prácticas en el proceso de modelado.

Por otra parte, OOPS! define un grado de impacto de cada uno de estos errores o pitfalls. Son considerados como *errores críticos,* aquellos que afectan la consistencia y razonamiento del modelo ontológico, *errores importantes* los que no afectan la función de la ontología pero sería deseable su corrección, y *errores menores*, los que no presentan un problema crucial para la funcionalidad de la ontología pero afectan la comprensión de la conceptualización que representa.

En esta herramienta se pueden seleccionar individualmente los errores a detectar en el análisis de la ontología según los pitfall de evaluación o se puede seleccionar por categoría.

En la clasificación de errores por categoría que se indican en la Figura 5-2 se distinguen dos grandes grupos: la clasificación por dimensión y la clasificación por criterio de evaluación. En el primer caso, se puede analizar una ontología según 3 enfoques: estructural, funcional y sobre perfiles de usabilidad. En el segundo caso,

se puede considerar la consistencia, la completitud y la cualidad de concisa de la ontología en análisis.

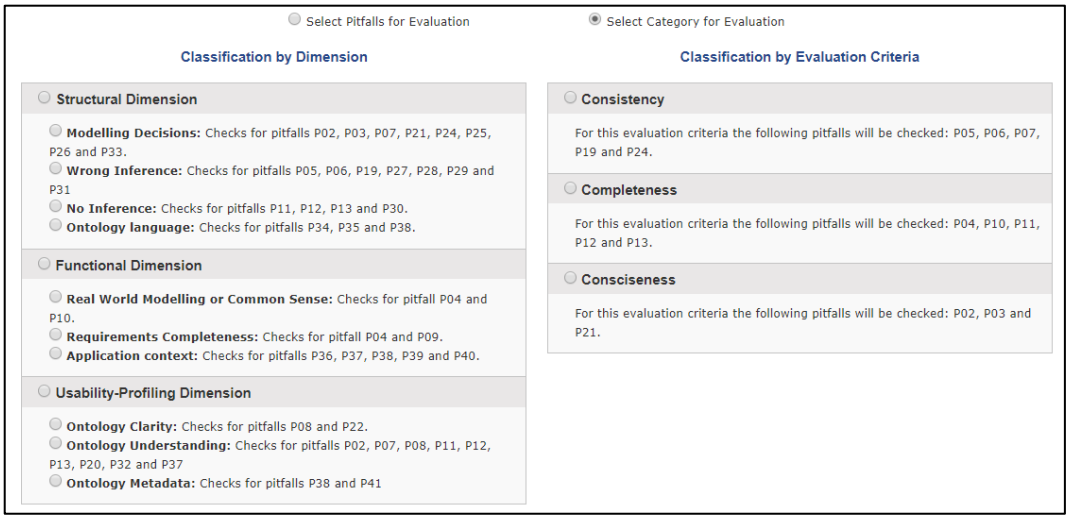


Figura 5-2: Catálogo de errores de OOPS! clasificado por Categoría

La clasificación por dimensiones responde a la categorización de métricas introducida por (Gangemi, Catenacci, Ciaramita, Lehmann, & Gil, 2005). En particular, la dimensión estructural aborda la evaluación de la topología, las propiedades lógicas y la semántica formal de la ontología en análisis, mientras que la dimensión funcional se ocupa de revisar la semántica de la ontología, en cuanto al uso y representación del conocimiento pretendido para la misma. Por último, la dimensión sobre el perfil de la usabilidad apunta a identificar errores referentes a las anotaciones y metadatos que debe contener la ontología para facilitar su implementación y comprensión.

Al considerar la clasificación por criterios de evaluación, que también se indican en la Figura 5-2, se pueden considerar el mismo conjunto de errores, pero asociados de acuerdo a los criterios de consistencia, completitud y concisión. Así por ejemplo, si se quiere validar expresamente la consistencia de la ontología, deberá atenderse a los errores identificados como P05 (Definición de relaciones inversas incorrectas), P06 (inclusión de jerarquías de clases cíclicas), P07 (mezclar conceptos distinto en una misma clase), P19 (definición de de múltiples dominios o rangos) y P24 (uso de definiciones recursivas).

Las características descriptas de OOPS! muestran la potencialidad de esta herramienta para ayudar en la validación de la ontología, atendiendo a la detección de errores constructivos comunes durante el desarrollo del modelo ontológico. En

particular, la primera vez que se utilizó OOPS! para validar OntoFoCE, permitió detectar las siguientes malas prácticas:

- Error P08: Anotaciones faltantes
- Error P24: Uso de definiciones recursivas
- Error P34: Clases no declaradas
- Error P41: Licencia no declarada

A continuación se describe cada uno de estos errores, y la solución encontrada para superarlos.

La Figura 5-3 muestra el **Error P08** de OOPS!, considerado como un error menor. Este tipo de error, que corresponde a la capacidad de usabilidad de la ontología, hace referencia a la ausencia de notas descriptivas de los distintos elementos de la misma, los cuales se completaron adecuadamente informando una breve descripción en el campo *rdfs:comment* de las Object Property y de los DataProperty de la implementación en Protégé.

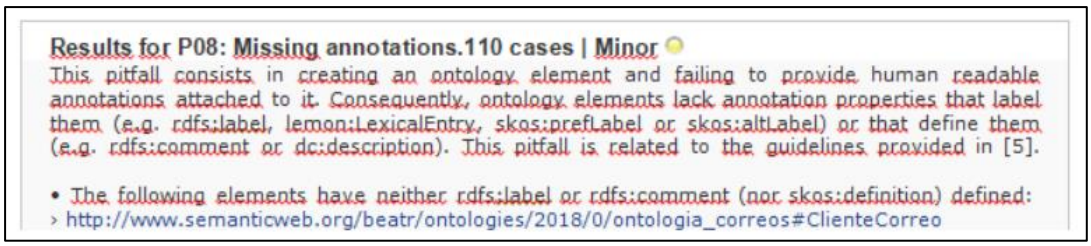


Figura 5-3: Error P08 sobre Anotaciones Faltantes

En la Figura 5-4 se muestra el mensaje de **Error P24** de OOPS! considerado como error importante. El error P24 señala errores en las decisiones de modelado (dimensión estructural) y problemas de consistencia del código. En este caso particular, la herramienta indica que la clase *OcurrenciaDeTransmision* está definida recursivamente.

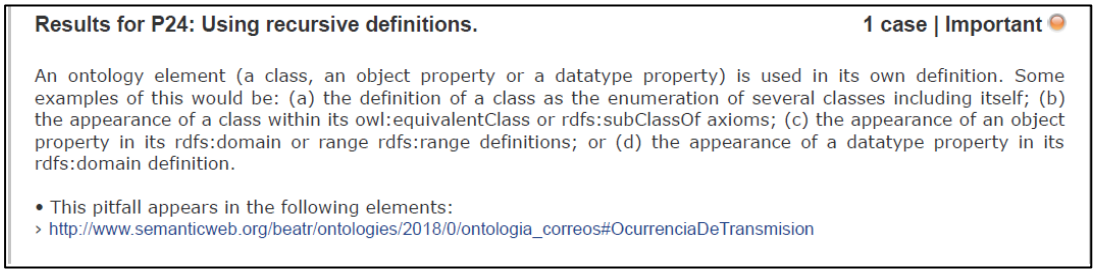


Figura 5-4: Error P24 sobre uso de definiciones recursivas

La clase *OcurrenciaDeTransmision* tiene características de recursividad dado que representa las ocurrencias intermedias dentro de un hilo, los axiomas de la sección 3.4.2.3 del Capítulo 3 definen la clase *OcurrenciaDeTransmision* en base a

las relaciones *esAnteriorA* y *esSiguienteDe* señalando que "*Una Ocurrencia de Transmisión es aquella que es anterior a otra Ocurrencia de Transmisión o a una Ocurrencia de Recepción o es siguiente de una Ocurrencia de Emisión"*. Esto se expresa con el siguiente axioma de clase:

Class: OcurrenciaDeTransmision
EquivalentTo: Ocurrencia and
((esAnteriorA exactly 1 OcurrenciaDeTransmision) or (esAnteriorA exactly 1 OcurrenciaDeRecepcion)) and (esAnteriorA only (OcurrenciaDeRecepcion or OcurrenciaDeTransmision)) and
((esSiguienteDe exactly 1 OcurrenciaDeTransmision) or (esSiguienteDe exactly 1 OcurrenciaDeEmision))and (esSiguienteDe only (OcurrenciaDeEmision or OcurrenciaDeTransmision))

Este axioma señala que una instancia de ***OcurrenciaDeTransmision*** es una instancia de ***Ocurrencia,*** que cumple con los siguientes criterios:

- Se asocia mediante la relación *esAnteriorA* con una instancia de ***OcurrenciaDeTransmision*** o con una única instancia de ***OcurrenciaDeRecepcion***.
- Se vinculará mediante la relación *esAnteriorA*, solamente con instancias de las clases ***OcurrenciaDeRecepcion*** u ***OcurrenciaDeTransmision***
- Se relaciona mediante la relación *esSiguienteDe* con una instancia de la clase ***OcurrenciaDeTransmision*** o con una instancia de ***OcurrenciaDeEmision***
- Se asocia mediante la relación *esSiguienteDe* con una instancia de ***OcurrenciaDeTransmision*** u ***OcurrenciaDeRecepcion***.

Si bien no se aconsejan las definiciones recursivas y se entienden como mala práctica, hay ocasiones, como en este caso, en que no se pueden evitar.

En la Figura 5-5 se muestra el mensaje del **Error P34** emitido por la herramienta OOPS!:

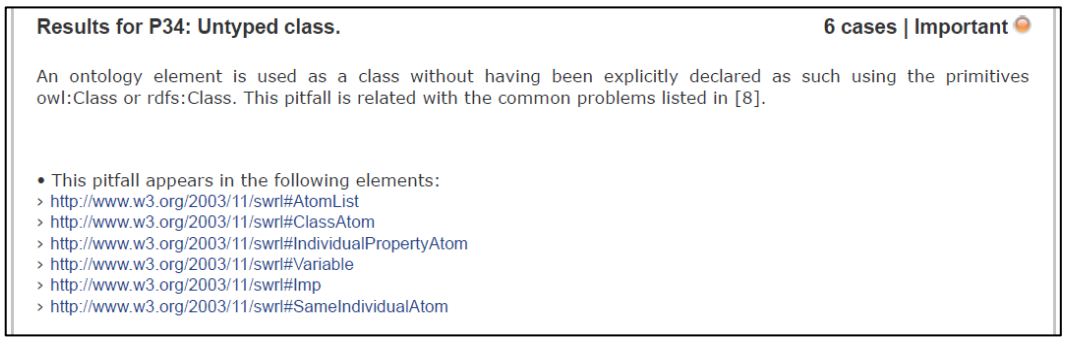


Figura 5-5: Error P34 sobre Clases no declaradas

El **Error P34** hace referencia a la definición de clases para las que no se define un tipo. En particular, las clases para las que se detectó este problema corresponden a clases definidas en la especificación de SWRL y que son utilizadas en las reglas

propuestas para OntoFoCE. No es posible modificar las definiciones de estas clases en nuestra ontología, por lo cual este error queda sin resolver en OOPS!.

Por último, el **Error P41**, hace referencia a los metadatos de la ontología. Particularmente señala la falta de declaración de una licencia. Para subsanar este error se completó el conjunto de metadatos de OntoFoCE, registrando la licencia CreativeCommons V4.0 https://creativecommons.org/licenses/by/4.0/legalcode en la propiedad *dcterms;license* de la ontología implementada en Protégé.

OOPS! permitió validar OntoFoCE resolviendo los errores señalados, con excepción de la definición de clases marcada por el Error P34 y la definición recursiva de la Clase *OcurrenciaDeTransmision* señalada en el Error P24 de esa herramienta.

Asimismo, se decidió validar el código con la herramienta OWL Validator que permite seleccionar OWL 2 como perfil de la evaluación. Esta herramienta, diseñada por la Universidad de Manchester y definida como aplicación web de uso libre, realiza el análisis sintáctico de una ontología y verifica si la misma se ajusta a un perfil determinado (OWL 2, OWL 2 DL, OWL 2 EL, OWL 2 QL u OWL 2 RL), por otra parte, es posible seleccionar el tipo de reporte de sintaxis a generar: OWL Manchester, DL o Funcional. La Figura 5-6 muestra la pantalla de carga de la aplicación, con un recuadro para ingresar el código OWL a validar y los parámetros que pueden ajustarse según la sintaxis que se desea considerar.

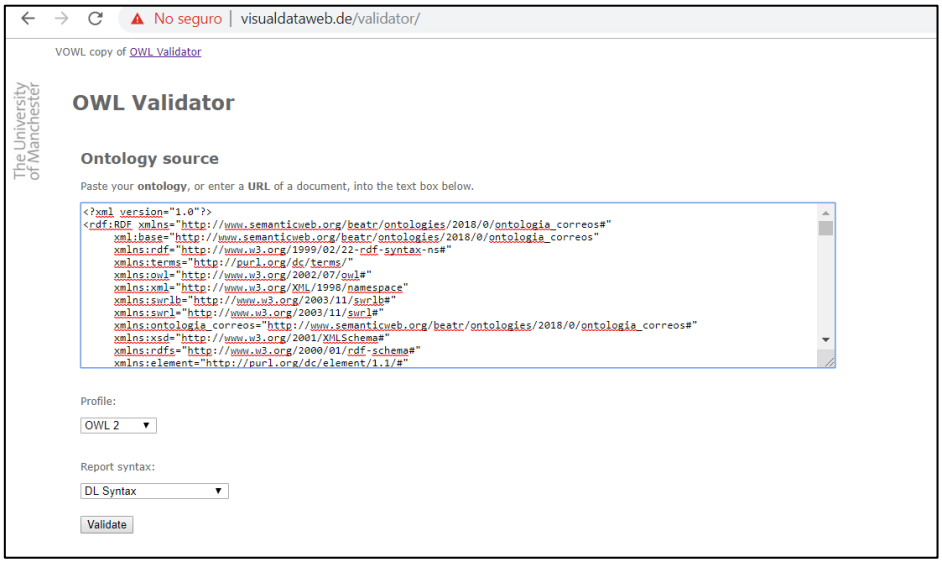


Figura 5-6: Pantalla de ingreso del código de OntoFoCE en OWL Validator

Allí se observa que se cargó el código OWL de OntoFoCE, se seleccionó el valor *OWL2* para el parámetro *Profile* y la opción de la sintaxis en *Description Logic (DL)* para el parámetro *Report Sintax.*

Luego de seleccionar el botón de *"Validate"* la aplicación procesa el código ingresado y devuelve una pantalla de respuesta que se indica en la Figura 5-7. Allí se observa el reporte de la herramienta que no muestra errores en la sintaxis del código ingresado. Eso se comprueba con la frase "La ontología y los componentes importados poseen el perfil de OWL 2" indicado en el apartado *Summary* de la pantalla.

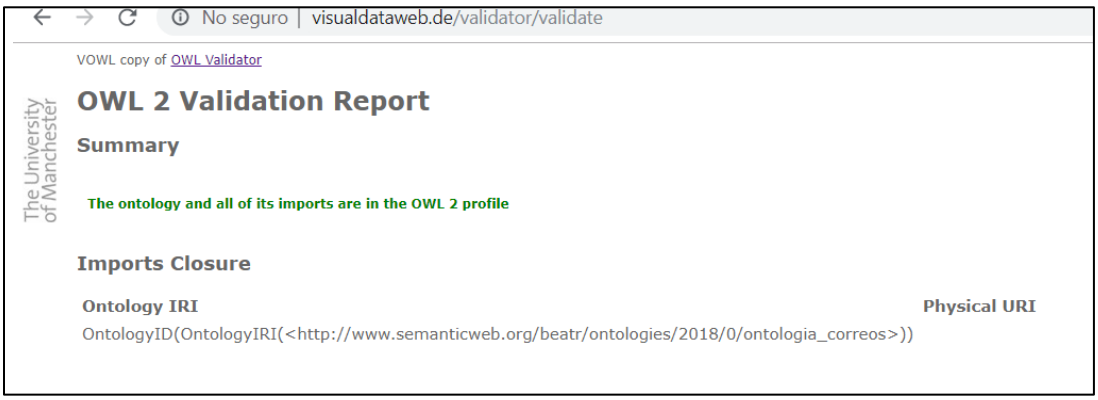


Figura 5-7: Resultado de la evaluación de OntoFoCE con OWL Validator

Con la incorporación de ambas herramientas semiautomáticas para validación del código, OOPS! y OWL Validator, se pudo realizar una revisión exhaustiva del lenguaje utilizado en OntoFoCE, obteniéndose un grado de validación adecuado y suficiente para avanzar en el siguiente criterio de validación propuesto por (Ramos et al., 2009), aun considerando los pitfalls marcados por OOPS! y justificados desde el diseño de la ontología.

5.2.2 Validación de la Exactitud de la estructura taxonómica

La revisión de la estructura taxonómica se basa en el análisis de la representatividad que tienen los conceptos de la ontología, en cuanto a las definiciones jerárquicas y semánticas, las cuales deben ajustarse a la realidad. Esta revisión solo la puede aportar el usuario experto en base al conocimiento que tenga sobre el dominio de la ontología, pero será necesario que dicho usuario experto tenga los conocimientos básicos sobre componentes ontológicos, a fin de comprender si los conceptos están bien representados en el modelo ontológico de OntoFoCE.

En cuanto a la definición taxonómica de la ontología, existen un conjunto de errores comunes que deben evitarse: clasificación semántica incorrecta (al proponer un concepto como subclase de una clase a la que no pertenece), clases e instancias

con distintos nombres pero definiciones similares, ausencia de clases, redundancia de relaciones, entre otros.

En el caso de OntoFoCE, se invitó a los usuarios expertos a que revisaran la estructura jerárquica utilizada para representar el conocimiento, revisar las propiedades de los conceptos y las relaciones establecidas entre ellos, así como para analizar la generalización y especialización de clases del modelo. Esta revisión se realizó en un taller de trabajo al que fueron invitados peritos informáticos integrantes del Grupo de Investigación de Forensia Digital de la UCASAL. En dicho taller, realizado en sucesivos encuentros que se desarrollaron durante la etapa de construcción de la ontología, se explicó a los asistentes acerca del modelo ontológico de OntoFoCE, abordando además la definición y significado de cada componente ontológico. Estos expertos, con conocimiento en la disciplina informática y en Forensia Digital, realizaron sugerencias sobre la representación de los siguientes conceptos y relaciones:

- En las primeras versiones del modelo lógico, se utilizó el concepto de *Usuario* para representar los datos de las cuentas de correo utilizadas, los expertos sugirieron ajustar este concepto denominándolo *Cuenta,* ya que los datos obtenidos no son suficientes para identificar al usuario, es decir la persona que utiliza una cuenta de correo, sino que son definitorios de la cuenta de correo en sí misma.
- De su experiencia como peritos, los expertos señalaron que no siempre es posible realizar el análisis forense de correos electrónicos, ya que la evidencia digital obtenida contiene datos incompletos u ocultos generados mediante scripts malintencionados que adulteran el contenido de la cabecera, específicamente ocultando o borrando direcciones IP. Esto derivo en la definición de la subclase *CorreoFactible* que especializada la clase *Correo*¸ para contemplar los requerimientos mínimos que debe cumplir una cabecera a fin de que el análisis forense realizado sea posible.
- También se discutió sobre la relación entre los elementos del correo (Asunto, Cuerpo, Adjunto y Cabecera), en cuanto a que si debían vincularse al concepto *Correo* o al concepto *Ocurrencia.* Si bien parece natural que estos elementos se asocien al *Correo* ya que son parte del mismo, al recurrir a la definición del concepto *Ocurrencia* para representar el proceso de transmisión, se observó que en realidad, estos elementos (Asunto, Cuerpo, Adjunto y Cabecera) *viajan* con el

correo y se almacenan en los equipos y servidores de paso correspondientes, así es más adecuado asociarlos al concepto de *Ocurrencia* que al de *Correo.*

- Por último, los expertos analizaron la representación de la dirección IP de un equipo o servidor, encontrando que los procesos de transmisión que controlan el envío almacenan en la cabecera la dirección IP explícita (en formato IPv4 o IPv6 según corresponda) o el dominio con que se identifica el servidor. Cuando ocurre esto último, no es posible averiguar con plena certeza cual fue la dirección IP de ese dominio al momento del proceso de transmisión, por ello, la clase *IdentificacionEquipo* se especializó en dos subclases: *IP* y *Hostname.*

Con los aportes de los usuarios expertos, se ajustó el modelo taxonómico inicialmente propuesto logrando una versión más acabada de la ontología en cuanto a la representatividad del análisis forense de correos electrónicos.

5.2.3 Validación del Vocabulario

Esta fase consiste en verificar que los términos codificados en la ontología existan y sean significativos en otras fuentes de conocimiento independiente, como por ejemplo el corpus del dominio.

Una vez identificado el corpus del dominio se puede trabajar con dos métricas: ***precisión*** y ***recall.***

La métrica de ***Precisión*** se define como el porcentaje de términos de la ontología que figuran en el corpus, en relación a la cantidad total de términos de la ontología, y se calcula mediante la siguiente fórmula:

$$Precisión = CO\_C / COnto$$

Siendo:

CO_C = Cantidad de términos que se solapan entre la ontología y el corpus.

COnto = Cantidad total de términos de la ontología (sumatoria de clases, subclases y atributos)

También se pueden considerar los valores del corpus y de la ontología definiendo una métrica referida al corpus, es decir, se define la métrica de ***Recall*** como el porcentaje de términos del corpus que aparecen en la ontología, con relación al total de términos del corpus. La fórmula de cálculo de esta métrica es la siguiente:

$$Recall = CO\_C / CCorpus$$

Siendo:

CO_C = Cantidad de términos que se solapan entre la ontología y el corpus.

CCorpus = Cantidad total de términos del Corpus

En el caso de la correos electrónicos, puede tomarse como referencia del corpus el trabajo elaborado por (Banday, 2011) en el que se describe la arquitectura de un correo electrónico, definiendo con claridad cada componente o parte identificable del correo, así como los procesos técnicos necesarios para describir cómo se realiza la transmisión del envío.

Cabe aclarar que OntoFoCE incluye conceptos que van más allá del análisis forense, y se incluyen en la ontología solo a los fines de derivar la trazabilidad del correo a partir de la representación del proceso de transmisión.

A partir del trabajo citado se elaboró la Tabla 5-1 que describe los diferentes elementos o conceptos sobre el correo electrónico, que pueden integrarse para conformar el *corpus* sobre el dominio correo electrónico. En dicha tabla se indica además cuales de esos elementos o conceptos están representados en OntoFoCE.

La tabla mencionada contiene una primera columna de identificación del término. En las segunda y tercera columna se muestra el nombre del término según la propuesta de (Banday, 2011) y la definición del tipo de elemento, respectivamente. Luego la cuarta columna presenta la descripción del término, la cual tiene por fin aclarar detalles que el propio nombre del término no expresa. Finalmente, en la última columna se indica el concepto o atributo de la OntoFoCE asociado a dicho término. Cada fila de la tabla representa un término según la propuesta de arquitectura del correo electrónico de (Banday, 2011).

Tabla 5-1: Corpus de términos del dominio Correo Electrónico

N°	Término	Tipo	Breve descripción	Concepto o atributo de OntoFoCE asociado al elemento
1	***Autor***	***Componente del correo***	***Responsable de crear el mensaje, su contenido y su lista de destinatarios***	***aliasUsuario***
2	***Receptor***	***Componente del correo***	***Consumidor del mensaje entregado***	***CuentaReceptor***
3	Controlador de devolución	Proceso de transmisión	Proceso que notifica sobre fallos en los envíos.	
4	Mediador	Proceso de transmisión	Proceso que recibe, agrega, reformula y redistribuye mensajes entre los autores y los destinatarios.	
5	Originador	Proceso de transmisión	Proceso que asegura la validez del mensaje y luego lo envía. Maneja cualquier problema sobre la transmisión del correo.	
6	Relay	Proceso de transmisión	Proceso responsable de realizar el enrutamiento del envío, y se encarga de agregar información de rastreo en la cabecera del correo.	

N°	Término	Tipo	Breve descripción	Concepto o atributo de OntoFoCE asociado al elemento
7	Gateway	Proceso de transmisión	Proceso que conecta servicios de correo heterogéneos a pesar de las diferencias en su sintaxis y semántica.	
8	Receptor	Proceso de transmisión	Proceso que realiza la entrega final o envía el mensaje a una dirección alternativa en caso de fallo en el envío.	
9	Edge	Proceso de transmisión	Proceso responsable de la transferencia en redes al borde de la Internet abierta.	
10	***Consumidor***	***Proceso de transmisión***	***Servicio perimetral, como es común para el acceso a correo electrónico basado en web***	***ClienteCorreo***
11	***Transit***	***Proceso de transmisión***	***Proveedores de servicios de correo electrónico (ESP)***	***ispIP***
12	SMTP	Protocolo y script	Protocolo de comunicación para la transferencia del correo entre el equipo emisor y el servidor de la cuenta emisora.	
13	SMTP*	Protocolo y script	Conjunto de protocolos SMTP ajustados a normas RFC	
14	HTTP	Protocolo y script	Protocolo de transferencia de hipertextos, que actúa sobre SSL y TLS	
15	INT	Protocolo y script	Protocolos y procedimientos específicos para la entrega interna de correo electrónico entre nodos del mismo dominio	
16	IMAP	Protocolo y script	Protocolo de comunicación para la transferencia del correo desde el servidor de la cuenta receptora.	
17	POP3	Protocolo y script	Protocolo para bajar el correo del servidor de la cuenta receptora	
18	***Message- ID***	***Componente del correo***	***Cadena única de identificación de mensaje generada cuando es enviado.***	***idCorreo***
19	***In-Reply-To***	***Componente del correo***	***Contiene el ID del mensaje original que se envía el mensaje de respuesta.***	***idCorreo***
20	References	Componente del correo	Identifica otros documentos relacionados con este mensaje, como otro mensaje de correo electrónico.	
21	***From***	***Componente del correo***	***Nombre y dirección de correo electrónico del autor del mensaje.***	***CuentaEmisor aliasUsuario***
22	***Sender***	***Componente del correo***	***Dirección responsable de enviar el mensaje en nombre del autor, si no se omite o es igual a lo especificado en el campo From***	***CuentaReceptor***
23	Reply-to	Componente del correo	Dirección de correo electrónico, que el autor desea que los destinatarios utilicen para las respuestas.	
24	***Date***	***Proceso de transmisión***	***Fecha y hora en que el mensaje estuvo disponible para entrega***	***fechaHoraOcurrencia***
25	***Subject***	***Componente del correo***	***Describe el tema o asunto del mensaje.***	***contenidoAsunto***
26	Comments	Componente del correo	Contiene comentarios resumidos sobre el mensaje.	
27	Keyword	Componente del correo	Contiene una lista de palabras clave separadas por comas que pueden ser útiles para los destinatarios, por ejemplo al buscar el correo.	
28	***TO***	***Componente del correo***	***Especifica una lista de direcciones de los destinatarios del mensaje.***	***CuentaReceptor***
29	***CC***	***Componente del correo***	***Lista de direcciones de envío adicionales a la principal***	***CuentaReceptor***
30	***BCC***	***Componente del correo***	***Lista oculta de direcciones de envío***	***CuentaReceptor***
31	***Resent Message- ID***	***Componente del correo***	***Cadena única de identificación de mensaje generada cuando es reenviado.***	***idCorreo***
32	***Resent-****	***Componente del correo***	***Al reenviar manualmente un mensaje, reenvíe los campos de encabezado refiriéndose al reenvío, no al mensaje original.***	***CuentaReceptor***

N°	Término	Tipo	Breve descripción	Concepto o atributo de OntoFoCE asociado al elemento
33	List-ID	Componente del correo	Identificación de una lista de distribución de correos	
34	***List-****	***Componente del correo***	***Conjunto de correos que integran una lista de distribución***	***CuentaReceptor***
35	***Received***	***Componente del correo***	***Contiene información de rastreo que incluye el host de origen, procesos mediadores, relay y nombres de dominio y / o direcciones IP***	***identificadorEquipo***
36	***Return-Path***	***Componente del correo***	***Contiene la dirección registrada en el From del correo enviado***	***cuentaEmisor***
37	DKIM Signature	Protocolo y script	La firma del correo electrónico se almacena en el campo DKIM-Signature del encabezado.	
38	Received-SPF	Protocolo y script	Contiene los resultados de validación de las políticas del remitente (SPF) para un dominio y sus servidores de correo, sobre qué máquinas están autorizadas a utilizar su dominio en las campos HELO y MAIL FROM..	
39	MIME Version	Componente del correo	Describe la versión del formato de mensaje MIME.	
40	***Content-****	***Componente del correo***	***Contiene una colección de campos MIME que describen varios aspectos del cuerpo del mensaje, incluyendo firmas.***	***contenidoCuerpo firmaUsuario***
41	HELO/ EHLO	Protocolo y script	Contiene datos sobre el dominio de alojamiento definidos en los comandos HELO y EHLO de SMTP.	
42	ENVID	Protocolo y script	Cadena incluida en el Servidor DSN para identificar la dirección de retorno del destinatario.	
43	***MailFrom***	***Componente del correo***	***Cadena que contiene la dirección de correo electrónico para control de devolución de mensajes.***	***CuentaEmisor***
44	***RcptTo***	***Componente del correo***	***Especifica la dirección de buzón de un destinatario.***	***CuentaReceptor***
45	ORCPT	Protocolo y script	Parámetro opcional para el comando RCPT, que indica la dirección original a la que se dirige el RCPT TO actual.	
46	***Source Address***	***Proceso de transmisión***	***Contiene la dirección de origen del host anterior al servidor receptor actual desde el cual se envió el datagrama IP (el mensaje de correo electrónico está fragmentado en paquetes IP).***	***identificadorEquipo***

Considerando la propuesta de (Banday, 2011) descripta en la Tabla 5-1 resulta necesario identificar aquellos componentes del corpus que se incluyen en OntoFoCE. Así, en función de la función que cumple cada uno se han definido diferentes tipos de elementos:

- Componentes del correo: aquellos referidos al propio correo electrónico (24 elementos),
- Proceso de Transmisión: aquellos que definen los procesos involucrados en el proceso de transmisión (11 elementos), y
- Protocolos y Script: los distintos protocolos y cadenas de comandos que rigen el proceso de envío (11 elementos).

Cabe mencionar que de estos grupos mencionados, solo son de interés para el análisis forense de correos electrónicos 18 elementos incluidos en el primer apartado (señalados en negrita y cursiva en la tabla), los restantes no se consideran porque no contienen datos que aporten a la validez o existencia del correo electrónico, estos componentes que se descartan son los siguientes:

- *References:* este término usualmente contiene información para identificar otros documentos relacionados con este mensaje, como otro mensaje de correo electrónico, pero cuando el análisis forense requiere de estudios de correlación entre varios correos electrónicos, ésta se trata siempre desde los datos de emisión/recepción del correo, no desde el contenido del mismo.
- *Reply To:* este término informa la cuenta de correo electrónico, que el autor desea que los destinatarios utilicen para las respuestas. Para el análisis forense del correo, esta información es complementaria, ya que el proceso de envío se analiza considerando la cuenta emisor y la cuenta receptor.
- *Comments:* aquí figuran comentarios resumidos sobre el mensaje, que –al igual que el contenido del cuerpo del mensaje- no forma parte del análisis forense para identificar la validez o existencia del correo electrónico.
- *Keyword:* el término contiene una lista de palabras clave separadas por comas que pueden ser útiles para los destinatarios, por ejemplo al buscar el correo. No forma parte del análisis forense, por idénticas razones a las señaladas para el término anterior.
- *List-ID*: este término contiene una identificación de la lista de distribución, no el conjunto de cuentas que integran esa lista de distribución. Por esa razón, no se considera en el análisis forense.
- *MIME Version*: en este elemento se hace alusión a la versión del formato de mensaje MIME, no es de interés para el análisis forense.

De los 11 elementos identificados como del Proceso de Transmisión, solo se utilizan 4 en el análisis forense (señalados en negrita y cursiva en la tabla), los restantes (*Controlador de devolución, Mediador, Originador, Relay, Gateway, Receptor y Edge)* hacen referencia a procesos que llevan adelante durante la transmisión del correo electrónico, pero que no son considerados para el análisis forense debido a que no impacta en el almacenamiento del correo durante el proceso de transmisión, es decir, no impactan en las ocurrencias. Estos datos, al igual que los

referidos a encriptación y protección antivirus que son generados por los clientes de correo, se ignoran durante el análisis forense.

De acuerdo a la metodología propuesta por (Ramos et al., 2009), la definición de un corpus del dominio y su relación con los términos de la ontología a validar, expresada en términos comparativos como en la Tabla 5.1, se toman como insumo para considerar las métricas de ***Precisión y Recall.*** Es decir que, una vez considerado el corpus del dominio con un conjunto finito de términos, y luego de identificar entre ellos aquellos que se representan en la ontología, se está en condiciones de trabajar las métricas propuestas por (Ramos et al., 2009).

En el caso de OntoFoCE estos valores son los siguientes:

CO_C = 18 términos de Componentes del correo +
4 términos del Proceso de Transmisión
= 22 (señalados en negrita y cursiva en la tabla)

COnto = 14 clases +
13 subclases +
23 atributos de clase = 50

Así la precisión se expresa como:

Precisión = 22/50 = 44%

Respecto de la métrica de *Recall*, en el caso de OntoFoCE estos valores son los siguientes:

CO-C = 22

CCorpus = 46

Así la métrica de Recall se expresa como:

Recall= 22/46 = 48%

La interpretación de estos valores indica la validez del vocabulario.

El valor del 44% obtenido para la *Precisión* señala que en la ontología se ha representado casi la mitad del corpus definido para el correo electrónico, y esto es lógico, ya que los restantes términos de OntoFoCE se definieron solo a los fines de derivar la trazabilidad del correo a partir de la representación del proceso de transmisión. Es decir, la ontología toma los elementos principales del correo electrónico –aquellos que permiten obtener una precisión del 44% respecto del corpus- y se agregan los restantes términos requeridos para validar la existencia del correo emitido a través de la trazabilidad del proceso de transmisión. Hubiera sido deseable que el valor de la métrica de *Precisión* fuera más alta, pero solo sería

posible si se tomara en consideración un corpus más completo, que incluyera los términos relativos a la forensia digital. Por ejemplo, el trabajo de (Al-Zarouni, 2004) describe en todo detalle el procedimiento de análisis forense a partir de la cabecera del correo electrónico pero no incluye una definición formal de los conceptos, términos o elementos que se utilizan, sin embargo incluye términos como los utilizados en OntoFoCE, como ser: *cabecera, trazabilidad, logs de servidores de paso, gestores de correos locales,* entre otros.

Por otra parte, el valor obtenido del 48% para el *Recall,* indica que del corpus considerado se toma menos de la mitad de los términos, pero si de este corpus se considera que 22 elementos hacen referencia a los subprocesos, protocolos y cadenas de comandos que se establecen para realizar el proceso de transmisión, el cual no es de interés para el análisis forense del correo electrónico, entonces el porcentaje de *Recall* llega a 92%.

Se puede concluir que el vocabulario definido en OntoFoCE es válido, en el marco de las restricciones encontradas para el corpus definido para el correo electrónico.

5.2.4 Validación de la Adecuación de la Ontología a los requerimientos

Según la propuesta de (Ramos et al., 2009) las actividades para evaluar la adecuación a los requerimientos son básicamente dos: verificar que las especificaciones del documento de requerimientos se cumplan, y verificar que las respuestas proporcionadas por la ontología a las preguntas de competencias sean correctas y pertinentes.

En una primera instancia se trabajó en la verificación del cumplimiento del objetivo general planteado para OntoFoCE: "*Comprobar la autenticidad del correo electrónico como prueba digital y en consecuencia la condición de no repudiabilidad de la prueba*", proponiendo como objetivos específicos los siguientes: "*Representar el proceso de transmisión de correo electrónico y derivar de ella la trazabilidad del envío"* e *"Identificar la información correspondiente a las cuentas y equipos que intervienen en la transmisión".*

El cumplimiento de estos objetivos se realiza logrando obtener respuestas válidas y ciertas sobre el análisis realizado a la evidencia digital, plasmadas en las repuestas de las preguntas de competencia.

Respecto de la representación del proceso de transmisión, en la sección 3.4.2 del capítulo 3, se describe detalladamente cómo se realiza esta representación y la derivación de la trazabilidad del envío.

Ello se explica mediante la instanciación de las cabeceras de ejemplo para los Escenarios 1, 2 y 3 del capítulo 3, en donde se describen las clases, relaciones y propiedades que se instanciaron. En cada ejemplo instanciado se realiza la identificación de los datos correspondientes a las cuentas y equipos que intervienen en la transmisión.

En la segunda etapa de esta misma fase, hay que verificar que las respuestas proporcionadas por la ontología a las preguntas de competencias sean correctas y pertinentes.

Para ello, en los distintos ejemplos descriptos en los escenarios del capítulo 3 se explica cómo se obtiene los datos necesarios para responder al punto de pericia solicitado por el Juez de las respuestas que brindan las preguntas de competencia.

Allí se detalla a nivel de código (consultas SPARQL) las respuestas obtenidos para el caso ejemplo, así como la verificación de las respuestas a las preguntas de competencia referidas específicamente a la trazabilidad.

Si se considera el punto de pericia formulado para el ejemplo en cuestión, que dice: *Determinar la existencia y veracidad de envío y recepción de los mails detallados en estos obrados,* se observa que las respuestas de las preguntas de competencia le permiten al perito responder con plena certeza acerca de lo solicitado en el punto de pericia, ya que las respuestas obtenidas a partir de la instanciación de las cabeceras identifican con claridad los datos que permiten certificar "…*la existencia y veracidad de envío y recepción…"* tal como se pide.

Las respuestas brindadas por las preguntas de competencia son los mismos datos que el perito buscaría en las cabeceras al realizar la pericia de manera manual, ya que la *existencia* de un correo se verifica cuando se identifica el equipo y la cuenta emisora así como el equipo y la cuenta receptora utilizada, y éstas son las respuestas brindadas por las preguntas de competencia P01, P02, P04 y P05.

En cuanto a la *veracidad de envío y recepción* ésta se comprueba cuando se establece el proceso seguido por el correo electrónico desde que sale de la cuenta emisora y hasta que es recibido en la cuenta receptora, y la respuesta de la P09 detalla el proceso de transmisión completo, indicando expresamente los servidores de paso.

Las preguntas de competencia se formalizan mediante consultas SPARQL y la ejecución de estas consultas permite obtener respuestas a dichas preguntas, cuyos resultados se muestran desde varios escenarios de pericias.

En el capítulo 3 sección 3.4 se describe el modelo conceptual de OntoFoCE y se particularizan las consultas SPARQL que representan cada pregunta de competencia utilizándose para ello la funcionalidad del editor SPARQL que ofrece Protégé, y considerando los ejemplos de cada escenario.

Así, se responden algunas preguntas desde el análisis forense de un correo en particular, y otros desde un conjunto de correos sobre los cuales se indican valores *filtro* para las preguntas de competencia que seleccionan y visualizan aquellas cabeceras que cumplen con la condición de filtrado.

En el capítulo 3 se ejemplifica las preguntas de competencia desde el resultado obtenido en el visor de consultas SPARQL de Protégé, mientras que en el capítulo 4 sección 4.7 se describe la funcionalidad de ObE Forensic, la aplicación basada en OntoFoCE, que muestran los resultados de cada pregunta de competencia tal como las visualiza el perito al utilizar la aplicación, siguiendo los mismos ejemplos para cada escenario definido en el Capítulo 3.

Vale mencionar en esta instancia que OntoFoCE se probó con tres escenarios diferentes, comenzando por el Escenario 1 con el ejemplo más simple (un correo con un emisor y un receptor), agregándole cierta complejidad en el Escenario 2 (un correo con un emisor y varios receptores) y multiplicando la cantidad de emisores y receptores en el Escenario 3.

Luego, en el capítulo 4 se agregó un Escenario 4 que consistió en probar el prototipo con un conjunto de 576 cabeceras provenientes de dos cuentas distintas. En todos los ejemplos, se pudo validar las preguntas de competencia formuladas en los escenarios, arribándose a los resultados esperados.

5.3 Validación de ObE Forensic por parte de Usuarios Expertos

Para reforzar la validación realizada según el método propuesto por (Ramos et al., 2009), se trabajó además en la validación de la aplicación informática ObE Forensics que permite realizar el análisis forense de un correo electrónico, y que se basa en OntoFoCE.

Las funcionalidades y características de ObE Forensic se describen en detalle en el Capítulo 4. La validación de dicha aplicación también fue realizada por un equipo de usuarios expertos a quienes se convocó expresamente para que utilizaran el prototipo inicialmente desarrollado, y con sus aportes y comentarios fue posible mejorar la calidad de respuesta de la aplicación.

En este caso los usuarios expertos son aquellos profesionales informáticos que actúan además como peritos.

Una vez desarrollada la aplicación en su primera versión, se invitó a los usuarios expertos a que la probaran a partir de casos reales de pericias de correos electrónicos en los que ellos hubieran participado, y de los cuales tienen además los resultados del análisis forense.

De este modo, se contrastó los resultados emitidos por ObE Forensic con los obtenidos por los peritos en cada caso.

Para utilizar datos de pericias reales, se pidió a los usuarios expertos que anonimizaran las cabeceras de los correos electrónicos que utilizarían para las pruebas, ya que las mismas son evidencia digital y es necesario preservar la identidad de las personas y/o instituciones que pudieran estar comprometidas.

Una vez que los peritos aceptaron probar ObE Forensic, se les entregó un Formulario de Experimentación que debían completar cada vez que validaran la aplicación, el cual se muestra en la Figura 5-8.

De los resultados de las pruebas experimentales realizadas se obtuvieron aportes, críticas y propuestas de mejoras, que se incorporaron en la aplicación a medida que los expertos iban participando de la prueba del prototipo.

Los aportes de los usuarios expertos se muestran en relación a los siguientes criterios de análisis:

- Validez de las preguntas de competencia
- Validez de los resultados brindados por ObE Forensic
- Valoración de la utilidad de la aplicación
- Valoración de las dificultades para realizar el experimento
- Consideraciones para optimizar la aplicación

A continuación se describe cada uno de ellos de los criterios de análisis, con los aportes destacados de los distintos usuarios y las modificaciones y/o agregados que se realizaron para mejorar la respuesta o funcionalidad de ObE Forensics.

UNIVERSIDAD CATÓLICA DE SALTA
FACULTAD DE INGENIERÍA
INSTITUTO DE ESTUDIOS INTERDISCIPLINARIOS DE INGENIERÍA (IEsIING)

PROYECTO DE INVESTIGACIÓN: **APLICACIÓN DE TECNOLOGÍAS SEMÁNTICAS A LA FORENSIA DIGITAL – ANÁLISIS FORENSE DE CORREOS ELECTRÓNICOS**

PROTOCOLO DE EXPERIMENTACIÓN

Requerimiento de Confidencialidad:
El experimento a realizar se encuentra sujeto a las disposiciones de confidencialidad y reserva de información, estipuladas en el PROTOCOLO OPERATIVO COMPLEMENTARIO N° 1, aprobado por RR N° 1068/18 de la UCASAL.

Breve descripción del experimento:
Se trata de probar una aplicación informática para realizar el análisis forense de correos electrónicos. La aplicación recibe como insumo o entrada la cabecera de un correo electrónico en formato de texto plano (*.txt)*, se realiza el proceso de instanciación en la ontología, y produce como resultado la respuesta a 21 preguntas de competencia que deberían permitir contestar los puntos de pericia.

Secuencia de pasos a seguir:
1) Ingresar a la dirección https://digilab.ucasal.edu.ar y acceder según el usuario y contraseña asignada.
2) Cargar el archivo .txt que debe contener la cabecera del correo (o correos a analizar).
3) Observar los resultados de las preguntas de competencia.

Registro de Resultados:
Contestar las siguientes consignas:
1) Fecha y hora de inicio y finalización del experimento.
2) Cuales preguntas de competencia responden a los puntos de pericia del caso?
3) Los datos obtenidos coinciden con el análisis forense realizado previamente para el caso?
4) Considerando la escala de 1 a 5, siendo 1 el valor de "nada" y 5 el valor de "totalmente", en que grado le resultó útil la aplicación?
5) Describa las dificultades que tuvo para realizar el experimento
6) En su opinión, describa los aspectos o partes que deberían mejorarse de la aplicación

Respuestas:

Figura 5-8: Formulario de Experimentación

Validez de las preguntas de competencia

Del conjunto de 21 preguntas de competencia, los peritos seleccionaron las siguientes para obtener los datos necesarios para dar respuesta al punto pericial correspondiente a cada caso.

- PC01: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?
- PC02: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?
- PC03: Dado un correo CE ¿A qué cuentas se remitió el correo?
- PC04: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del emisor?

- PC05: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del Receptor?
- PC06: Dado un correo CE ¿Cuál fue el cliente de correo utilizado por cada usuario?
- PC07: Dado un correo CE ¿Cuál fue el equipo desde el cual se emitió el correo?
- PC08: Dado un correo CE ¿Cuál fue el equipo en el que se recibió el correo?
- PC12: Dado una cuenta C1 ¿se ha emitido un correo hacia la cuenta C2?
- PC13: Dado una cuenta C1 ¿se ha recibido un correo desde la cuenta C2?
- PC21: ¿Cuáles son los correos intercambiados entre las cuentas C1 y C2 en un rango de fechas dado?

Uno de los peritos informó que "*…no fue posible aplicar el software puesto que no se encontró una funcionalidad que diera el soporte necesario para responder al punto de pericia en cuestión (Comprobar la autenticidad de dichos correos). Se buscó alguna opción que contemplara el análisis de validez de correos electrónicos a través de mecanismos como DKIM, SPF o DMARC pero no fue hallado...* ".

Los mecanismos DKIM, SPF o DMARC[68] que indicó el usuario hacen referencia a técnicas de autenticación de correo electrónico que permite al receptor comprobar que un correo electrónico fue realmente enviado y autorizado por el propietario del dominio de la cuenta emisora.

Si bien se consideran estas técnicas como adecuadas para comprobar la autenticidad de un envío, no invalida la capacidad de autentificación a partir de la trazabilidad del proceso de comunicación del correo electrónico, sino que la refuerza. Se está estudiando la viabilidad de implementar el mecanismo DKIM en ObE Forensic.

Validez de los resultados brindados por ObE Forensic

Si bien en la mayoría de los experimentos los peritos informan que los datos obtenidos sí coinciden con el análisis forense realizado previamente para el caso, resulta conveniente atender los siguientes casos:

Uno de los peritos informa que "*… si bien no hubo ningún caso en el que el sistema informe algo erróneo (falso positivo), sí existieron casos en los que no encontró información que sí estaba presente…*". El perito hace referencia aquí a que

[68] https://www.dmarcanalyzer.com/es/dkim-3/dkim-record-check/

el algoritmo no identificó una ocurrencia de transmisión intermedia, por lo que se realizó la prueba paso a paso de la ejecución del algoritmo para ese caso particular, identificando el error y ajustándolo debidamente para que no volviera a generarse el inconveniente.

Otro de los peritos informa que "*...Solo en algunos casos. Por ejemplo no responden ninguna de las preguntas...*". Durante la entrevista realizada con el experto a fin de identificar cuáles eran las preguntas que no respondió la aplicación, se observó que en realidad, la crítica estaba más orientada a las cuestiones de usabilidad de la aplicación. Las interfaces de comunicación con el usuario no eran lo suficientemente claras respecto de los botones habilitados para ver las preguntas de competencia para un único correo, respecto de los botones habilitados cuando se analiza un conjunto de correos. Es decir, hay preguntas de competencia (las indicadas como 10 a 21) que requieren de al menos dos cabeceras ingresadas, de lo contrario no se pueden responder. Esto no estaba suficientemente expresado en las interfaces de comunicación de la aplicación, situación que se modificó de inmediato.

Valoración de la utilidad de la aplicación

En su mayoría los peritos consideran la aplicación como provechosa, asignándole un puntaje promedio de 3 a 4 puntos, en el rango indicado de 1 a 5 (con 5 como valor máximo de utilidad).

Valoración de las dificultades para realizar el experimento

No se indicaron dificultades más allá del aprendizaje propio de encontrarse con una nueva aplicación.

Uno de los peritos indica que "*... Cuando se ingresa un mail sin cabecera la aplicación no genera un error...*". A este comentario se dio respuesta cuando se implementó la instanciación de la clase *CorreoFactible* que previamente había sido omitida en la primera versión de la aplicación.

Otro de los peritos indicó que "…*No permite cerrar la sesión del usuario logueado*…", situación que fue solucionada de inmediato ajustando la funcionalidad de la aplicación a este requerimiento.

También se informó acerca de "… *En la carga de cualquier dato, posiblemente tenga una vulnerabilidad de SQL INJECTION ya que cualquier dato que antecede con un " ' " provoca un error y por ello luego ya no se puede cargar. Un ejemplo de*

ello es en la carga de los datos de hardware y software. Dependiendo del lenguaje utilizado se debería prevenir a fin de evitar el robo de datos…".

Se tomó en cuenta este comentario para reforzar las reglas de validación de ingreso de los datos. Muchas de esas vulnerabilidades ya están resueltas por el framework Laravel. La más crítica de ellas, SQL INJECTION, no se aplicaría en este caso, ya que no se utiliza una base de datos SQL, sino un conjunto de tripletas RDF. Pero este comentario advierte acerca de si no es posible un caso de "inyección SPARQL", respecto de lo cual hay opiniones encontradas[69]. Teniendo presente esta situación, se dispuso ajustar las consultas no utilizando la función SELECT con el parámetro "*" de manera que el usuario solo podría inyectar el código pero no vería su resultado.

Por otra parte, la aplicación cuenta con un espacio de almacenamiento propio por cada usuario, con lo cual se restringe el acceso a los datos de otros usuarios.

Consideraciones para optimizar la aplicación

Se realizaron varias sugerencias, entre las que se destacan las siguientes:

- *La aplicación podría enriquecer su funcionalidad permitiendo concatenar filtros para el tratamiento de los correos. Por ejemplo "filtrar aquellos que correspondan al período X" concatenado con "de lo filtrado, extraer aquellos correos que hayan transitado por la IP xx.xx.xx.xx".*
- *Si los correos electrónicos no se encuentran en la misma carpeta, la aplicación no permite elegir más de uno.*

Ambas propuestas se toman para trabajar a futuro, previo a la puesta en producción de ObE Forensics.

Como conclusión de esta validación, puede decirse que la ObE Forensic responde a los objetivos que persigue: representar la trazabilidad del proceso de transmisión e identificar los datos de las cuentas intervinientes.

Aun así, es necesario atender las sugerencias de mejora aportadas por los usuarios expertos, así como continuar realizando pruebas de validación, particularmente para considerar casos de estudio reales que permitan validar el resto de las 21 preguntas de competencia, que solo se respondieron con el banco de

[69] https://www.owasp.org/images/0/0f/Onofri-NapolitanoOWASPDayItaly2012.pdf

pruebas hasta ahora utilizado en los distintos escenarios ejemplificados en los capítulos 3 y 4.

Existe la posibilidad de poner en producción ObE Forensics, a través de la figura de asistencia técnica y servicios a terceros, y ofrecerla al área del Ministerio Público y Fiscalía de los ámbitos de la justicia provincial y federal, para lo cual, se ha previsto continuar con la validación de la aplicación por parte de usuarios expertos con casos reales, a fin de completar la aplicación y fortalecer la propuesta.

5.4 Consideraciones sobre la completitud de los resultados que brinda ObE Forensic

Una cuestión que debe destacarse es que, en las primeras pruebas de efectividad de ObE Forensics, frente a otras herramientas existentes para el análisis forense de correos electrónicos, se encontró que los resultados emitidos por la aplicación aquí propuesta son más completos que los emitidos por otras herramientas.

Tomando el conjunto de herramientas no propietarias de uso habitual por parte de los peritos, el software *MailXaminer* es el que mejor se comporta al momento de realizar una pericia de este tipo de evidencia digital, de modo que se consideró realizar una prueba comparativa entre dicho software y ObE Forensics, considerando el caso de estudio planteado para el Escenario 1 del capítulo 4.

Así, en este primer caso comparativo, se encontraron dos elementos en los que ObE Forensics trabaja mejor. El primero de ellos se refiere a la cantidad de ocurrencias que se identificaron y el segundo a la identificación de la cuenta receptora.

Para el primer caso la Figura 5-9 muestra los resultados obtenidos al analizar la cabecera de ejemplo del Escenario 1 en la herramienta *MailXaminer*. Obsérvese que dicha herramienta identifica 6 ocurrencias, mostrando con claridad la igualdad entre el valor del sub-atributo *by* de una línea con el valor del sub-atributo *from* de la línea anterior.

Pero ObE Forensics identifica 7 ocurrencias (ver Figura 4-43), también teniendo presente el valor del sub-atributo *by,* es decir en la Figura 5-9 la herramienta *MailXaminer* no considera todas las apariciones del sub-atributo *by* que figuran en la cabecera.

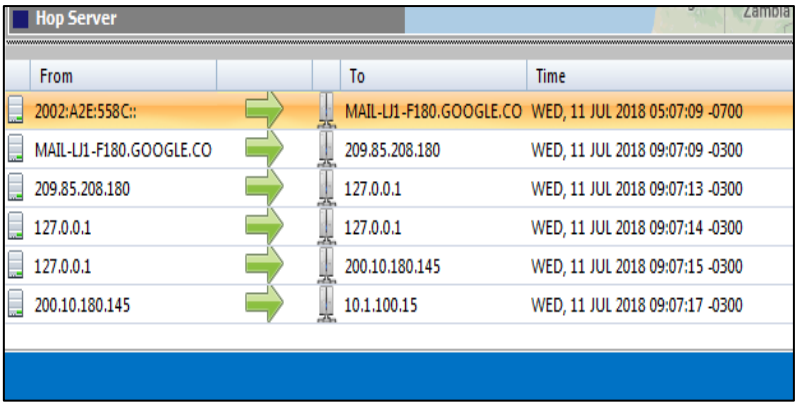


Figura 5-9: Ocurrencias identificadas por *MailXaminer* para el caso ejemplo

Para esta cabecera ObE Forensics identifica 12 ocurrencias, que se muestran en la Figura 5-10.

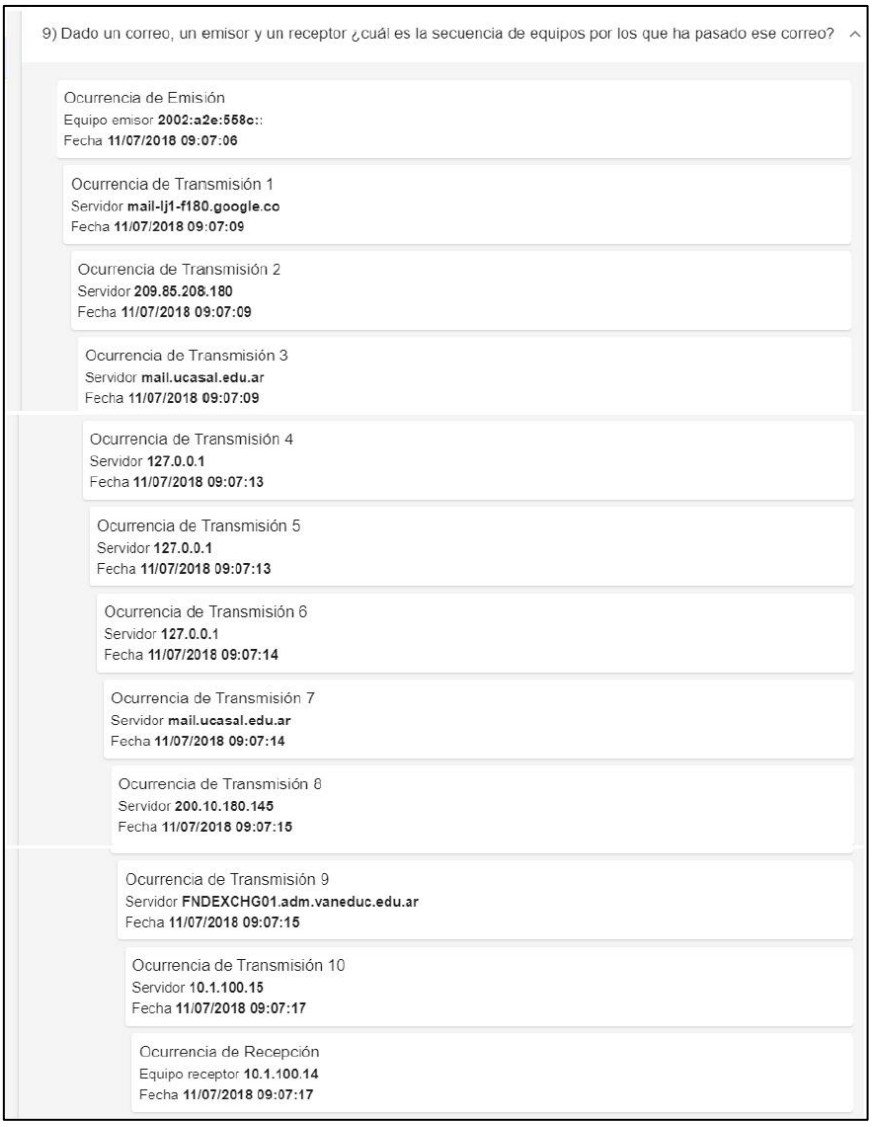


Figura 5-10: Ocurrencias identificadas por ObE Forensics para el caso ejemplo

En la Figura 5-10 se observa que aun si se considera que algunas ocurrencias podría tomarse como idénticas, de todos modos la aplicación propuesta en esta tesis

identifica dos nombres de dominio (*mail.ucasal.edu.ar* y *FNDEXCHG01.adm.vaneduc.edu.ar)* y una dirección IP (10.1.100.14), valores éstos que *MailXaminer* no informa.

Por otra parte, tampoco hay coincidencia respecto de la identificación del equipo receptor, ya que ObE Forensics identifica la dirección IP 10.100.1.14 como la correspondiente al equipo receptor, mientras que *MailXaminer* identifica la dirección IP 10.100.1.15.

Asimismo, ObE Forensics identifica la cuenta receptora, aun cuando ésta no figura en los parámetros *Delivered-To o To,* y cuando están ausentes en la cabecera, el algoritmo recorre las primeras ocurrencias hasta encontrar el sub-atributo *for* que contiene como valor la cuenta que actúa como receptora. La Figura 5-11 muestra los datos de identificación del correo emitidos por *MailXaminer*, mientras que la Figura 5-12 señala los mismos valores para ObE Forensics.

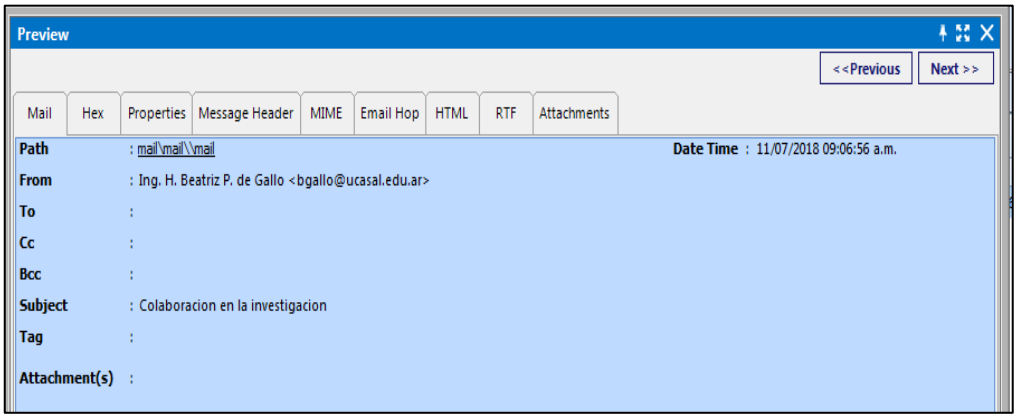


Figura 5-11: Datos de Identificación mostrados por *MailXaminer*

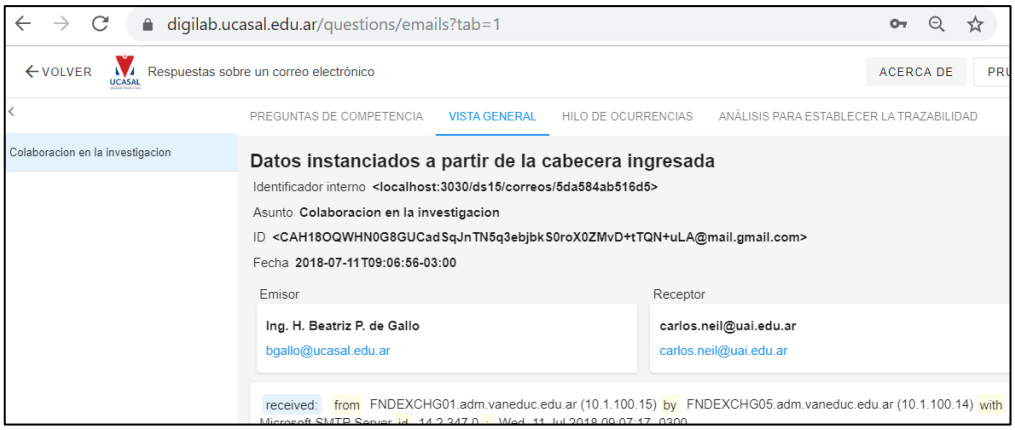


Figura 5-12: Datos de Identificación mostrados por ObE Forensics

Por supuesto que estos hallazgos no son concluyentes, ya que solo se avanzó en una prueba de ejemplo, pero se prevé investigar con otros escenarios, profundizando el análisis no solo en cantidad de correos, sino también considerando los escenarios

descriptos en el capítulo 3 y 4, y otros que pudieran plantearse. Asimismo, sería interesante obtener índices de rendimiento de ObE Forensics respecto de otras herramientas de análisis forense de correos electrónicos.

5.5 Conclusiones del Capítulo

Como conclusión puede decirse que el método de evaluación de ontologías propuesto por (Ramos et al., 2009) resultó un elemento útil para validar OntoFoCE desde los cuatro criterios propuestos. La validación realizada respecto del uso correcto del lenguaje, como de la taxonomía, vocabulario y adecuación a los requerimientos resultó pertinente para lograr el grado de madurez actual de la ontología, sin perjuicio de que pudieran existir mejoras que solo se pueden percibir con la puesta en producción del modelo ontológico para nutrirse de experiencias y casos de uso superadores de los considerados hasta ahora.

Por otra parte, la validación de ObE Forensics realizada por usuarios expertos permitió ajustar aspectos de usabilidad, además de mejorar los criterios de seguridad informática requeridos por el tipo de información que se procesa en la aplicación.

Y el análisis de rendimiento comparativo entre ObE Forensics y MailXaminer presenta un panorama de desafío para iniciar acciones concretas, mediante protocolos de prueba formales, destinados a medir la eficiencia de la herramienta propuesta en esta tesis.

CAPÍTULO 6. CONCLUSIONES

En este capítulo corresponde señalar las conclusiones más relevantes que resultan del trabajo de investigación realizado, destacando los resultados obtenidos en la Tesis y formular las líneas de investigación que se derivan de la misma.

6.1 Principales Contribuciones de la Tesis

Los aportes que esta Tesis brinda se pueden describir desde cada uno de los capítulos que muestran el proceso de definición y construcción de OntoFoCE y ObE Forensics.

En el **Capítulo 2**, dedicado a la definición del marco teórico de la investigación se abordó el estado del arte de las Tecnologías Semánticas aplicadas a la Forensia Digital, con foco en el análisis forense del correo electrónico.

A modo de síntesis se puede tomar el trabajo de (Garfinkel, 2010) en el que señala los desafíos que enfrenta la Forensia Digital: diseño de las herramientas orientadas a la evidencia; modelo de visibilidad, filtro e informe que proponen; problemas estructurales que presentan estas herramientas; características de abstracción y modularidad que ofrecen; y enfoque en la identidad del individuo. El autor señala dos características importantes en estas herramientas: la falta de integración con estrategias de procesos como la ingeniería reversa que permitirían reducir tiempos y costos, y los inconvenientes de visualización de resultados en términos fácilmente interpretables por los profesionales de la justicia, por otra parte, se observó que –en su mayoría- las herramientas forenses actuales se orientan a la búsqueda de la evidencia digital pero no ayudan en la presentación o análisis de correlaciones entre los datos encontrados.

Al respecto, la herramienta forense propuesta en esta Tesis responde a los retos indicados por (Garfinkel, 2010). Por una parte, el diseño de ObE Forensics está enfocado a la evidencia, en particular al correo electrónico que se aporta como prueba judicial. La herramienta ofrece una estructura de visibilidad, filtro e informe que brinda al Perito un ambiente idóneo para establecer vínculos y/o relaciones de interés, mostrando los datos mediante una interface de comunicación sencilla.

ObE Forensics es una aplicación desarrollada expresamente para el análisis forense de correos electrónicos y responde a las necesidades propias de este tipo de evidencia digital. Su arquitectura de procesamiento se basa en herramientas no propietarias, cumpliendo con criterios de abstracción y modularidad.

Por último, ObE Forensics considera el conjunto integral de datos del análisis forense de correos electrónicos, enfocados hacia la identidad del individuo, considerando los valores de cuentas y equipos que pueden asociarse a los usuarios partícipes del correo electrónico que se aporta como prueba.

El marco teórico profundiza el estudio de las ontologías. Esto permitió desarrollar un producto que cumpla con los criterios que identifican una ontología según los autores considerados:

- *"Una ontología define las condiciones básicas y relaciones que comprenden el vocabulario de un área del tema así como las reglas para combinar condiciones y las relaciones para definir extensiones del vocabulario"* (Neches et al., 1991).
- *"Una especificación explícita de una conceptualización, es decir, que proporciona una estructura y contenidos de forma explícita que codifica las reglas implícitas de una parte de la realidad; estas declaraciones explícitas son independientes del fin y del dominio de la aplicación en el que se usarán o reutilizarán sus definiciones"* (Gruber, 1993b).
- *"La ontología describe una cierta realidad con un vocabulario específico, usando un conjunto de premisas de acuerdo con un sentido intencional de palabras del vocabulario"* (Guarino, 1997).

Así, OntoFoCE responde a las condiciones básicas y relaciones que comprenden el vocabulario referido al correo electrónico como objeto de estudio; cuenta con la estructura y contenidos explícitamente definidos a partir de la realidad que se está modelando o sea, el correo electrónico como evidencia digital; y utiliza un conjunto de premisas que respetan el sentido intencional de los términos del vocabulario que representa, esto último se observa particularmente en el modo en que se representó el proceso de transmisión del cual se deriva la trazabilidad del correo.

Considerando el correo electrónico en sí mismo, en el marco teórico se profundizó la arquitectura de procesamiento, así como el procedimiento de análisis forense de este artefacto.

En esta área, se puede tomar como principal aporte de la investigación realizada, la definición y formalización de la *trazabilidad del proceso de transmisión* como base para considerar la validez de esta evidencia digital.

Asimismo, el estudio realizado sobre los puntos de pericia para sistematizarlos y relacionarlos con las preguntas de competencia de OntoFoCE también es meritorio, pues sienta un antecedente único en su tipo, no solo por el resultado de vinculación de 86 puntos de pericia iniciales en 21 preguntas de competencia, sino por la participación de usuarios expertos para la identificación de los puntos de pericia sobre correos electrónicos.

Por último, debe destacarse que la revisión bibliográfica realizada señala el interés de la comunidad científica en estudiar la Forensia Digital, principalmente por la necesidad de formalizar las metodologías y herramientas generadas desde la experiencia del análisis forense, agregándoles un sustento científico válido para sustentar la validez de la tarea pericial.

Por otra parte, de la revisión realizada se muestra los principales aspectos encontrados:

- En la comunidad científica existe una marcada preocupación por desarrollar una respuesta adecuada al avance del cibercrimen o ciberdelito.
- Las tecnologías semánticas están siendo utilizadas por los investigadores del tema para formular modelos que soporten científicamente la actividad pericial.
- Existen numerosos estudios sobre diversos aspectos del análisis forense, entre los que se destacan: estudios comparativos de herramientas, formulación de técnicas o métodos forenses, estudio de casos de uso concretos, herramientas para el análisis forense de dispositivos varios, entre otros.
- No se encontraron estudios que aborden directamente la verificación o validación del proceso de transmisión del correo electrónico, siendo que en la generalidad de las veces, ésta es la actividad técnica central del perito, y sobre este proceso se sustenta la no repudiabilidad de la evidencia digital.
- Las herramientas de análisis forense se limitan a generar los resultados en términos técnicos, dejando a cargo del perito la interpretación de los mismos a la luz de los puntos periciales.

El **Capítulo 3** aborda la construcción de OntoFoCE, considerando Methontology como metodología guía para su desarrollo, cumpliendo con las fases de especificación, formalización, conceptualización, implementación y validación del modelo ontológico en un esquema iterativo que permitió el ajuste gradual y consistente de OntoFoCE hasta su versión actual.

Del proceso de desarrollo se destaca particularmente la especificación del dominio realizada con la participación de un grupo de usuarios expertos, todos ellos profesionales informáticos dedicados a la forensia digital, que colaboraron principalmente en la definición de los puntos de pericia que luego se derivan en las preguntas de competencia.

En cuando a OntoFoCE -el aporte sustancial de esta tesis- se desarrolló a partir de la conjunción de las Tecnologías Semánticas con la Forensia Digital, cumpliendo con el objetivo propuesto de *comprobar la autenticidad del correo electrónico como prueba digital y en consecuencia la condición de no repudiabilidad de la prueba*, basado en la representatividad del proceso de transmisión del correo electrónico, del cual se deriva la trazabilidad del envío.

Esta representatividad se muestra en las sucesivas vistas que describen las 3 fases de la transmisión: envío, trayecto y recepción del correo electrónico.

Asimismo se observa que si bien en varios trabajos se aborda el sistema de transmisión del correo electrónico considerando el encabezado del mismo, en ninguno se aplica el concepto de *trazabilidad* como elemento vinculante de los distintos equipos o servidores utilizados en la transmisión.

Las guías procesales para peritar correos electrónicos establecen que se debe acceder al correo recibido y verificar los equipos de emisión y recepción mediante la dirección IP que figura en la cabecera del mismo, realizando un recorrido inverso del proceso de transmisión para llegar desde el equipo receptor al equipo emisor.

Este recorrido inverso puede sostenerse técnicamente si se aplica el concepto de *trazabilidad del proceso de transmisión*, y puede sostenerse científicamente si dicha trazabilidad se representa mediante una ontología.

En este capítulo se trabajó particularmente el tema de la identificación de todas las direcciones IP y hostnames que pudieran identificar a un equipo/servidor utilizado. Este proceso se considera un aporte destacado de esta tesis, toda vez que permite una completitud del análisis de la cabecera que es mucha utilidad para el Perito.

También se abordó en detalle las preguntas de competencia que brinda OntoFoCE, considerando 3 escenarios diferentes de pericias sobre correos electrónicos, aunque se agrega un último escenario más complejo en el capítulo siguiente.

Desde la más sencilla –considerando un único correo- hasta la más compleja –considerando una cuenta de correo-, los distintos ejemplos sirvieron para describir el código SPARQL correspondiente a cada una de las 21 preguntas de competencia, y se comprobó el resultado mediante la generación de las instancias de las clases que representan cada ejemplo.

A partir de una instanciación manual, se probaron las preguntas de competencia con un banco de pruebas inicial de 8 cabeceras, que actuaban sobre 5 cuentas de correo, luego, en el capítulo 4, se trabajó en el último escenario simulando una pericia sobre una cuenta con 576 correos electrónicos.

Del **Capítulo 4** se puede identificar como aporte más destacado, la construcción de la aplicación informática denominada ObE Forensics, basada en OntoFoCE. Esta herramienta para el análisis forense, cuenta con las características propias de un desarrollo web, incluidos los requerimientos de seguridad informática, que se refuerzan debido al carácter de datos reservados que se toman como insumo para el análisis forense de correos electrónicos.

Otra característica destacable de ObE Forensics, es que a través de las 21 preguntas de competencia que se responden, se contestan a la gran variedad de puntos de pericia relevados.

Los ejemplos trabajados en los distintos escenarios que se describen en este capítulo, muestran la facilidad de uso de la herramienta, así como la versatilidad para procesar y visualizar los datos de la cabecera analizada. Estas ventajas se resaltan cuando el Perito requiere procesar un conjunto masivo de correos electrónicos.

La independencia de los gestores de correos y de la estructura interna que éstos definen para el procesamiento de los correos, es una de las principales ventajas de ObE Forensics, y esto es así porque la ingeniería ontológica permite la representatividad del correo electrónico como objeto de estudio, con abstracción de toda restricción de software y/o estructura de datos.

Por último, se debe destacar que la aplicación ObE Forensics fue desarrollada y funciona en un contexto de herramientas no propietarias. Ello impactó de manera

positiva, no solo en los costos de desarrollo del prototipo, sino además en la factibilidad técnica de la puesta en producción del mismo.

La instancia de validación de OntoFoCE y ObE Forensics, que se describe en el **Capítulo 5,** muestra la conformidad de estas herramientas, ajustadas a normas que verifican la calidad de la ontología.

Por una parte, la utilización de una metodología integrada para validar el uso del lenguaje, la estructura taxonómica, la validez del vocabulario y el grado de adecuación a los requerimientos, permitió lograr la validación necesaria y suficiente de OntoFoCE.

La validación se completa con la prueba del prototipo a cargo de usuarios expertos, con casos de uso reales, mediante un protocolo de experimentación en los que se evalúan diferentes características de la aplicación, que luego se tomaron como insumo para la mejora de la misma.

Por último el capítulo 5 incluye consideraciones acerca de la completitud de la herramienta propuesta en esta tesis, presentando un estudio comparativo entre los resultados obtenidos con *MailXaminer* y los obtenidos con ObE Forensics para el mismo caso de estudio, en el cual se observa que esta última herramienta brinda más resultados y con mejor detalle que *MailXaminer*.

6.2 Trabajos Futuros

En particular, y enfocados en OntoFoCE y ObE Forensics, la investigación realizada debe profundizarse para trabajar distintas características y/o componentes que no se incluyeron en la presente Tesis, pero que se vislumbran como trabajo a futuro:

- OntoFoCE representa un modelo del dominio de la forensia de correos electrónicos, basado en los 86 puntos de pericia que luego derivaron en 21 preguntas de competencia. Sería oportuno considerar la inclusión de otras preguntas –resultantes de nuevos puntos de pericia- que no se consideraron en el modelo inicial.

A eso se puede dar respuesta con un *Editor* que haga posible armar o generar nuevas preguntas de competencia a partir de la estructura que hoy tiene OntoFoCE, es necesario discutir en ese momento si el actual modelo es suficiente para soportar cualquier pregunta de competencia.

- Un aspecto que no se abordó en esta tesis es la reusabilidad de OntoFoCE, ya sea como parte de otros modelos que estudien las evidencias digitales desde las tecnologías semánticas, como de la incorporación de otras ontologías que aborden métodos, taxonomías y otros componentes vinculados al derecho y a la forensia digital. En particular, OntoFoCE permite representar la trazabilidad del proceso de transmisión de un correo electrónico, y el modelo podría ser válido para cualquier otro tipo de artefacto forense (chat, redes sociales, entre otros) en el cual se encuentre presente un proceso comunicacional (emisor, receptor, canal, mensaje). Y con los ajustes necesarios, podría utilizarse para representar la trazabilidad de procesos de transmisión o envío de otras entidades u objetos tales como documentos, fluidos, objetos varios.

- Si bien OntoFoCE incluye la búsqueda de *palabras claves* solicitadas en los puntos de pericia, indagando en el Asunto, Cuerpo y Adjuntos del correo electrónico, la aplicación ObE Forensics no incluye esta funcionalidad.
  Debido principalmente a que se debe estudiar en profundidad tres características de los tres componentes citados: los métodos de encriptación y cifrado utilizado por los gestores de cuentas de correo al momento de enviar el mismo y sus adjuntos; los diversos formatos de archivo que puede contener los adjuntos del correo electrónico y el volumen de datos que dichos archivos suponen.
  Desde el punto de vista práctico, la diversidad de alternativas tecnológicas que presenta este campo, amerita un estudio profundo y específico que supera los objetivos propuestos para esta Tesis.

- En esta Tesis se plantea un principio básico para representar la autenticidad del correo electrónico que dice: *Un correo electrónico es auténtico cuando se identifican los datos del remitente (cuenta de correo y dirección IP), la*

trazabilidad del mismo (diferentes dispositivos que intervienen en la transmisión) y los datos del destinatario (cuenta de correo y dirección IP).

Este principio se puede reforzar si se considera el carácter de *inalterabilidad* del mensaje, es decir, si se comprueba que el contenido del mensaje recibido no es distinto al contenido del mensaje enviado.

Cuando se comenzó a investigar acerca de representar en la ontología este carácter de inalterabilidad, se encontró que si bien –en esencia- se trata de un proceso de comparación de contenidos de archivos, el impacto que en ellos tienen la estructura de datos que utilizan los diferentes gestores de cuentas de correo para almacenar los correos enviados y recibidos, no hace sencilla la tarea. A lo que se agrega, la selección cuidadosa de las metodológicas de encriptación de datos y de comparación de contenidos que deberán estudiarse para seleccionar la más conveniente.

- Relacionado con este último punto, se discutió acerca de la factibilidad de obtener datos sobre los servidores de paso utilizados durante el proceso de transmisión. Por una parte, en la actualidad, es muy poco probable que los propietarios de dichos equipos pongan a disposición de la justicia los contenidos que albergan, debido principalmente a la maraña de instancias judiciales de carácter internacional que ello implicaría.

  Por la otra, las cuestiones relativas a la seguridad informática y a la protección de datos de los contenidos de estos servidores, también imponen barreras tecnológicas que no son sencillas de abordar.

  Durante el desarrollo de esta tesis se discutió acerca del *grado de certeza* que se puede brindar al Juez respecto de la relación entre la dirección IP y el servidor físico que supuestamente está asociada a dicha dirección, como ejemplo, se planteó el caso de la imposibilidad de aseverar con plena certeza que si la cabecera señala dos ocurrencias con la misma dirección IP, ambas se encuentran en el mismo servidor.

  La representación de esta situación en OntoFoCE está salvada, porque el caso se probó en distintas instancias y el modelo reacciona correctamente.

  Pero queda la incertidumbre de probar si cuando las ocurrencias señalan a la misma IP, ésta corresponde al mismo servidor físico.

Tal vez si se aborda el estudio del *criterio de inalterabilidad* propuesto para el correo electrónico entonces pueda dilucidarse esta cuestión.

- ObE Forensics está en etapa de prototipo, de la puesta en producción para la utilización continua de esta herramienta seguramente surgirán cuestiones de usabilidad y respuesta a nuevas situaciones que se deberán atender.
  Por otra parte, para la puesta en producción se debe generar un contexto totalmente seguro dado que se procesa evidencia digital, que puede impactar en la libertad de las personas, por ello será importante trabajar todas las normas de calidad relativas a la construcción y puesta en funcionamiento del laboratorio forense en donde funcionará ObE Forensics.

- Se iniciaron trabajos comparativos sobre la eficiencia de ObE Forensics frente a otras herramientas específicas para el análisis forense de correos electrónicos, si bien no se arribaron a resultados concluyentes, los avances indican que es posible que se puedan mejorar los tiempos de procesamiento de la aplicación, aun considerando que al ser un producto web, los tiempos de conectividad inciden en los rendimientos que se puedan obtener, así como avanzar con técnicas de experimentación controladas para comparar el funcionamiento de ObE Forensics frente a otras herramientas.

- La instanciación dinámica de la ontología de correos electrónicos propuesta por (Kota, 2012) se debería estudiar en profundidad, a fin de aprovechar esta experiencia para mejorar OntoFoCE.

- Otro criterio de mejora sobre la aplicación, se puede observar en el trabajo de (Ovens & Morison, 2016) sobre “Identification and Analysis of Email and Contacts Artefacts on iOS and OS X” en el que se consideran múltiples dispositivos que comparten una única cuenta.

La revisión bibliográfica realizada para esta Tesis permitió identificar áreas de vacancia, particularmente en la aplicación de las Tecnologías Semánticas a la Forensia Digital.

Se encontraron investigaciones con contribuciones destacadas para la formalización científica de la Forensia Digital, focalizadas en cuatro áreas de interés: direccionamiento de cuentas, desarrollo de herramientas forenses, estrategias para mitigación de ataques y representación semántica de artefactos forenses.

La revisión bibliográfica permitió la identificación de tres áreas de interés. El estudio del *Trafico de Redes*, particularmente en la identificación de direcciones IP; aplicaciones de la *Minería de Datos* con énfasis en la minería de textos y técnicas de extracción de datos; y estudios referidos a la *Seguridad Informática* que se enfoca principalmente en la atención de ataques cibernéticos.

Mientras que las áreas con escasa aplicación en la forensia de correos electrónicos serían las siguientes: *Big Data*, *Tecnologías Semánticas* y *Procesamiento de Lenguaje Natural*, encontrándose pocos trabajos orientados a la definición de ontologías de correos electrónicos.

Particularmente se prestó atención a las escasas investigaciones dirigidas a estudiar el análisis forense de correos electrónicos basados en los metadatos de la cabecera, y de este grupo, solo en un trabajo se abordó detalladamente el proceso de transmisión.

Este estado del arte permite definir futuras líneas de investigación sobre Forensia Digital, entre las que se destacan las siguientes.

Si bien se encontraron estudios sobre la aplicación de las Tecnologías Semánticas a la Forensia de artefactos forenses usuales, como celulares, redes sociales y dispositivos móviles, aún son incipientes las investigaciones referidas a la aplicación de las Tecnologías Semánticas a los *nuevos* artefactos forenses, como sería el caso de Internet de las Cosas, en donde se conjugan todos los artefactos forenses ya conocidos más el agregado de sensores, actuadores y tecnologías de identificación y etiquetado como RFID, por citar algunos de los componentes de comunicación que se integran a la Forensia Digital a partir de Internet de las Cosas.

Desde el lado de la Forensia Digital, es necesario seguir trabajando la formalización científica de herramientas y métodos, así como la atención de *nuevos*

artefactos forenses. El Informe Garnet[70] 2019 señala las tecnologías emergentes que se destacan para los años venideros.

De las diez señaladas, algunas son de especial interés para la Forensia Digital porque son los futuros artefactos forenses sobre los que descansará la evidencia digital, por ejemplo:

- Cosas Autónomas, hace referencia a cinco tipos de componentes (robótica, vehículos, drones, accesorios y agentes) manejados por la Inteligencia Artificial, y que se desarrollan en 4 ambientes: mar, tierra, aire y digital.
- Gemelos digitales, nombre que recibe la representación digital que refleja un objeto, proceso o sistema de la vida real.
- Espacios Inteligentes, referido al entorno físico o digital en el que los humanos y los sistemas habilitados por la tecnología interactúan en ecosistemas cada vez más abiertos, conectados, coordinados e inteligentes

[70] Extraído de https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2019/, consultado el 23/07/19.

ANEXO I: INVESTIGACIÓN BIBLIOGRÁFICA

Esta sección incluye el trabajo desarrollado en la búsqueda del estado del arte de las temáticas involucradas en esta tesis: Forensia Digital, Ontologías y Trazabilidad. Mediante un sistema de búsqueda y selección de textos, basado en normas sistemáticas y de buen orden, se explica la labor bibliográfica realizada.

I.1 INTRODUCCIÓN

En el caso particular de los correos electrónicos, el análisis forense se realiza sobre la cabecera del mail, obteniéndose un volumen de datos técnicos de difícil interpretación para el lego, y deben seleccionarse y mostrarse en el marco del resto de las pruebas de la causa judicial, ofreciendo un informe técnico que permita la interpretación de los resultados a la luz de la causa, por parte de los profesionales de la criminalística y el derecho.

Se requiere mucho más que la identificación de una dirección IP del correo electrónico. Hoy en día se exige que estos datos se presenten sistemáticamente y semánticamente en el marco de la causa judicial, en el mismo espacio de análisis que el resto de los elementos probatorios. Y en particular, las ontologías resultan una herramienta pluridisciplinar para facilitar el análisis de la prueba documental, por parte de todos los actores (abogados, jueces, investigadores y peritos).

En este contexto resulta de interés realizar un estudio sistemático del estado del arte en ambas áreas: ontologías y forensia digital, y particularmente sobre la aplicación de tecnologías semánticas a la forensia de correos electrónicos.

Respecto del tipo de estudio, se pretende efectuar una investigación exploratoria para identificar el estado del arte sobre la aplicación de las tecnologías semánticas a la Forensia Digital, particularmente en la forensia de correos electrónicos, realizando un estudio crítico y ajustando el alcance del mismo a los objetivos propuestos. Éstos últimos se mencionan a continuación:

- Identificar y estudiar los aportes investigativos más actualizados sobre Ontologías y Forensia Digital.

- Establecer las áreas de vacancia sobre la aplicación de las tecnologías semánticas a la Forensia Digital
- Relacionar los trabajos desde atributos de cercanía (o distancia) con la aplicación de tecnologías semánticas para el análisis forense de correos electrónicos.

I.2 MÉTODO DE REVISIÓN BIBLIOMÉTRICA

A fin de realizar una revisión bibliográfica ordenada y ajustada a norma, se analizaron diversas metodologías de revisiones de trabajos científicos, identificando los criterios más útiles para la temática en estudio.

(Brereton, Kitchenham, Budgen, Turner, & Khalil, 2007) definieron una metodología para la revisión de la literatura vinculada a la ingeniería de software, basándose en métodos de revisión provenientes de la medicina. Su principal aporte está en la definición del paradigma basado en la evidencia, que promueve la evaluación objetiva y la búsqueda de resultados empíricos relevantes sobre un tema de investigación. Siguiendo esta misma línea de generar métodos para la revisión bibliográfica en la ingeniería de software, (Biolchini, Mian, Natali, & Travassos, 2005) presenta su método de 3 pasos, que permite transitar de los conceptos hacia los estudios que pueden proporcionar evidencia sobre el tema en cuestión (fase 1), luego se analizan comparativamente los contenidos de esas publicaciones, para generar nuevo tipo de evidencia si fuera posible (fase 2), y por último se arriba a las conclusiones, que podría significar la obtención de nuevos conocimientos.

También se consultó la metodología propuesta por (Grant & Booth, 2009), en la que se realiza un análisis comparativo entre 14 tipos diferentes de revisión bibliográfica (siempre en el área de la Medicina), identificado por una parte los tipos de revisiones (crítica, literaria, mapeo sistemático, meta-análisis, estudios mixtos, de visión general, revisión por alcance, etc.), mostrando para cada una las fortalezas, debilidades y conveniencia de utilización.

Por su parte (Velásquez, 2015) sintetiza las metodologías propuestas por tres autores (Kitchenham, Sorrell y Tranfield) identificando las tres fases más comunes involucradas en una revisión: planeamiento (en la que se propone la justificación, motivación y diseño del protocolo de búsqueda), la ejecución (que incluye los

procesos de búsqueda, selección, evaluación de calidad, extracción y síntesis de resultados) y el reporte final de la revisión realizada.

Para (Medina Lopez et al., 2010) la búsqueda bibliográfica ajustada a norma implica considerar al menos 5(cinco) fases de trabajo:

1) Identificación del campo de estudio y del período a analizar;

2) Selección de las fuentes de información;

3) Realización de la búsqueda (qué, dónde y cómo);

4) Gestión y depuración de los resultados de la búsqueda; y

5) Análisis de los resultados.

Durante el desarrollo de estas etapas se deben considerar además criterios de éxito, tales como: establecer con claridad el objetivo que se persigue, documentar el proceso, establecer parámetros de cualificación comparables, entre otros.

En (Portugal et al., 2018) los autores proponen definir restricciones para la inclusión de trabajos a revisar, con el objetivo de limitar el alcance del estudio en función de la estructura de las publicaciones (descripción de un experimento por ejemplo). Y también proponen definir criterios de exclusión referidos al tipo de trabajo a considerar (tesis, patentes, etc.).

Comparando las distintas metodologías analizadas, se encontró que la propuesta de (Medina Lopez et al., 2010) profundiza lo dicho por (Velásquez, 2015) y supera a las metodologías de (Brereton et al., 2007) y de (Biolchini et al., 2005) al presentar un esquema más estructurado para realizar la revisión; mientras que la propuesta de (Grant & Booth, 2009) sirvió para definir con mejor criterio el tipo de estudio a realizar.

A partir de lo dicho sobre el tipo de estudio, los objetivos que se persiguen y las metodologías analizadas, se optó por tomar como guía la metodología propuesta por (Medina Lopez et al., 2010) y lo dicho en (Portugal et al., 2018) para realizar una revisión del estado del arte de la aplicación de las ontologías en problemáticas de forensia digital de correos electrónicos. Así, se propuso un método de revisión sistemática para el estudio que se introduce en el presente trabajo basado en las siguientes fases:

1) Definición del Marco de Estudio y Alcance de la Revisión y

2) Procesos de Búsqueda y Selección.

I.2.1 Definición del Marco de Estudio y Alcance de la Revisión

A fin de cumplir con los objetivos propuestos para la revisión, resulta necesario definir los atributos o palabras claves que delimiten el *marco de estudio*, siendo términos con suficiente fuerza como para guiar los procesos de búsqueda.

Se partió de la conjunción de dos temáticas principales: las tecnologías semánticas y la forensia digital. Ambas áreas, de amplísimo desarrollo por sí solas, se fusionan en trabajos de investigación particulares que cuentan cada uno de ellos con sus propios objetivos, en los que –en la generalidad- se observa la aplicación de ontologías en la resolución de problemas de la forensia digital.

Aun considerando el marco teórico específico de estudios de aplicación de ontologías a la forensia digital, se definió un siguiente nivel de detalle centrado en la entidad u objeto de la pericia. Particularmente, interesaba identificar los últimos aportes referidos a los métodos, herramientas y artefactos forenses (se denomina así a los distintos componentes en los que reside la evidencia digital, sea éste un dispositivo de hardware o software), vinculados con correos electrónicos.

Se definieron los criterios de búsqueda mediante reglas de decisión enfocadas al marco teórico indicado. Los términos iniciales para la búsqueda son los siguientes: *forensia, ontologías, correo electrónico y cabecera del correo electrónico*. De la conjunción de estos términos se puede deducir los criterios de búsqueda:

- CB1: "*ontology AND forensic AND electronic mail*": que permite ahondar en la aplicación de ontologías a la forensia digital de correos electrónicos;
- CB2: "*forensic AND email header*": para identificar los trabajos relacionados a forensia de correos electrónicos en los que se aborden los métodos y herramientas utilizadas a partir del análisis forense de la cabecera del correo electrónico,

La literatura sobre revisión bibliográfica aconseja realizar una prueba piloto de los criterios de búsqueda seleccionados, de modo de afinarlos y adecuarlos a conveniencia. El objetivo de la prueba piloto es tener pocos falsos positivos (artículos que han sido seleccionados por la búsqueda automática pero que realmente no responden a los objetivos del estudio) y pocos falsos negativos (artículos no detectados por la estrategia de búsqueda establecida pero que son de provecho para el estudio). Por ejemplo: la palabra *email* no es una palabra clave exitosa por sí misma ya que las búsquedas automáticas devuelven textos que contienen *email*

como referencia del correo electrónico de los autores, y no como tema de estudio del artículo. Así, se decidió agregar las palabras *forensic* u *ontology* para orientar la indagación.

Por otra parte, se observó que el buscador debía ajustarse particularmente en lo siguiente:

- la publicación puede contener la palabra clave *ontology* o su plural (*ontologies*),
- el término *forensic* puede figurar bajo un sinónimo (*investigation*),
- la palabra clave *electronic mail* puede resumirse como *email*, *mail o e-mail*.

Atendiendo a estas consideraciones se definieron los criterios de búsqueda señalados en la Tabla I-1.

Tabla I-1: Criterios de Búsqueda Automática

Criterio	Regla de Búsqueda
N° 1	ontology AND forensic AND email AND year>=2014 AND year<=2018
N° 2	forensic AND header email AND year>=2014 AND year<=2018

Se aconseja que la *profundidad temporal* a considerar en un estudio de tipo bibliométrico abarque un período de 5 a 15 años, pero en el presente caso se debe considerar como condicionante el avance continuo y sin pausa que ha tenido la Forensia Digital durante los últimos 5 años debido al progreso de las tecnologías de la información. Por ello, se propuso incluir en esta revisión los trabajos de investigación producidos en el período 2014-2018, ya que es de interés identificar el estado actual de las temáticas más que su evolución en el tiempo.

Se recurrió a las siguientes *fuentes de información*: revistas científicas especializadas y artículos publicados en congresos sobre las temáticas de estudio. La selección de revistas científicas especializadas no es un tema menor, debe cuidarse los aspectos de reconocimiento de la publicación en el contexto científico y factor de impacto o medida de la importancia de la revista. Las actas de congresos son de utilidad cuando se trabaja en áreas de investigación emergentes, cual es el caso de la Forensia Digital en general, y de la Forensia de Correos Electrónicos en particular. Así, en el estudio realizado se consideraron las siguientes bibliotecas electrónicas: IEEE Xplore Digital Library, ScienceDirect, Scopus, Scholar Google, The Journal of Digital Forensics, Security and Law (JDFSL), y ACM Library.

En base a estas consideraciones, y con el objetivo de fijar los *límites de la revisión*, se definió como restricciones iniciales de la búsqueda las siguientes:

- R1: Se incluyen trabajos referidos a la aplicación de las tecnologías semánticas para definir o mejorar metodologías de trabajo, herramientas forenses y análisis forense de artefactos forenses.
- R2: se considera un espacio temporal de 5 años, tomando el período 2014-2018[71].

Los criterios de exclusión definidos son los siguientes:

- CE1: Se excluyen del estudio los libros, capítulos de libros, cartas, notas, tesis de grado o posgrado y patentes.
- CE2: Se excluyen publicaciones impresas en papel, considerando solo textos electrónicas, y de éstos, aquellos que cuentan con acceso público o a los que se puede acceder mediante las vías institucionales disponibles.
- CE3: Se excluyen las publicaciones escritas en otros idiomas que no sean en inglés.
- CE4: En caso de que un mismo estudio se repitan en dos o más búsquedas, se lo considera una única vez.
- CE5: Se excluyen los trabajos de los que no puede acceder al texto completo del mismo.
- CE6: Se excluyen artículos que no estén en formato de texto portátil (PDF) y que superen los 15Mb de tamaño.

I.2.2 Procesos de Búsqueda y Selección

El proceso de búsqueda y selección se dividió en tres fases claramente identificables: Búsqueda Inicial, Preselección por Conteo de Palabras Claves y Selección Final.

La fase de ***Búsqueda Inicial*** consistió en el uso de los buscadores automáticos de las bibliotecas digitales visitadas, activando las palabras claves definidas en la Tabla I-1 como parámetros de las consultas.

Es importante destacar que en el caso de Scholar Google el criterio de búsqueda genera una cantidad masiva de artículos obtenidos, por ellos se decidió tomar los 100

[71] Debe considerarse que existen trabajos del año 2018 que no estaban publicados al momento de realizar esta revisión, como por ejemplo el trabajo LAZIĆ, L., & BOGDANOSKI, M. E-MAIL FORENSICS: TECHNIQUES AND TOOLS FOR FORENSIC INVESTIGATION. *Univerzitet Metropolitan Beograd 20. oktobar 2018. godine*, 25, que si bien se presentó en un evento realizado en el año 2018, la publicación del mismo fue realizada en el 2019 (Ver: http://bisec.rs/files/2018.pdf#page=27)

primeros únicamente, entendiendo que el resto no es de interés para la investigación aprovechando el rankeo por relevancia que Scholar Google define para su algoritmo de búsqueda.

Una vez realizadas las búsquedas automáticas de nivel inicial, o sea, aquellas generadas por los buscadores de las propias bibliotecas parametrizados según los criterios señalados en la Tabla I-1, se obtuvieron 1091 trabajos.

La fase dos, denominada ***Preselección por Conteo de Palabras Claves***, toma como insumo los 1091 textos encontrados en la fase de Búsqueda Inicial, y se realizó la preselección por conteo de palabras claves, aplicando el algoritmo que se muestra en la Figura I-1.

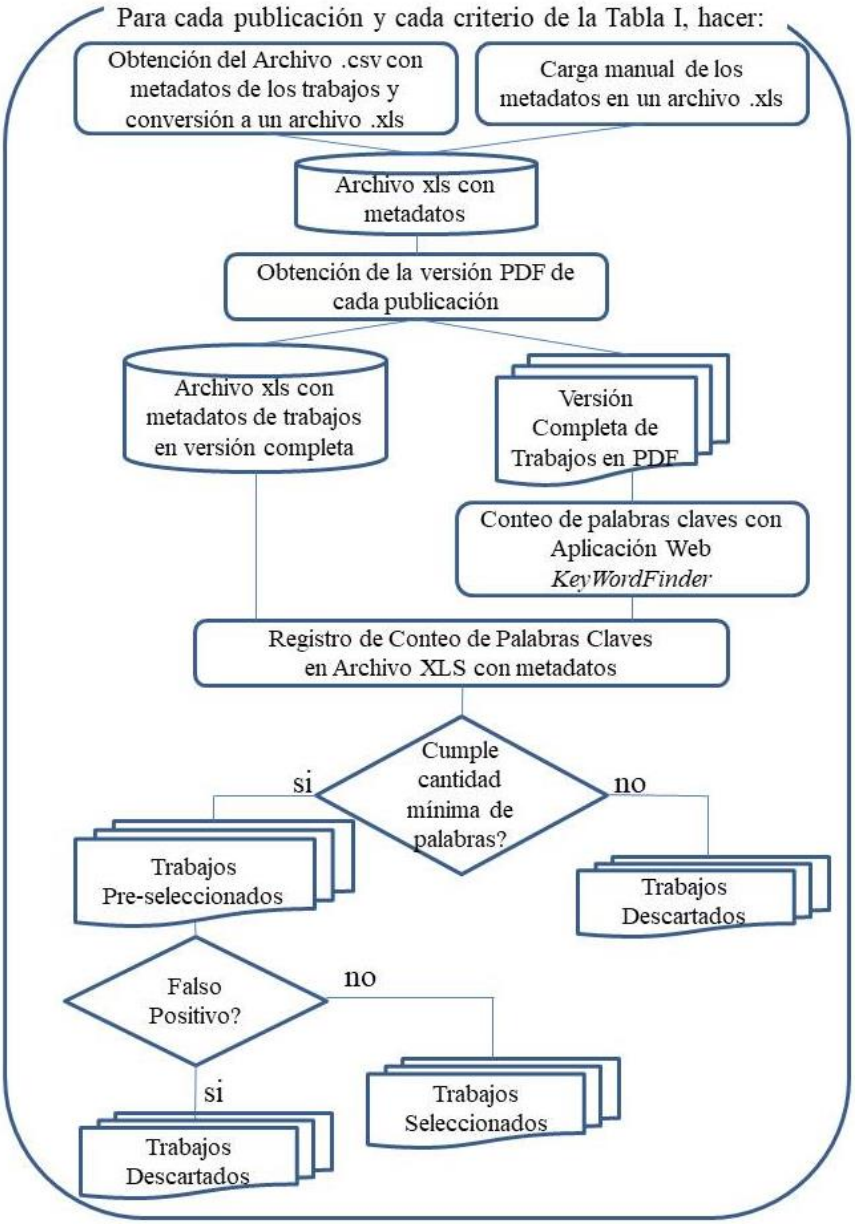


Figura I-1: Algoritmo de Preselección por Conteo de Palabras Claves

A continuación se explica el algoritmo de conteo de palabras claves, diseñado expresamente para este caso.

El procedimiento -de carácter semiautomático- se basa en un proceso ETL (Extraction, Transformation and Load), en el cual se uniformiza los metadatos de las publicaciones, respetando el formato que ofrece cada biblioteca, y contemplando una carga manual cuando la misma no permite exportar la búsqueda.

Primeramente se debe conformar una hoja de cálculo con los metadas de las publicaciones (Título, Autores, Palabras Claves, Año de Publicación, etc.), que -de acuerdo a las posibilidades de los buscadores de las bibliotecas consultadas- se pueden exportar en formato CSV desde la página de resultados obtenidos y de allí a una hoja de cálculo, o bien los metadatos se cargan manualmente en una hoja de cálculo.

La siguiente tarea consiste en acceder a los archivos PDF de las publicaciones seleccionadas y contar la cantidad de palabras claves que figuran en el texto completo del trabajo.

Para esta actividad se desarrolló una aplicación web denominada *KeyWordFinder* (disponible en https://digilab.ucasal.edu.ar/keywordfinder) que permite la carga de un texto en formato portátil (PDF) y realiza un proceso de barrido del texto contando el número total de veces que aparece una palabra clave previamente ingresada.

Se considera que un trabajo aborda la temática de estudio cuando la palabra clave figura un mínimo de tres veces en el texto.

Las palabras claves que se cuentan son las mismas que las propuestas como criterios de búsqueda de la Tabla I-1 y, además, deben figurar todas las palabras del criterio en una cantidad de 3 o más.

En particular, la palabra *mail* y sus sinónimos (*email, e-mail, electronic mail*), pueden figurar como dato de identificación de los autores, por ello, se incluye aquellos trabajos que superan en 5 el conteo de esa palabra.

Por último, se aplica una función lógica para indicar la condición de pre-seleccionado/descartado de cada trabajo, registrando esta situación en la hoja de cálculo de metadatos.

La revisión de los textos en formato portátil se realiza para los dos criterios de búsqueda de la Tabla I-1, considerando todas las bibliotecas señaladas, obteniéndose así un total de 88 publicaciones preseleccionadas.

El último paso del procedimiento, denominado ***Selección Final***, consiste en leer los trabajos preseleccionados en la fase anterior para confirmar si efectivamente resultan de interés para el estudio.

De este modo, se analiza cada trabajo teniendo presente su identificación con los objetivos de la revisión, se revisa el cumplimiento de los criterios de exclusión definidos y se confirma la condición final de *seleccionado/descartado*, incluyendo la identificación de 12 trabajos como falsos positivos.

Por último, en esta fase de revisión exhaustiva se aprovecha para identificar el área temática de cada trabajo, su objeto de estudio y separar los que tratan sobre correos electrónicos.

Así, de los 1091 trabajos hallados en la búsqueda inicial, se pre-seleccionaron 88 aplicando el algoritmo de conteo de palabras claves, y de éstos últimos, se seleccionaron 76 a partir de la lectura individual de los textos.

El resultado final de la selección, se indica en Tabla I-2 que detalla –por fuente bibliográfica- la cantidad de trabajos encontrados en cada una de las 3 fases de selección.

La tabla se completa con la referencia bibliográfica de cada uno de las publicaciones seleccionadas.

Tabla I-2: Resultados del procedimiento de búsqueda

Publicación	Criterios	Búsqueda Inicial	Preselección por Conteo	Selección Final	Referencias de los trabajos seleccionados
IEEE Xplore Digital Library	N° 1	56	4	4	(A. Gupta, Dasgupta, & Bagchi, 2017) (Amato et al., 2018), (Carvalho et al., 2016), (Mavroeidis, 2018),
	N° 2	185	25	24	(Amro, Almuhammadi, & Zhioua, 2017), (Choi, Lee, Choi, Kim, & Kim, 2016), (Duman, Kalkan-Cakmakci, Egele, Robertson, & Kirda, 2016), (Eshete & Venkatakrishnan, 2017), (Ghasem, Frommholz, & Maple, 2015a), (H. Pieterse, Olivier, & Van Heerden, 2016), (Heloise Pieterse, Olivier, & Van Heerden, 2015), (Fleurbaaij, Scanlon, & Le-Khac, 2017), (Iyer, Atrey, Varshney, & Misra, 2017), (Jayan & Dija, 2015), (Jo, Kim, & Choi, 2015),, (Kaur & Singh, 2014), (L. Chen & Mao, 2017), (Msongaleli, 2018), (Patil & Meshram, 2018), (Ramisch & Rieger, 2015), (Riaz & Tahir, 2018), (Rudd, Harang, & Saxe, 2018), (S. Gupta, Pilli, Mishra, Pundir, & Joshi, 2014), (Samuel, Graham, & Hinds, 2018), (Subedi, Budhathoki, & Dasgupta, 2018), (Tanwar & Poonia, 2014) (Y.-Y. Teing, Ali, Choo, Abdullah, & Muda, 2017), (Z. Chen et al., 2017),
Science Direct	N° 1	104	2	2	(Casey et al., 2015) (Settanni et al., 2017),

Publicación	Criterios	Búsqueda Inicial	Preselección por Conteo	Selección Final	Referencias de los trabajos seleccionados
	N° 2	104	21	18	(Aleroud & Zhou, 2017), (Bhardwaj & Goundar, 2017), (Bjelland, Franke, & Årnes, 2014), (Casey, 2018), (Clarke, Li, & Furnell, 2017), (Gray & Debreceny, 2014), (Iqbal et al., 2016), (Jahanirad, Wahab, & Anuar, 2016), (Nikkel, 2017), (Park, 2018), (Scanlon, Farina, & Kechadi, 2015), (Schmid, Iqbal, & Fung, 2015), (Senthivel, Ahmed, & Roussev, 2017), (Ullah et al., 2018), (Y. Y. Teing, Dehghantanha, Choo, & Yang, 2017), (Koven et al., 2016), (S. Yu, 2015) (X. Zhang, Baggili, & Breitinger, 2017),
Scopus	N° 1	120	2	2	(Karie & Venter, 2014) (Mehta, 2017),
	N° 2	5	1	1	(Mazurczyk & Caviglione, 2015)
Scholar Google	N° 1	84	8	5	(Carvalho & Carvalho, n.d.), (Ellison, Venter, & Adeyemi, 2017), (Jusas, Birvinskas, & Gahramanov, 2017) (Kalemi & Yildirim-Yayilgan, 2016), (Maake, Kebande, & Karie, 2017),
	N° 2	18	8	8	(Brown, 2015), (Ghasem, Frommholz, & Maple, 2015b), (Nurse, Erola, Goldsmith, & Creese, 2015) (Pluskal et al., n.d.), (Romaios et al., 2016), (Shashidhar & Manjaiah, 2014), (Umar et al., 2018), (Yang, Dehghantanha, Choo, & Muda, 2016),
The Journal of Digital Forensics, Security and Law (JDFSL)	N° 1	62	3	3	(Alzaabi, Martin, Taha, & Jones, 2015) (Mohammed, Clarke, & Li, 2017), (Wimmer et al., 2018),
	N° 2	78	8	5	(Breitinger & Baggili, 2017), (Devendran et al., 2015), (Jeong, Kang, & Lee, 2017) (Khan, Mizan, Hasan, & Sprague, 2017), (Wu et al., 2018),
ACM Library	N° 1	77	4	2	(Akremi, Sallay, Rouached, Bouaziz, & Abid, 2015), (Turner, 2017)
	N° 2	198	2	2	(Matic, Kotzias, & Caballero, 2015), (Varma, Walls, Lynn, & Levine, 2014)
		1091	88	76	Total: 76 trabajos

I.3 ANÁLISIS Y DISCUSIÓN DE LOS RESULTADOS

A nivel descriptivo, se pueden analizar los trabajos seleccionados en su totalidad, para luego realizar un análisis particular de cada investigación.

Si se tiene en cuenta el *año de publicación*, se observa que los trabajos se distribuyen de manera pareja en el quinquenio considerado (ver Figura I-2), y no se encuentra un incremento de investigaciones de un año a otro como sería de esperarse a fin conformar un marco científico que sustente la Forensia Digital.

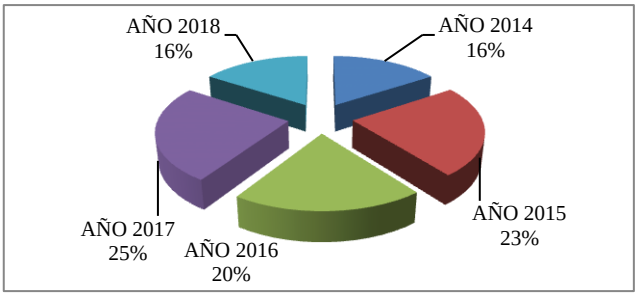


Figura I-2: Distribución de Trabajos por Año de Publicación

Respecto de la *temática principal* abordada en cada publicación, se observó que los trabajos se enfocan hacia Tráfico de Redes, Ontologías, Data Mining, Big Data, Seguridad Informática y otras áreas diversas como lenguaje natural y máquinas virtuales entre otros. La Figura I-3 muestra la distribución porcentual de las publicaciones por área temática.

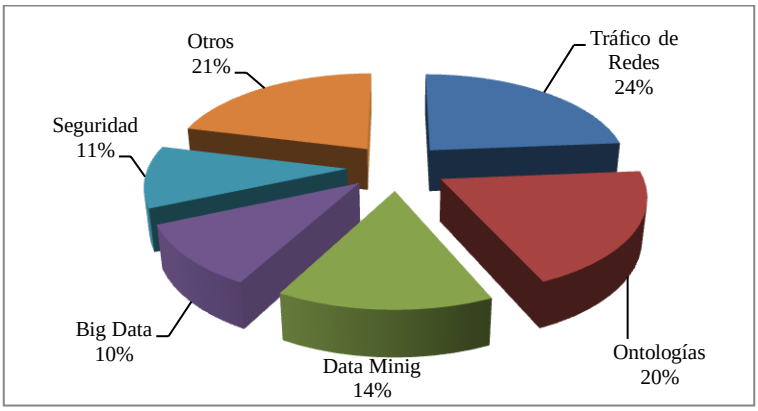


Figura I-3: Distribución de Trabajos por Área Temática

También se pueden clasificar los trabajos en función de su aporte a la mejora o estudio de los *métodos de análisis forense, las herramientas y los artefactos forenses*. En ese sentido se detectó que el 45% de los estudios tratan sobre métodos para el análisis forense, 38% abordan el tema de herramientas utilizadas para dicho análisis y el 17% restante trata sobre forensia en dispositivos o artefactos forenses (Figura I-4).

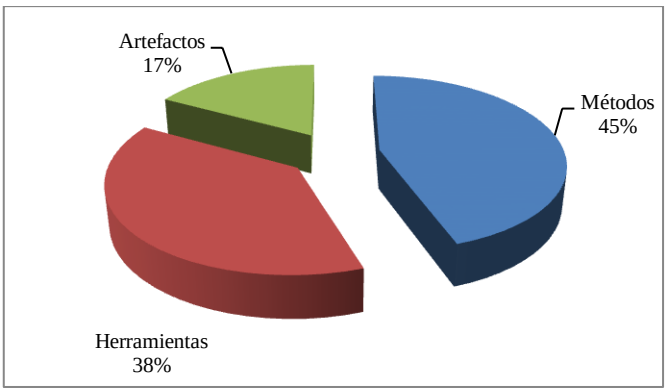


Figura I-4: Distribución de Publicaciones por Objeto de Estudio

En particular, sobre *correos electrónicos*, se encontró que 27 trabajos estudian y describen el análisis forense de correos electrónicos, y de éstos, cinco analizan ataques cibernéticos utilizando correos electrónicos. En cinco se aborda el estudio de la cabecera del correo electrónico y de éstos sólo en uno se estudia expresamente el proceso de transmisión.

En función del resumen enunciado, se pueden describir los aportes más importantes de cada trabajo, en base a tres grupos:

- *Análisis por área temática*: considerando las investigaciones que abordan el Tráfico de Redes; Ontologías; técnicas y herramientas de Data Mining; o de Big Data; Seguridad Informática y Otras Áreas de la informática.
- *Análisis por objeto de estudio*: identificando los estudios sobre Métodos de Análisis Forense; Herramientas de Análisis Forense y sobre Forensia de Dispositivos y/o Artefactos.
- *Análisis de trabajos referidos a Forensia de Correos Electrónicos*, detallando los estudios sobre ataques cibernéticos vía correo electrónico, aquellos que abordan la cabecera del correo electrónico y los que consideran el proceso de transmisión. Seguidamente se detallan los trabajos incluidos en cada apartado.

I.3.1 Análisis por Área Temática

En lo referente a *Tráfico de Redes* se destacan estudios que se pueden agrupar según dos enfoques: ataques cibernéticos y forensia de redes de datos.

Se encontraron estudios referidos a *ataques cibernéticos* en las siguientes investigaciones: en (Choi et al., 2016) los autores proponen TRIG un sistema de almacenamiento de tráfico y generación de información relacionada, que puede almacenar el tráfico de red de 20 Gbps en tiempo real; en (Settanni et al., 2017) se presenta un enfoque de colaboración para la gestión de la información de incidentes cibernéticos en procesos industriales; en (Breitinger & Baggili, 2017) se abordaron los sistemas de prevención de fuga de datos (DLPS) que analizan el tráfico de red y alertan en caso de una fuga de datos. El trabajo (Shashidhar & Manjaiah, 2014) analiza los ataques cibernéticos con ayuda de herramientas forenses digitales como *WinHex* mientras que el trabajo (Patil & Meshram, 2018) describe una metodología

de análisis de paquetes de red para detectar espionaje industrial a partir del tráfico de red de una organización.

En referencia a la *Forensia de Redes de Datos*, la investigación (Pluskal et al., n.d.) presenta NetFox, una herramienta para el análisis forense de redes con posibilidades avanzadas de reconstrucción de aplicaciones y técnicas avanzadas para detectar actividades ilegales o no autorizadas; por su parte, el trabajo (Senthivel et al., 2017) trata sobre el análisis forense completo del tráfico de red en sistemas SCADA; en (Scanlon et al., 2015) se estudia la herramienta de intercambio de archivos P2P BitTorrent versión 2.x y propone métodos para recuperar potenciales evidencias forense en dispositivos que ejecutan Windows, Mac OS, Ubuntu, iOS y dispositivos Android.

Considerando ahora la aplicación de *Ontologías*, se encontraron trabajos referidos a dos enfoques: trabajos que presentan ontologías expresamente desarrolladas y aplicación de las tecnologías semánticas como marcos de trabajo.

Los trabajos que presentan *desarrollo de ontologías* son los siguientes estudios: la herramienta CyBox propuesta en el estudio (Casey et al., 2015) basada en la ontología DFAX permite representar e intercambiar información forense digital, incorporando aspectos procesales de la forensia, como ser la cadena de custodia, manejo de casos y procesamiento forense, mientras que en (Karie & Venter, 2014) se propone DF Disciplines una ontología para categorizar las disciplinas forenses digitales, así como algunas metodologías que puedan ofrecer orientación en diferentes áreas de la forense digital. Mediante la ontología OBMO (Online Banking Malware Ontology) que se aborda en (Carvalho et al., 2016), que a su vez se basa en ontologías prediseñadas OSCAF (Open Semantic Collaboration Architecture Foundation) se define un método para modelar las organizaciones criminales intervinientes e identificar a los desarrolladores de malware; esta misma ontología OBMO se trabaja en (Carvalho & Carvalho, n.d.) para la investigación de malware bancario en línea con nuevos enfoques que correlacionen la evidencia digital. En (Wimmer et al., 2018) se describe una ontología que cataloga herramientas forenses digitales comunes mientras que el estudio (Kalemi & Yildirim-Yayilgan, 2016) presenta una revisión exhaustiva de las soluciones y modelos existentes para recopilar información y utilizarla para resolver crímenes mediante una ontología prototipo llamada SC-Ont mientras que el estudio (Akremi et al., 2015) describe una ontología para servicios web (Fi4SOA) propuesta para fusionar propiedades forenses

con requisitos comerciales en la fase de diseño del servicio. El estudio (Mavroeidis, 2018) presenta PSO como marco ontológico para mejorar la seguridad física y la detección de amenazas internas; en este mismo espacio (Ellison et al., 2017) presenta INFORENSIC Ver2 para la gestión del conocimiento en seguridad y análisis forense digital. También se analizó el trabajo (Alzaabi et al., 2015) que estudia los teléfonos celulares, proponiendo la ontología F-DOC para modelar cada componente del contenido de un teléfono inteligente con el propósito de realizar el análisis forense.

Otros trabajos abordan la *aplicación de las tecnologías semánticas* en diferentes contextos: en (Amato et al., 2018) se describe un marco ontológico que recupera y muestra modelos de evidencias obtenidos a través de diferentes herramientas forenses, presentando un sistema capaz de añadir afirmación semántica a los datos generados por las herramientas de análisis forense durante los procesos de extracción. Los autores del trabajo (Iqbal et al., 2016) proponen una taxonomía de ataques de seguridad en la nube y las posibles estrategias de mitigación; mientras que en (Maake et al., 2017) se describe un marco para integrar la Ingeniería de Requisitos (RE) con Ontologías Forenses Digitales (SDFO) científicamente optimizadas.

En la mayoría de estos trabajos no se indica la metodología utilizada para construir la ontología, salvo (Mavroeidis, 2018) y (Amato et al., 2018) que presentan una metodología propia; (Ellison et al., 2017) y (Alzaabi et al., 2015) que utilizan Methontology con algunas variantes; en el trabajo (Akremi et al., 2015) se utiliza la metodología Sherwood Applied Business Security (SABSA) y en el estudio (Carvalho & Carvalho, n.d.) se recurre a NeOn.

Muy pocos artículos ((Mavroeidis, 2018), (Amato et al., 2018), (Carvalho & Carvalho, n.d.), (Wimmer et al., 2018), (Kalemi & Yildirim-Yayilgan, 2016)) describen la utilización de herramientas y lenguajes utilizados en la propuesta. Entre los lenguajes y herramientas utilizados se encuentran: RDF, RDFS, SPARQL, OWL, OWL 2, SWLR y SPARQL Endpoint.

Solo describen gráficamente la taxonomía definida en (Carvalho et al., 2016), (Iqbal et al., 2016), (Carvalho & Carvalho, n.d.), (Wimmer et al., 2018), (Kalemi & Yildirim-Yayilgan, 2016), (Alzaabi et al., 2015) y (Karie & Venter, 2014); mientras que (Mavroeidis, 2018) y (Carvalho et al., 2016) incluyen casos de uso y en (Ellison et al., 2017), (Carvalho et al., 2016), (Casey et al., 2015) se menciona la reutilización de otras ontologías.

Por otra parte, son varias las investigaciones abordadas desde *Data Mining*. Aplicando aprendizaje automático se encontraron varios trabajos: en (Ghasem et al., 2015a) se propone un enfoque híbrido para detectar, filtrar y archivar pruebas provenientes de correos electrónicos; en tanto la herramienta DYNAMINER es introducida en (Eshete & Venkatakrishnan, 2017) para la identificación de malware como un problema de aprendizaje basado en análisis de tráfico de red. También se encontraron trabajos relacionados a redes neuronales para la identificación de Malware en el estudio (Rudd et al., 2018).

Las herramientas relacionadas a la *extracción de datos* son utilizadas en (Jahanirad et al., 2016) para la definición de mejores atributos de una cámara de video, mediante la aplicación de técnicas de procesamiento de imágenes y extracción de datos y así aumentar la precisión de detección, la solidez y la eficiencia computacional del dispositivo de video. El estudio detallado en (Fleurbaaij et al., 2017) describe el manejo de datos privilegiados mediante técnicas de extracción de datos, basadas en un script dentro de la herramienta forense digital Nuix. Por su parte, en (Gray & Debreceny, 2014) se estudia la aplicación de técnicas de extracción de datos para la detección de fraudes en los estados financieros de auditoría y propone una taxonomía para apoyar y guiar la búsqueda de datos, considerando la inclusión en el análisis de correos electrónicos como evidencia digital, utilizando técnicas de minería de textos.

Se encontraron aplicaciones de la minería de datos tanto dirigidos a la Forensia como a la Seguridad Informática (Khan et al., 2017) utiliza técnicas de *clustering* para el agrupamiento de correos electrónicos no deseados, atendiendo a diferentes atributos que permiten acotar los tiempos de búsqueda en el repositorio de SPAM. En tanto que en (Kaur & Singh, 2014) se estudian la capacidad de propagación de gusanos que amenazan cada vez más a los hosts y servicios de Internet, en donde se recurre a la minería de datos para explotar vulnerabilidades desconocidas y además estudiar sus propias representaciones cambiantes. Finalmente, en (Ullah et al., 2018) se identifican y analizan los vectores de ataque de exfiltración de datos (fuga de datos confidenciales o privados a una entidad no autorizada) y las contramedidas vigentes.

Por otra parte, se detectaron algunos trabajos que estudian analíticas sobre grandes volúmenes de datos o *Big Data*. En (Mohammed et al., 2017) los autores describen como realizar el análisis forense digital de un repositorio de Big Data con

datos heterogéneos provenientes de diversas fuentes. En tanto, el estudio (Bjelland et al., 2014) muestra cómo aplicar Fuzzy Hashing en investigaciones forenses y así identificar datos complejos y no estructurados que tienen cierta similitud de nivel de bytes.

También se describen herramientas analíticas para Big Data. En (Varma et al., 2014) se presenta LIFTR que permite priorizar información selectiva recuperada de los teléfonos Android. El estudio (A. Gupta et al., 2017) muestra PROFORMA, un sistema prototipo que evalúa continuamente la confiabilidad y el riesgo de las comunicaciones sociales, en el cual el usuario otorga permiso explícito para acceder a las redes sociales. Por su parte, (Msongaleli, 2018) estudia los remanentes de datos de valor forense del servicio de almacenamiento en la nube privada de Syncany, un motor de almacenamiento popular para plataformas de big data, orientando el estudio a la reducción de tiempos de búsqueda de la evidencia. En (Amro et al., 2017) se propone una herramienta de Big Data para analizar el tráfico de Internet y extraer información de alto nivel, como enlaces visitados, credenciales de usuario y cookies de sesión de los protocolos de red utilizados.

En la revisión realizada se hallaron trabajos relacionados a la *seguridad informática*. Entre estos trabajos se destacan la investigación descripta en el trabajo (Schmid et al., 2015) donde se identifican las lagunas legales y las tecnologías que facilitan la comisión de actos delictivos cibernéticos; así como la propuesta (X. Zhang et al., 2017) que describe las técnicas de ingeniería inversa que pueden utilizarse para acceder a datos encriptados.

Por su parte, en el estudio (Jusas et al., 2017) se muestra una clasificación de la evidencia digital en función del momento de obtención de las pruebas (análisis en vivo o post mortem), análisis de dispositivos móviles y herramientas de análisis forense.

Por otra parte, también se abordaron *técnicas de ataque y mitigación de amenazas*. En tal sentido el trabajo (Mazurczyk & Caviglione, 2015) presenta un análisis para detectar y mitigar los ataques o amenazas estenográficas en teléfonos celulares mientras que el estudio señalado en (Subedi et al., 2018) aborda técnicas de análisis correlacional para identificar familias de ransomware.

Entre *otras áreas temáticas* identificadas se puede mencionar estos trabajos: forensia de máquinas virtuales ((Riaz & Tahir, 2018) y (Park, 2018)); análisis forense de archivos en memoria o discos ((Romaios et al., 2016), (Iyer et al., 2017),

(Z. Chen et al., 2017), (L. Chen & Mao, 2017), (Y. Y. Teing et al., 2017), (Samuel et al., 2018)), y aplicación de técnicas de procesamiento de lenguaje natural. Relacionado con esta última temática, el artículo (Turner, 2017) propone una herramienta para la traducción automática de políticas de seguridad escritas en lenguaje natural a lenguaje de máquina, mediante el lenguaje ABAC (Attribute Based Access Control).

El análisis de la revisión según la clasificación por áreas temáticas se completa con aquellos trabajos referidos a la aplicación de dichas áreas en la forensia de correos electrónicos, los cuales se detallan en la sección II.3.3.

I.3.2 Análisis por Objeto de Estudio

Si se considera el elemento central del estudio, se pueden identificar trabajos que desarrollan *métodos o metodologías específicas* para determinadas actividades. Así, los estudios se pueden agrupar en métodos para:

- gestionar incidentes cibernéticos ((Settanni et al., 2017), (Breitinger & Baggili, 2017), (Brown, 2015), (Tanwar & Poonia, 2014), (Mavroeidis, 2018), (Iqbal et al., 2016), (Ghasem et al., 2015b), (Kaur & Singh, 2014)),
- analizar el tráfico de red ((Clarke et al., 2017), (Msongaleli, 2018), (Patil & Meshram, 2018), (Yang et al., 2016)),
- detectar datos falsificados ((Jayan & Dija, 2015), (Duman et al., 2016), (S. Yu, 2015), (S. Gupta et al., 2014)),
- identificar virus ((Samuel et al., 2018), (Rudd et al., 2018), (Khan et al., 2017)),
- así como identificar y procesar datos complejos y no estructurados ((Bjelland et al., 2014), (Casey, 2018), (Amato et al., 2018), (Mohammed et al., 2017), (Jusas et al., 2017), (Park, 2018), (Jo et al., 2015), (Jeong et al., 2017)).

También se encontraron algunos métodos innovadores o de atención para el análisis forense, como por ejemplo: en (Shashidhar & Manjaiah, 2014) se detallan los procesos de investigación sobre fuentes de ataques cibernéticos y en (Fleurbaaij et al., 2017) se proponen métodos para privilegiar datos forenses y minimizar la exposición de los contenidos al investigador forense.

Se analizaron 5 trabajos que describen *herramientas* orientadas al análisis de trafico de redes (CARONTE (Matic et al., 2015), NETFOX (Pluskal et al., n.d.),

TRIG (Choi et al., 2016), RDAP (Nikkel, 2017), RSLogix (Senthivel et al., 2017)) y una herramienta orientada al seguimiento de incidentes de seguridad (DYNAMINER (Eshete & Venkatakrishnan, 2017)).

Asimismo, en esta revisión se identificaron un conjunto de herramientas basadas en las tecnologías semánticas: INFORENSIC Ver 2 (Ellison et al., 2017), OBMO (Carvalho et al., 2016) y (Carvalho & Carvalho, n.d.), Fi4SOA (Akremi et al., 2015), CyBox (Casey et al., 2015), OTM (Maake et al., 2017), SC-Ont (Kalemi & Yildirim-Yayilgan, 2016), PROFORMA (A. Gupta et al., 2017), F-DOC (Alzaabi et al., 2015) y DF (Karie & Venter, 2014). También se encontraron herramientas basadas en procesamiento de lenguaje natural ((Turner, 2017)), en técnicas de minería de datos ((Gray & Debreceny, 2014), (Jahanirad et al., 2016), (Schmid et al., 2015)) y en técnicas de análisis de grandes volúmenes de datos ((Ullah et al., 2018), (Subedi et al., 2018), (Amro et al., 2017), (Y.-Y. Teing et al., 2017)).

En cuanto a los trabajos que abordan el análisis forense de *artefactos*, se pueden citar los siguientes: estudios para análisis de dispositivos móviles ((H. Pieterse et al., 2016), (X. Zhang et al., 2017), (Varma et al., 2014), (Mazurczyk & Caviglione, 2015)), para análisis de computadoras ((Y.-Y. Teing et al., 2017), (Kalemi & Yildirim-Yayilgan, 2016), (Z. Chen et al., 2017), (Samuel et al., 2018)) y para análisis de redes ((Scanlon et al., 2015), (Patil & Meshram, 2018), (Bhardwaj & Goundar, 2017), (Wu et al., 2018)) considerando diversos sistemas operativos y software de base; estudios sobre análisis de memorias ((Romaios et al., 2016) y (Iyer et al., 2017)) y máquinas virtuales ((Riaz & Tahir, 2018); sobre análisis de imágenes ((Wu et al., 2018)) y mensajería instantánea en redes sociales ((Yang et al., 2016)).

En el caso particular de los estudios que tratan sobre métodos, herramientas y artefactos que involucran correos electrónicos, se detallan en la sección siguiente.

I.3.3 Análisis de Estudios Sobre Correos Electrónicos

En este tercer criterio de clasificación se detallan los trabajos que abordan el análisis de correos electrónicos, detallando particularmente como se encararon desde las áreas temáticas señaladas en la clasificación de la sección I.3.1, y como se identificaron según el objeto de estudio según la clasificación de la sección I.3.2.

Así, se pueden considerar los trabajos relacionados a *tráfico de redes*, en donde particularmente interesan aquellos estudios centrados en el *análisis de la dirección IP*. En (Clarke et al., 2017) se presenta un algoritmo para reducir el volumen de direcciones IP a analizar considerando la identificación e interacción de los usuarios mediante el análisis de metadatos del tráfico de red. El trabajo (Matic et al., 2015) propone un método para recuperar la dirección IP oculta en servicios de TOR (The Onion Router). En tanto, el estudio (Nikkel, 2017) desarrollaron la herramienta RDAP que resuelve consultas sobre registro de direcciones IP, nombres de dominio, sistemas autónomos, características de seguridad e internacionalización para una dirección IP determinada.

Es de destacar que solo se encontró un trabajo que recurre a las *tecnologías semánticas* aplicadas a correos electrónicos. El estudio (Mehta, 2017) describe una ontología que representa el direccionamiento semántico del correo electrónico, que permite a los usuarios dirigir correos electrónicos a grupos especificados semánticamente, proporcionando autenticación segura a grupos de cuentas de correo.

Desde la *minería de datos* se realizaron varias investigaciones sobre correos electrónicos. En (Khan et al., 2017) se recurre a la agrupación de correos electrónicos no deseados en función de sus diferentes atributos para conformar clústeres que permiten acortar los tiempos de búsqueda en el repositorio de correos spams. El estudio (Ghasem et al., 2015a) utiliza *aprendizaje automático* para detectar, filtrar y archivar evidencia proveniente de correos electrónicos, que permitan identificar a quienes actúan en el ciberespacio cometiendo delitos. Los autores del trabajo (Gray & Debreceny, 2014) proponen una aplicación de *técnicas de extracción de datos y minería de textos* para la detección de fraudes en los estados financieros de auditoría y propone una taxonomía para apoyar y guiar la búsqueda de datos, considerando la inclusión en el análisis de correos electrónicos como evidencia digital. Finalmente, (Ghasem et al., 2015b) apela a la minería de textos para generar un marco de detección de acciones de ciberacoso en mensajes; servicio de mensajes cortos, servicio de mensajes multimedia, chat, mensajes de instancia y correos electrónicos.

En (Koven et al., 2016) se proponen técnicas analíticas de *Big Data* para la búsqueda de evidencia en grandes conjuntos de datos de correo electrónico.

Desde la *Seguridad Informática* se encontraron varios trabajos que se pueden identificar en base al tipo de ataque cibernético planteados desde correos

electrónicos. (S. Yu, 2015) analiza cinco escenarios (S. Yu, 2015)[130](S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)(S. Yu, 2015)en los que el investigador forense tiende a pasar por alto la información incriminatoria crucial que se ha disfrazado de spam. También referido a la seguridad informática, se analizó el trabajo de (Patil & Meshram, 2018) en el cual se realiza una encuesta a los usuarios finales y se concluye que es habitual la violación de las políticas de seguridad, particularmente con el uso indebido del correo electrónico. El estudio (Duman et al., 2016) aborda el Spearphishing (variación de Phishing que ataca a organizaciones especialmente seleccionadas) con un novedoso enfoque automatizado basado en modelos probabilísticos de metadatos de correo electrónico y características estilométricas del contenido de correo electrónico. Acerca de investigaciones sobre Phishing se encontró el estudio (Aleroud & Zhou, 2017) que trata el correo electrónico como vehículo para este tipo de amenaza de seguridad, en dicho estudio se describe una encuesta que investiga los ataques de phishing y las técnicas antiphishing desarrolladas no solo en entornos tradicionales, como correos electrónicos y sitios web, sino también en entornos nuevos, como las redes sociales y teléfonos móviles.

Entre los trabajos que abordan la aplicación de *otras áreas de la informática* aplicada a la forensia de correos electrónicos, se encontró el estudio (Schmid et al., 2015) que propone técnicas de procesamiento de lenguaje natural (clasificación asociativa personalizada) para abordar el problema de atribución de autoría de correos electrónicos a partir de las características que definen el estilo de escritura de una persona.

Considerando los *métodos forenses* es de interés el estudio (Mazurczyk & Caviglione, 2015) en el que los autores definieron un método forense basado en las directrices disponibles preparadas por el Instituto Nacional de Estándares y Tecnología (NIST) para forensia de correos electrónicos. Por su parte, en (Tanwar & Poonia, 2014) se propone un algoritmo para encontrar vínculos y repeticiones de datos en un contexto de investigación forense, se evaluó la efectividad del algoritmo buscando similitud en cadenas de correos electrónicos y permitió segregar direcciones de correo similares de las no similares.

A continuación se detallan las *herramientas forenses* para el procesamiento de datos de correos electrónicos. Se identificaron las siguientes: Sistema de

Visualización de correlaciones para Foxmail (Z. Chen et al., 2017) que permite extraer la información del archivo de evidencia de correo mostrando gráficamente la asociación entre los contactos y permite la búsqueda en el cuerpo del correo así como el archivo adjunto mediante la recuperación de texto completo; EMAILFINDER (L. Chen & Mao, 2017) para acceder a información de correos electrónicos residente en la memoria de teléfonos móviles; INVEST (Koven et al., 2016) que posibilita la búsqueda de evidencia en grandes conjuntos de datos de correo electrónico; en el estudio (Ramisch & Rieger, 2015) se presenta un script en SQLite Index Recovery que se probó con datos de la aplicación Apple Mail.

Considerando el correo electrónico como *artefacto forense*, la revisión permitió identificar investigaciones de interés. El estudio (Iyer et al., 2017) refiere al comportamiento de las diferentes aplicaciones cliente de correo electrónico mientras recibe los correos electrónicos falsificados del remitente, el estudio plantea un algoritmo para la identificación de direcciones falsas mediante el análisis de determinados campos. Los autores de (Bhardwaj & Goundar, 2017) investigan acerca de las ventajas y amenazas de la infraestructura de correo electrónico basada en la nube (comúnmente conocida como web mail) y analizan las amenazas de ataque a los servidores de correo corporativos que pueden impactar en los negocios e incluso en la pérdida de reputación o el espionaje industrial. Finalmente, el estudio (Nurse et al., 2015) considera los correos electrónicos salientes y la fuga de información involuntaria considerando los metadatos que son una parte natural de los encabezados de los correos electrónicos, marcando un nivel notable de exposición de la información de identidad personal y organizativa que puede quedar a disposición de un atacante.

Particularmente, los trabajos (Jayan & Dija, 2015), (Msongaleli, 2018), (Romaios et al., 2016), (Umar et al., 2018) y (Devendran et al., 2015) abordan el análisis forense de correos electrónicos desde el encabezado del mismo. El último trabajo mencionado, compara las cinco herramientas forenses de código abierto más populares para forensia de correos electrónicos, todas ellas basadas en el análisis del encabezado. El estudio (Jayan & Dija, 2015) discute acerca de diferentes métodos para detectar la falsificación de correos electrónicos (spoofing) analizando el encabezado del correo. En (Umar et al., 2018) se define un método forense para teléfonos Android basado en las directrices preparadas por el Instituto Nacional de Estándares y Tecnología (NIST) para forensia de correos electrónicos a partir de los

metadatos de la cabecera. Por su parte, (Romaios et al., 2016) analiza varias herramientas forenses que se basan en los registros del encabezado de los correos electrónicos, con énfasis en la delincuencia en línea y las restricciones legales, examinando la amplitud de la información que se puede obtener con esas herramientas. Por su parte, en el trabajo (Msongaleli, 2018) se propone un algoritmo de tres niveles para identificar correos maliciosos a partir de los registros de servidores y dispositivos locales que se encuentran en el encabezado de los correos.

Respecto del *proceso de transmisión* de correos electrónicos, los trabajos precitados en el párrafo anterior dan por sentado que al considerar las direcciones IP del encabezado se analiza el proceso de transmisión pero no lo enfocan de manera directa y concreta, salvo en el trabajo (Msongaleli, 2018).

I.4 APORTES DE LA REVISIÓN REALIZADA

En esta sección se resumen los resultados de la revisión, desde su contribución a los objetivos planteados para la revisión bibliográfica.

Así, respecto del primer objetivo, referido a *Identificar y estudiar los aportes investigativos más actualizados sobre Ontologías y Forensia Digital*, se puede decir que el mismo se cumplió ya que se encontraron publicaciones con contribuciones destacadas para la formalización científica de la Forensia Digital, particularmente, aquellos vinculados a las tecnologías semánticas, que en un total de 13 trabajos, representan el 17% de las investigaciones consideradas, y están enfocados a 4 temáticas particulares: direccionamiento de cuentas, desarrollo de herramientas forenses, estrategias para mitigación de ataques y representación semántica de artefactos forenses.

El segundo objetivo de esta revisión es la propuesta de *establecer las áreas de vacancia sobre la aplicación de las tecnologías semánticas al análisis forense de correos electrónicos*. Al respecto, se observó que los estudios realizados se enfocaron en tres áreas temáticas: el estudio del *Trafico de Redes*, particularmente en la *identificación de direcciones IP*; en la *Minería de Datos* con énfasis en la *minería de textos y técnicas de extracción de datos*; y en la *Seguridad Informática* que se enfoca principalmente en la atención *de ataques cibernéticos*. Mientras que las áreas

con escasa aplicación en la forensia de correos electrónicos serían las siguientes: *Big Data, tecnologías semánticas y procesamiento de lenguaje natural.*

En particular, la conjunción entre tecnologías semánticas y forensia digital aplicada a correos electrónicos solo se encontró en un trabajo de investigación: el estudio (Mehta, 2017) en el que se propone el direccionamiento de correo electrónico semántico (SEA) para dirigir correos a grupos de usuarios especificados semánticamente; pero se debe destacar que este trabajo no aborda el desarrollo de una ontología que modele específicamente la cabecera del correo electrónico.

El último objetivo de la revisión, referido a *relacionar estos trabajos desde atributos de cercanía (o distancia) con la aplicación de tecnologías semánticas para el análisis forense de correos electrónicos,* puede decirse que se cumplió, ya que la revisión se desarrolló con el detalle suficiente como para encontrar dos resultados concretos: a) las tecnologías semánticas se utilizan para el desarrollo de herramientas de forensia digital (se observó esto en 10 trabajos, que representan el 13% de las publicaciones revisadas); y b) solo un trabajo trata sobre la aplicación de las tecnologías semánticas a la forensia de correos electrónicos en particular.

Particularmente se prestó atención a las escasas investigaciones (solo 5 que representa el 7% de las publicaciones revisadas), dirigidas a estudiar el análisis forense de correos electrónicos basados en los metadatos de la cabecera, y de este grupo, solo en un trabajo se abordó detalladamente el proceso de transmisión.

I.5 CONCLUSIONES

Los objetivos propuestos para la revisión se cumplieron, ya que se pudieron identificar investigaciones actualizadas sobre aplicación de las tecnologías semánticas a la Forensia Digital, se encontraron áreas de vacancia de interés para trabajar desde el espacio de la investigación académica, y se pudo relacionar las publicaciones revisadas en términos de cercanía o distancia en la aplicación de las tecnologías semánticas.

Son pocos los estudios sobre Big Data y Procesamiento de Lenguaje Natural que abordan cuestiones de la forensia digital. Y respecto de las tecnologías semánticas, aunque fue considerada por un conjunto importante de investigadores sobre Forensia Digital, se abordó escasamente en el análisis pericial de correos electrónicos. De lo

dicho, se puede concluir que son varios los ámbitos en los que sería importante generar investigaciones sobre Forensia Digital.

Por una parte, las técnicas y herramientas analíticas de Big Data son adecuadas para procesar grandes volúmenes de datos no estructurados, como los contenidos en las cuentas de correo electrónico, para realizar estudios sobre el volumen, variedad y valor de esos datos, considerando además las velocidades de procesamiento que permiten estas herramientas analíticas.

El correo electrónico cuenta con un componente textual de mucho interés para la Forensia Digital, particularmente porque la documentación impresa fue reemplazada poco a poco por los mensajes de correo electrónico, en los que también se observan características sobre el estilo de escritura, sintaxis y semántica de las palabras. Desde este enfoque, los métodos y herramientas propias del Procesamiento de Lenguaje Natural pueden aportar un marco científico adecuado.

En la búsqueda bibliográfica realizada se encontraron trabajos asociados a una o más de estas dos temáticas: *ontologías y forensia digital*, en los que se observan modelos, criterios o componentes que son útiles para la formulación de herramientas y marcos ontológicos para resolver problemáticas referidas a la forensia de correos electrónicos.

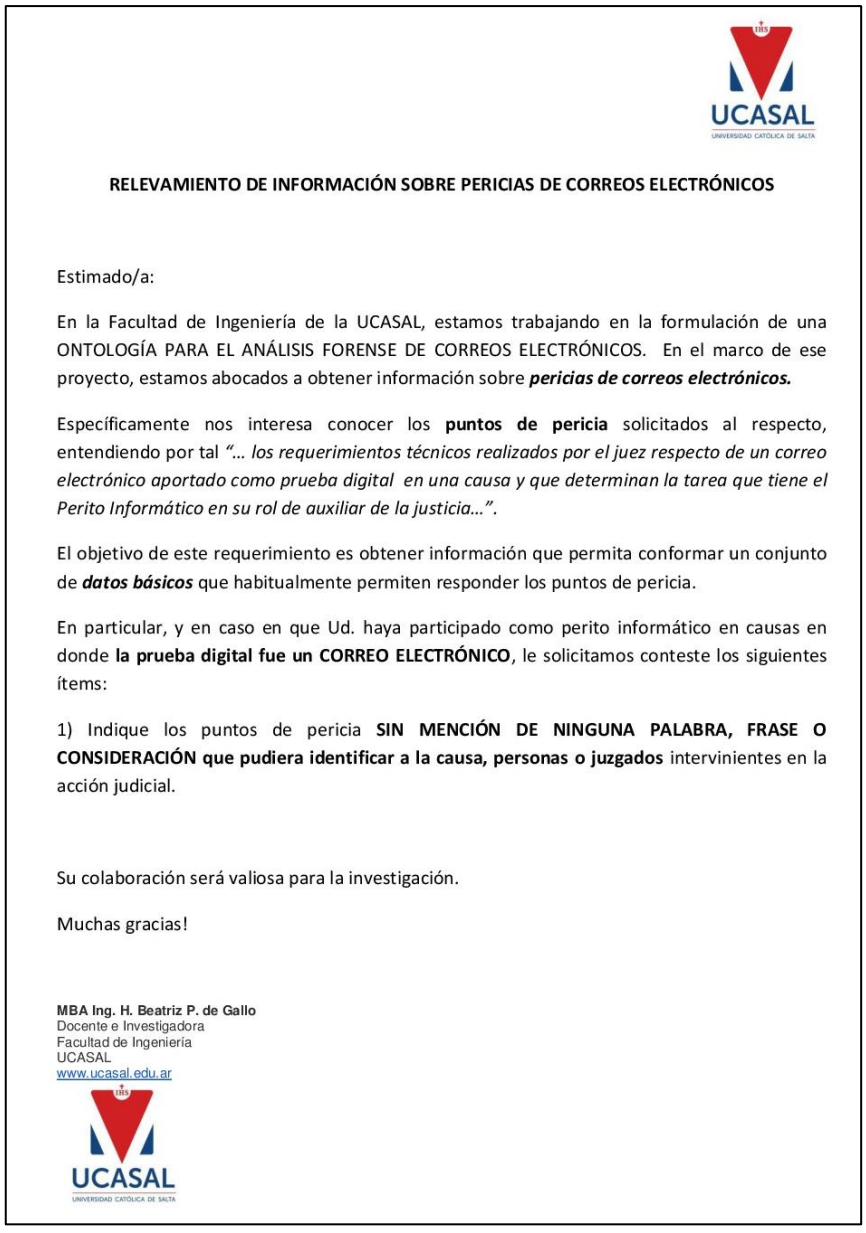
Pero también se debe destacar que no se encontraron trabajos sobre el caso particular de aplicación de la trazabilidad para validar correos electrónicos, basado en una representación ontológica, lo cual evidencia la necesidad del desarrollo de criterios científicos que certifiquen y formalicen las actividades de la forensia digital.

La ausencia de este tipo de trabajos abre las posibilidades para la investigación en esta línea. En última instancia, esta características –la comprobación de la existencia del correo electrónico- es la que permite sostener la condición de no repudio de esta evidencia digital.

ANEXO II: PUNTOS DE PERICIA

Este apartado contiene toda la información auxiliar referida a los PUNTOS DE PERICIA, incluyendo tanto el formulario de relevamiento implementado en la búsqueda de dichos requerimientos judiciales, como el análisis y tablas de resultados arribados durante el procesamiento de la información recabada. Todo ello con el objetivo de generar en última instancia, las preguntas de competencia a las que debe responder OntoFoCE.

A continuación se muestra el formulario utilizado para el RELEVAMIENTO SOBRE PERICIAS DE CORREOS ELECTRÓNICOS (Figura II-1)



RELEVAMIENTO DE INFORMACIÓN SOBRE PERICIAS DE CORREOS ELECTRÓNICOS

Estimado/a:

En la Facultad de Ingeniería de la UCASAL, estamos trabajando en la formulación de una ONTOLOGÍA PARA EL ANÁLISIS FORENSE DE CORREOS ELECTRÓNICOS. En el marco de ese proyecto, estamos abocados a obtener información sobre ***pericias de correos electrónicos.***

Específicamente nos interesa conocer los **puntos de pericia** solicitados al respecto, entendiendo por tal *"... los requerimientos técnicos realizados por el juez respecto de un correo electrónico aportado como prueba digital en una causa y que determinan la tarea que tiene el Perito Informático en su rol de auxiliar de la justicia..."*.

El objetivo de este requerimiento es obtener información que permita conformar un conjunto de ***datos básicos*** que habitualmente permiten responder los puntos de pericia.

En particular, y en caso en que Ud. haya participado como perito informático en causas en donde **la prueba digital fue un CORREO ELECTRÓNICO**, le solicitamos conteste los siguientes ítems:

1) Indique los puntos de pericia **SIN MENCIÓN DE NINGUNA PALABRA, FRASE O CONSIDERACIÓN que pudiera identificar a la causa, personas o juzgados** intervinientes en la acción judicial.

Su colaboración será valiosa para la investigación.

Muchas gracias!

MBA Ing. H. Beatriz P. de Gallo
Docente e Investigadora
Facultad de Ingeniería
UCASAL
www.ucasal.edu.ar

Figura II-1: Relevamiento de Puntos de Pericias de correos electrónicos

Esta acción de consulta a los usuarios expertos permitió identificar 86 puntos de pericia iniciales, que se enuncian en la Tabla II-1.

Tabla II-1: Puntos de Pericia Relevados

N°	Punto de Pericia Relevado
01	¿Quién es el titular de la casilla sss@yahoo.com.ar , ¿Cuándo se habilitó la misma?, señale si en el período de la causa fue el sistema de contacto del actor con la demandada o sus casillas de email; persona1@dominio.com.ar, usuario@yahoo.com.ar, persona2@dominio. com.ar, persona3@dominio.com.ar y persona4@dominio.com.ar e imprimir ó transcribir los mails recepcionados y emitidos desde esa bandeja de servicios entre los siguientes sujetos y fecha.
02	A fs. 33 vta. se solicita: "analice y dictamine si se encuentran constancias de autenticidad, veracidad y verosimilitud de las personas y cuentas remitentes, fecha, hora y contenido de los e-mails dirigidos a XXX, que se detallan en la demanda y que se adjuntan impresos para su cotejo".
03	A fs. 34 se solicita: "Identifique cuales fueron los equipos de origen y de destino del mensaje. Además, deberá recabar datos de utilidad que permitan determinar el contenido del e-mail en cuestión. Los servidores o proveedores de Internet – que operan a modo de enlace entre el remitente y el destinatario del e-mail- cuentan con información inherente a la fecha en que se produjo el envío o reenvío del e-mail, duración de conexión, archivos adjuntos, número de identificación de los equipos de origen y destino de las comunicaciones – datos de tráfico-.
04	A fs. 42 se solicita: "…para que se constituya en el domicilio de la demandada y extraiga copia de la información del disco rígido o la base de datos informática en donde se archivan los correos electrónicos correspondientes a las cuentas de XXX y ZZZ en Argentina; copia de la cuenta personal del Correo electrónico CCC@empresa.com.ar de XXX durante el período 2007/2009.
05	Acceder al equipo y extraer toda información de la que pudieran surgir los correos electrónicos en los que la casilla "xxxxxxx@xxxx.com.ar" figure como remitente o destinatario
06	Acorde a lo ordenado en las presentes actuaciones, en fs. 47 y que tomando en cuenta los registros informáticos existentes en la sucursal de la entidad demandada y los elementos probatorios agregados en autos, detalle los trámites efectuados a nombre del suscripto desde fecha 24/05/2011 al 15/06/2011. Todo otro dato de interés para un total esclarecimiento de los hechos.
07	Acorde a lo ordenado en las presentes actuaciones, se debe comprobar la autenticidad de los mails que obran reservados en autos. En relación a ellos determinar si los mismos son auténticos y pertenecen a sus titulares.
08	Agregue al informe los correos electrónicos enviados y recibidos desde esas cuentas entre su creación y MES de AÑO inclusive
09	Analizando el servidor de la demandada verifique la existencia actual o pasada del usuario o cuenta xx@xx e indique fecha de creación
10	Certificar o informar respecto a que los e-mail antes reseñados son verdaderos y remitidos por el servicio de atención al cliente de la firma XXX.
11	Comprobar la autenticidad de los mails que obran reservados en autos. En relación a ellos determinar si los mismos son auténticos y pertenecen a sus titulares.
12	Constate la existencia, fecha y contenido de los mismos en la cuenta de DOMINIO
13	Del análisis de todos los equipos del personal de la contraria, determine la veracidad o no de los correos que mi parte atribuye a la patronal
14	Determinar el lugar físico de origen de los correos extraídos
15	Determinar el servidor de origen y destino involucrados en la comunicación entre las casillas de correo xx@empresa.com,ar y ff@empresa2.com.ar
16	Determinar la autenticidad de la impresión de la página web y un correo electrónico adjuntado a las presentes actuaciones.
17	Determinar la autenticidad de la impresión de un correo electrónico adjuntado a las presentes actuaciones en fjs. 48.
18	Determinar la autenticidad de un correo electrónico adjuntado a las presentes actuaciones
19	Determinar la existencia del dominio xxxxxxxx y la casilla de correo electrónico que surge de los mismos
20	Determinar la existencia y veracidad de envío y recepción de los mails detallados en estos obrados.

N°	Punto de Pericia Relevado
21	Determinar la veracidad de los correos electrónicos enviados y recibidos por el Sr. XX con la empresa demandada. Y todo otro dato que considere de interés en la causa.
22	Determinar la veracidad de los correos electrónicos enviados y recibidos por el Sr. XX con la empresa demandada
23	Determinar la veracidad del correo electrónico
24	Determinar si el correo electrónico que se adjunta como prueba fue modificado o alterado y en ese caso, si la adulteración se realizó desde la computadora secuestrada.
25	Determinar si los correos que se adjuntan como prueba se encuentran en los equipos informáticos de la demandada.
26	Determinar si los correos que se envían desde un servidor validado, como por ejemplo los utilizados aquí por las partes, x@x.com o y@y.com o z@z.com pueden ser detectados desde su fuente y destino
27	Determine el origen, autenticidad y cronología de cada uno de esos mensajes.
28	Determine la existencia de las casillas de correo individualizadas en la documental incorporada.
29	Determine si el día x/x/xxxx fueron enviados del dispositivo móvil x@x.com al correo electrónico y@y.com los siguientes archivos adjuntos: a, b, c
30	Dictamine si se encuentran constancias de autenticidad, veracidad y verosimilitud de las personas y cuentas remitentes, fecha, hora y contenido de los e-mails dirigidos a USUARIO
31	El perito deberá indicar, si es posible que los mensajes enviados o recibidos, como los textos adjuntos que allí figuran, pueden haber sido alterados en sus fechas y horas de emisión o recepción.
32	En caso de desconocimiento de 4 Correos Electrónicos recibidos por la actora de contribuyentes, se practique pericia informática en el correo electrónico: usuario@dominio.com
33	En caso de desconocimiento de la veracidad de los mails solicito se designe perito informático a fin de que determine si los mismos fueron remitidos por el actor o por su padre o hermano, en sus cuentas de email.
34	En caso de desconocimiento se designe perito informático a fin de que constate la existencia, fecha y contenido de los mismos en la cuenta de Gmail.
35	En caso de negativa de todos los emails ofrecidos como prueba, solicito se realice una Pericia informática a los fines de que informe si los emails recibidos y enviados pertenecen a la Empresa y/o personas demandadas en el presente.
36	En presencia de los titulares de las cuentas de correos electrónicos involucrados, quienes deberán aportar las contraseñas correspondientes, informen si tales correos electrónicos fueron emitidos y recibidos desde y por las direcciones de correos electrónicos consignados en tal documentación.
37	Extraer todo correo electrónico que contenga uno o más datos contenidos en la siguiente lista: Nombre: XX, DNI: XXX, Teléfono: XXXX, Dirección: XXXXX.
38	Identificar la fecha y hora de recepción de los correos
39	Identificar la IP de procedencia de los correos
40	Identificar las direcciones de mail del emisor y del receptor de tales correos
41	Identifique cuales fueron los equipos de origen y de destino del mensaje. Además, deberá recabar datos de utilidad que permitan determinar el contenido del e-mail en cuestión. Los servidores o proveedores de Internet – que operan a modo de enlace entre el remitente y el destinatario del e-mail- cuentan con información inherente a la fecha en que se produjo el envío o reenvío del e-mail, duración de conexión, archivos adjuntos, número de identificación de los equipos de origen y destino de las comunicaciones – datos de tráfico.
42	Indicar si el correo que se adjunta como prueba fue remitido a más de un destinatario y quienes fueron ellos.
43	Indicar, si es posible que los mensajes enviados o recibidos, como los textos adjuntos que allí figuran, pueden haber sido alterados en sus fechas y horas de emisión o recepción
44	Indique de que dominio provinieron los emails
45	Indique el IP de donde provinieron los emails.
46	Indique fecha de creación
47	Indique la localización del IP remitente de las constancias ofrecidas.
48	Indique las fechas en que fueron enviados y recibidos los referidos correos.
49	Indique si era habitual y frecuente la comunicación entre el correo x@x.com con alguna cuenta o persona de ZZZZ.
50	Indique si puede determinar la autenticidad de los mails acompañados como prueba del derecho

N°	Punto de Pericia Relevado
	de mi mandante a la causa.
51	Informe si alguno de los mensajes enviados tenían como "asunto" temas relacionados con YYYYYYYYYYYYY
52	Informe si durante la relación contractual y aun luego de finalizada esta, las partes se enviaron mensajes de correo electrónico desde la dirección x@x.com hacia las cuentas xx y zz
53	Informe si en la cuenta x@x.com se recibieron correos electrónicos enviados desde xx y/o zz
54	Informe si los correos intercambiados por las partes pueden ser objeto de adulteración
55	Informe si los emails recibidos y enviados pertenecen a la Empresa y/o personas demandadas en el presente
56	Informe si resulta habitual la comunicación con los clientes por correo electrónico
57	Informen si tales correos electrónicos fueron emitidos y recibidos desde y por las direcciones de correos electrónicos consignados
58	Ingresar a la cuenta del correo electrónico de la parte actora, investigue, analice y determine la validez y autenticidad del envío por parte de la empresa EMPRESA S.A. y/o COMERCIO de emails conteniendo la propuesta contractual. Certifique la validez de los emails impresos que se acompañan como prueba.
59	Manifieste si puede precisar si alguno de esos mensajes tenia archivos adjuntos con algún contenido relacionado
60	Para el caso de desconocimiento de constancia de mails se designe perito informático a los fines de que determine la fecha de emisión y veracidad de los correos que fueran cuestionados
61	Proceda al copiado digital de los correos que contengan la información vinculada a la causa en un medio de almacenamiento extraíble, constituyéndose en depositario del mismo y garantizando los pasos mínimos previstos por la informática forense (adquisición, preservación, obtención y presentación).
62	se debe comprobar la autenticidad efectuada del intercambio por correo electrónico
63	Se determine el número de IP de donde provinieron los emails.
64	Se expida sobre si la cuenta de correo electrónico usuario@dominio.com es utilizada por el Sr. USUARIO e indique si dicha cuenta pertenece a alguna de las firmas demandadas (FIRMA1 y/o FIRMA2),
65	Se provea pericial informática a fin de que el experto del análisis de todos los equipos del personal de la contraria, determine la veracidad o no de los correos que mi parte atribuye a la patronal.
66	Se solicita a V.S. se desinsacule perito ingeniero informático de la lista quien teniendo a la vista los correos electrónicos determine la autenticidad de los mismos, puerto de entrada y salida y todo otro dato de interés.
67	Se verifique y se constate de que dominio provinieron los emails
68	Se verifique y se constate la autenticidad de la documentación acompañada obrando requisitoria pericial a foja 62 hasta foja 107 del cuerpo principal de los autos citados.
69	Se verifique y se constate la autenticidad de la documentación acompañada
70	Si de los registros del mismo surge que la cuenta xx@xx y yy@yy.com, fue borrada en su caso indique fecha
71	Si dichos correos fueron mantenidos entre la cuenta de correo electrónico del Sr. USUARIO1 y la cuenta usuario@dominio.com perteneciente al Sr. USUARIO2
72	Si dichos e-mails han sido remitidos desde las direcciones que figuran en los mismos y si han sido remitidos hacia las direcciones que figuran en ellos.
73	Si el contenido de dichos correos, se condice con el contenido de los correos acompañados en la documental
74	Si en base a las consideraciones previas, puede dictaminar si los emails intercambiados por las partes y adjuntados en formato papel, son auténticos
75	si fueron recibidos en el destinatario o no
76	si ha detectado algún indicio de manipulación en los mismos
77	Si los "Textos Adjuntos" coinciden con la documentación acompañada por mi parte en el punto anterior
78	Si los correos electrónicos que se adjuntan impresos a la presente demanda, corresponden a la cuenta de correo electrónico usuario@dominio.com perteneciente al Sr. NOMBRE APELLIDO, tarea para lo cual esta parte ofrece poner a su disposición el acceso a la cuenta de correo electrónico referida
79	Si los correos electrónicos quedan documentados en los servidores

N°	Punto de Pericia Relevado
80	Si los correos que allí aparecen como enviados o recibidos son de las fechas que indica mi parte en su demanda y si son coincidentes con los que se mencionan en b.18 (Prueba de la actora) que se detallan como Sobre carpeta de mails y documentos adjuntos del correo de la accionante xx@yahoo.com.ar (está mencionado en la demanda y consta de 71 fojas)
81	Si los e-mails intercambiados que se adjuntan en soporte papel al expediente resultan coincidentes con los efectivamente remitidos y recepcionados informáticamente
82	Supletoriamente y para el hipotético caso de desconocimiento de 4 Correos Electrónicos recibidos por la actora de contribuyentes, se practique pericia informática en el correo electrónico: yo_estudio@hotmail.com.
83	Teniendo a la vista los correos electrónicos determine la autenticidad de los mismos, puerto de entrada y salida y todo otro dato de interés.
84	Verificar la autenticidad de los correos electrónicos extraídos
85	Verifique los mails fueron remitidos entre PADRE (padre@live.com), persona (persona@hotmail.com) e HIJO hijo@hotmail.com
86	Y si los textos que se encuentran reservados en Secretaría corresponden a los enviados por los servidores respectivos

Con estos resultados, se realizó un Focus Group con la participación de los investigadores del Grupo de Forensia Digital de la UCASAL, que además actúan como peritos, a fin de reducir los 86 puntos a un número adecuado considerando que muchos de esos puntos de pericia inicialmente encuestados expresaban lo mismo pero con diferentes palabras. Así, se obtuvieron 46 puntos de pericia que se indican a continuación (Tabla II-2).

Tabla II-2: Puntos de Pericia Resultantes

N°	PUNTO DE PERICIA
1)	¿Quién es el titular de la casilla salta@yahoo.com.ar? , ¿Cuándo se habilitó la misma?, si en el período de la causa fue el sistema de contacto del actor con la demandada o sus casillas de email; persona1@dominio.com.ar, usuario@yahoo.com.ar, persona2@dominio.com.ar, persona3@dominio.com.ar y persona4@dominio.com.ar e imprimir ó transcribir los mails recepcionados y emitidos desde esa bandeja de servicios entre esos sujetos y fecha.
2)	Acceder al equipo y extraer toda información de la que pudieran surgir los correos electrónicos en los que la casilla "xxxxxxx@xxxx.com.ar" figure como remitente o destinatario.
3)	Agregue al informe los correos electrónicos enviados y recibidos desde esas cuentas entre su creación y MES de AÑO inclusive.
4)	Constate la existencia, fecha y contenido de los correos en la cuenta de dominio "dominio.com".
5)	Determinar el lugar físico de origen de los correos extraídos.
6)	Determinar el servidor de origen y destino involucrados en la comunicación entre las casillas de correo usuario@empresa.com.ar y usuario@empresa2.com.ar.
7)	Determinar la existencia y veracidad de envío y recepción de los mails detallados en estos obrados
8)	Determinar la veracidad de los correos electrónicos enviados y recibidos por el Sr. XX con la empresa demandada.
9)	Determinar si el correo electrónico que se adjunta fue modificado o alterado y en ese caso, si la adulteración se realizó desde la computadora secuestrada.
10)	Determinar si los correos que se adjuntan como prueba se encuentran en los equipos informáticos de la demandada.
11)	Determine el origen, autenticidad y cronología de cada uno de esos mensajes.
12)	Determine la existencia de las casillas de correo individualizadas en la documental incorporada.
13)	Determine si el día x/x/xxxx fueron enviados del dispositivo móvil (desde la cuenta) x@x.com al correo electrónico y@y.com los siguientes archivos adjuntos.
14)	Dictamine si se encuentran constancias de autenticidad, veracidad y verosimilitud de las

N°	PUNTO DE PERICIA
	personas y cuentas remitentes, fecha, hora y contenido de los e-mails dirigidos a USUARIO.
15)	En caso de desconocimiento de 4 Correos Electrónicos recibidos por la actora de contribuyentes, se practique pericia informática en el correo electrónico: usuario@dominio.com.
16)	En caso de desconocimiento de la veracidad de los mails solicito se designe perito informático a fin de que determine si los mismos fueron remitidos por el actor o por su padre o hermano, en sus cuentas de email.
17)	Extraer todo correo electrónico que contenga uno o más datos contenidos en la siguiente lista: Nombre: XX, DNI: XXX, Teléfono: XXXX, Dirección: XXXXX.
18)	Identifique cuales fueron los equipos de origen y de destino del mensaje. Además, deberá recabar datos de utilidad que permitan determinar el contenido del e-mail en cuestión. Los servidores o proveedores de Internet – que operan a modo de enlace entre el remitente y el destinatario del e-mail- cuentan con información inherente a la fecha en que se produjo el envío o reenvío del e-mail, duración de conexión, archivos adjuntos, número de identificación de los equipos de origen y destino de las comunicaciones – datos de tráfico.
19)	Indicar si el correo que se adjunta como prueba fue remitido a más de un destinatario y quienes fueron ellos.
20)	Indique de que dominio provinieron los emails.
21)	Indique el IP de donde provinieron los emails.
22)	Indique la localización del IP remitente de las constancias ofrecidas.
23)	Indique las fechas en que fueron enviados y recibidos los referidos correos.
24)	Indique si era habitual y frecuente la comunicación entre el correo x@x.com con alguna cuenta o persona de ZZZZ.
25)	Informe si alguno de los mensajes enviados tenían como “asunto” temas relacionados con YYYYYYYYYYYY.
26)	Informe si durante la relación contractual y aun luego de finalizada esta, las partes se enviaron mensajes de correo electrónico desde la dirección x@x.com hacia las cuentas xx y zz.
27)	Informe si en la cuenta x@x.com se recibieron correos electrónicos enviados desde xx y/o zz.
28)	Informe si los correos intercambiados por las partes pueden ser objeto de adulteración.
29)	Informen si tales correos electrónicos fueron emitidos y recibidos desde y por las direcciones de correos electrónicos consignados.
30)	Ingresar a la cuenta del correo electrónico de la parte actora, investigue, analice y determine la validez y autenticidad del envío por parte de la empresa EMPRESA S.A. y/o COMERCIO de emails conteniendo la propuesta contractual. Certifique la validez de los emails impresos que se acompañan como prueba.
31)	Manifieste si puede precisar si alguno de esos mensajes tenía archivos adjuntos con algún contenido relacionado.
32)	Para el caso de desconocimiento de constancia de mails se designe perito informático a los fines de que determine la fecha de emisión y veracidad de los correos que fueran cuestionados.
33)	Se determine el número de IP de donde provinieron los emails.
34)	Se expida sobre si la cuenta de correo electrónico usuario@dominio.com es utilizada por el Sr. USUARIO e indique si dicha cuenta pertenece a alguna de las firmas demandadas (FIRMA1 y/o FIRMA2).
35)	Si (los correos electrónicos) fueron recibidos en el destinatario o no.
36)	Si dichos correos fueron mantenidos entre la cuenta de correo electrónico del Sr. USUARIO1 y la cuenta usuario@dominio.com perteneciente al Sr. USUARIO2.
37)	Si dichos e-mails se remitieron desde las direcciones que figuran en los mismos y si han sido remitidos hacia las direcciones que figuran en ellos.
38)	Si el contenido de dichos correos, se condice con el contenido de los correos acompañados en la documental.
39)	Si en base a las consideraciones previas, puede dictaminar si los emails intercambiados por las partes y adjuntados en formato papel, son auténticos.
40)	Si los correos electrónicos que se adjuntan impresos a la presente demanda, corresponden a la cuenta de correo electrónico usuario@dominio.com perteneciente al Sr. NOMBRE APELLIDO, tarea para lo cual esta parte ofrece poner a su disposición el acceso a la cuenta de correo electrónico referida.
41)	Si los correos que allí aparecen como enviados o recibidos son de las fechas que indica mi parte en su demanda y si son coincidentes con los que se mencionan en b.18 (Prueba de la actora)

N°	PUNTO DE PERICIA
	que se detallan como Sobre carpeta de mails y documentos adjuntos del correo de la accionante usuario@dominio.com.ar.
42)	Si los e-mails intercambiados que se adjuntan en soporte papel al expediente resultan coincidentes con los efectivamente remitidos y recepcionados informáticamente.
43)	Teniendo a la vista los correos electrónicos determine la autenticidad de los mismos, puerto de entrada y salida y todo otro dato de interés.
44)	Verificar la autenticidad de los correos electrónicos extraídos.
45)	Verifique los mails que fueron remitidos entre PADRE (padre@dominio.com), persona (persona@dominio.com) e HIJO hijo@dominio.com.
46)	Y si los textos que se encuentran reservados en Secretaría corresponden a los enviados por los servidores respectivos.

Estos 46 puntos de pericia se tomaron como requerimientos para la construcción de OntoFoCE, y se definieron 21 preguntas de competencia que permiten responderlos. Las preguntas de competencia son las siguientes:

PC01: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?

PC02: Dado un correo CE ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?

PC03: Dado un correo CE ¿A qué cuentas se remitió el correo?

PC04: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del emisor?

PC05: Dado un correo CE ¿Cuál es el alias de usuario y dirección de e-mail del Receptor?

PC06: Dado un correo CE ¿Cuál fue el cliente de correo utilizado por cada usuario?

PC07: Dado un correo CE ¿Cuál fue el equipo desde el cual se emitió el correo?

PC08: Dado un correo CE ¿Cuál fue el equipo en el que se recibió el correo?

PC09: Dado un correo, un emisor y un receptor ¿cuál es la secuencia de equipos por los que pasó ese correo?

PC10: Dado una cuenta C ¿cuáles son los correos que emitió?

PC11: Dado una cuenta C ¿cuáles son los correos que recibió?

PC12: Dado una cuenta C1 ¿se ha emitido un correo hacia la cuenta C2?

PC13: Dado una cuenta C1 ¿se ha recibido un correo desde la cuenta C2?

PC14: Dada una dirección IP ¿cuál sería la localización geográfica del mismo?

PC15: ¿Cuáles son los correos que han pasado por el dispositivo que posee una IP dada?

PC16: ¿Cuáles son los mails enviados desde una determinada cuenta en una fecha dada?

PC17: ¿Cuáles son los mails recibidos por una determinada cuenta en una fecha dada?

PC18: Dada una palabra clave ¿Figura en el asunto de un correo?

PC19: Dada una palabra clave ¿Figura en el cuerpo de un correo?

PC20: Dada una palabra clave ¿Figura en el adjunto de un correo?

PC21: ¿Cuáles son los correos intercambiados entre las cuentas C1 y C2 en un rango de fechas dado?

Luego se realizó un cruce entre ambos conjuntos (puntos de pericia y preguntas de competencia) a fin de verificar que cada punto de pericia identificado, tenga una o más preguntas de competencia asociadas que lo responde, esto se muestra en la Tabla II-3. En la que las sucesivas filas indican los puntos de pericia, las columnas señalan las preguntas de competencia que puede responder la ontología, y la intersección de cada fila/columna que se ha sombreado señala cuales son las preguntas de competencia que responden al punto de pericia.

Tabla II-3: Matriz de Relación Puntos de Pericia y Preguntas de Competencia

		PREGUNTAS DE COMPETENCIA																			
N°	PUNTO DE PERICIA	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
1)	¿Quién es el titular de la casilla salta @yahoo.com.ar?, ¿Cuándo se habilitó la misma?, si en el período de la causa fue el sistema de contacto del actor con la demandada o sus casillas de email; persona1@dominio.com.ar, usuario@ yahoo.com.ar, persona2@dominio.com .ar, persona3@dominio.com.ar y perso-na4@dominio.com.ar e imprimir ó transcribir los mails recepcionados y emitidos desde esa bandeja de servicios entre esos sujetos y fecha.																					
2)	Acceder al equipo y extraer toda información de la que pudieran surgir los correos electrónicos en los que la casilla “xxxxxxx@xxxx.com.ar” figure como remitente o destinatario.																					
3)	Agregue al informe los correos electrónicos enviados y recibidos desde esas cuentas entre su creación y MES de AÑO inclusive.																					
4)	Constate la existencia, fecha y contenido de los correos en la cuenta de dominio “dominio.com”.																					
5)	Determinar el lugar físico de origen de los correos extraídos.																					
6)	Determinar el servidor de origen y des-tino involucrados en la comunicación entre las casillas usuario@empresa .com,ar y usuario@empresa2. com.ar.																					
7)	Determinar la existencia y veracidad de envío y recepción de los mails detallados en estos obrados																					
8)	Determinar la veracidad de los correos electrónicos enviados y recibidos por el Sr. XX con la empresa demandada.																					
9)	Determinar si el correo electrónico que se adjunta fue modificado o alterado y en ese caso, si la adulteración se realizó desde la computadora secuestrada.																					

		PREGUNTAS DE COMPETENCIA																				
N°	PUNTO DE PERICIA	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
10)	Determinar si los correos que se adjuntan como prueba se encuentran en los equipos informáticos de la demandada.																					
11)	Determine el origen, autenticidad y cronología de cada uno de esos mensajes.																					
12)	Determine la existencia de las casillas de correo individualizadas en la documental incorporada.																					
13)	Determine si el día x/x/xxxx fueron enviados del dispositivo móvil (desde la cuenta) x@x.com al correo electrónico y@y.com los siguientes archivos adjuntos.																					
14)	Dictamine si se encuentran constancias de autenticidad, veracidad y verosimilitud de las personas y cuentas remitentes, fecha, hora y contenido de los e-mails dirigidos a USUARIO.																					
15)	En caso de desconocimiento de 4 Correos Electrónicos recibidos por la actora de contribuyentes, se practique pericia informática en el correo electrónico: usuario@dominio.com.																					
16)	En caso de desconocimiento de la veracidad de los mails solicito se designe perito informático a fin de que determine si los mismos fueron remitidos por el actor o por su padre o hermano, en sus cuentas de email.																					
17)	Extraer todo correo electrónico que contenga uno o más datos contenidos en la siguiente lista: Nombre: XX, DNI: XXX, Teléfono: XXXX, Dirección: XXXXX.																					
18)	Identifique cuales fueron los equipos de origen y de destino del mensaje. Además, deberá recabar datos de utilidad que permitan determinar el contenido del e-mail en cuestión. Los servidores o proveedores de Internet – que operan a modo de enlace entre el remitente y el																					

		PREGUNTAS DE COMPETENCIA																				
N°	PUNTO DE PERICIA	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
	destinatario del e-mail- cuentan con información inherente a la fecha en que se produjo el envío o reenvío del e-mail, duración de conexión, archivos adjuntos, número de identificación de los equipos de origen y destino de las comunicaciones – datos de tráfico.																					
19)	Indicar si el correo que se adjunta como prueba fue remitido a más de un destinatario y quienes fueron ellos.																					
20)	Indique de que dominio provinieron los emails.																					
21)	Indique el IP de donde provinieron los emails.																					
22)	Indique la localización del IP remitente de las constancias ofrecidas.																					
23)	Indique las fechas en que fueron enviados y recibidos los referidos correos.																					
24)	Indique si era habitual y frecuente la comunicación entre el correo x@x.com con alguna cuenta o persona de ZZZZ.																					
25)	Informe si alguno de los mensajes enviados tenían como “asunto” temas relacionados con YYYYYYYYYYYY.																					
26)	Informe si durante la relación contractual y aun luego de finalizada esta, las partes se enviaron mensajes de correo electrónico desde la dirección x@x.com hacia las cuentas xx y zz.																					
27)	Informe si en la cuenta x@x.com se recibieron correos electrónicos enviados desde xx y/o zz.																					
28)	Informe si los correos intercambiados por las partes pueden ser objeto de adulteración.																					
29)	Informen si tales correos electrónicos fueron emitidos y recibidos desde y por las direcciones de correos electrónicos consignados.																					

		PREGUNTAS DE COMPETENCIA																				
N°	PUNTO DE PERICIA	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
30)	Ingresar a la cuenta del correo electrónico de la parte actora, investigue, analice y determine la validez y autenticidad del envío por parte de la empresa EMPRESA S.A. y/o COMERCIO de emails conteniendo la propuesta contractual. Certifique la validez de los emails impresos que se acompañan como prueba.																					
31)	Manifieste si puede precisar si alguno de esos mensajes tenía archivos adjuntos con algún contenido relacionado.																					
32)	Para el caso de desconocimiento de constancia de mails se designe perito informático a los fines de que determine la fecha de emisión y veracidad de los correos que fueran cuestionados.																					
33)	Se determine el número de IP de donde provinieron los emails.																					
34)	Se expida sobre si la cuenta de correo electrónico usuario@dominio.com es utilizada por el Sr. USUARIO e indique si dicha cuenta pertenece a alguna de las firmas demandadas (FIRMA1 y/o FIRMA2).																					
35)	Si (los correos electrónicos) fueron recibidos en el destinatario o no.																					
36)	Si dichos correos fueron mantenidos entre la cuenta de correo electrónico del Sr. USUARIO1 y la cuenta usuario@dominio.com perteneciente al Sr. USUARIO2.																					
37)	Si dichos e-mails se remitieron desde las direcciones que figuran en los mismos y si han sido remitidos hacia las direcciones que figuran en ellos.																					
38)	Si el contenido de dichos correos, se condice con el contenido de los correos acompañados en la documental.																					
39)	Si en base a las consideraciones pre-vias, puede dictaminar si los emails in-																					

		PREGUNTAS DE COMPETENCIA																				
N°	PUNTO DE PERICIA	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
	tercambiados por las partes y adjuntados en formato papel, son auténticos.																					
40)	Si los correos electrónicos que se adjuntan impresos a la presente demanda, corresponden a la cuenta de correo electrónico usuario@dominio.com perteneciente al Sr. NOMBRE APELLIDO, tarea para lo cual esta parte ofrece poner a su disposición el acceso a la cuenta de correo electrónico referida.																					
41)	Si los correos que allí aparecen como enviados o recibidos son de las fechas que indica mi parte en su demanda y si son coincidentes con los que se mencionan en b.18 (Prueba de la actora) que se detallan como Sobre carpeta de mails y documentos adjuntos del correo de la accionante usuario@dominio.com.ar.																					
42)	Si los e-mails intercambiados que se adjuntan en soporte papel al expediente resultan coincidentes con los efectivamente remitidos y recepcionados informáticamente.																					
43)	Teniendo a la vista los correos electrónicos determine la autenticidad de los mismos, puerto de entrada y salida y todo otro dato de interés.																					
44)	Verificar la autenticidad de los correos electrónicos extraídos.																					
45)	Verifique los mails que fueron remitidos entre PADRE (padre@domi-nio.com), persona (persona@dominio.com) e HIJO hijo@dominio.com.																					
46)	Y si los textos que se encuentran reservados en Secretaría corresponden a los enviados por los servidores respectivos.																					

ANEXO III: REPRESENTACIONES INTERMEDIAS DE LA CONCEPTUALIZACIÓN DE OntoFoCE

En este apartado se incluyen las representaciones en forma de tablas descriptivas de los componentes de la ontología, que ayudan a mostrar los conceptos y relaciones del dominio de OntoFoCE. Mayor detalle acerca del modelo conceptual de OntoFoCE se describe en el Capítulo 3.

Se incluyen aquí:

- Tabla III-1: Diccionario de Conceptos
- Tabla III-2: Relaciones y sus inversas

El resto de las representaciones intermedias, en formato de Diagramas de Actividades y Diagramas de Objetos se incluyen en el texto principal de la tesis, Capítulos 3 y 4 fundamentalmente.

Tabla III-1: Diccionario de Conceptos

CLASE	DESCRIPCIÓN	ATRIBUTOS DE INSTANCIA	RELACIONES
Adjunto	Archivo asociado al correo electrónico con información complementaria al contenido del correo.	contenidoAdjunto	adjuntoContienePalabraClave palabraClaveEstaEnAdjunto adjuntoPerteneceAOcurrencia ocurrenciaTieneAdjunto
Asunto	Texto que expresa el tema del que trata el correo electrónico	contenidoAsunto	asuntoContienePalabraClave palabraClaveEstaEnAsunto asuntoPerteneceAOcurrencia ocurrenciaTieneAsunto
CabeceraCorreo	Bloque de texto plano que contiene información relativa al correo y al proceso de transmisión realizado.	contenidoCabeceraCorreo	cabeceraPerteneceAOcurrencia ocurrenciaTieneCabecera
ClienteCorreo	Aplicación informática que gestiona una cuenta de correo electrónico	nombreClienteCorreo	cCEsEjecutadoPorEquipoE equipoEEjecutaClienteC cCEsEjecutadoPorEquipoR equipoREjecutaClienteC emiteDesde cCSoportaCuentaE recibeEn cCSoportaCuentaR
ClienteLocal	Cliente de Correo residente en un dispositivo (PC, Teléfono, etc) y que guarda una copia de los correos enviados/recibidos en dicho dispositivo		
ClienteRemoto	Cliente de Correo al que se accede remotamente vía web		
Correo	Identificación del Correo Electrónico que está siendo analizado	idCorreo	correoEsEmitidoPorCuentaEmisor cuentaEmisorEmiteCorreo correoEsRecibidoPorCuentaReceptor cuentaReceptorRecibeCorreo correoTieneSecuencia secuenciaCorrespondeACorreo
CorreoFactible	Correo electrónico que cumple con los requisitos mínimos para ser analizado		
Cuenta	Servicio online que provee un espacio para la recepción, envío y almacenamiento de mensajes de correo electrónico.	aliasUsuario cuentaCorreo	cuentaVinculadaAExpediente expedienteVinculaCuenta
CuentaEmisor	Cuenta de correo del usuario que emite un correo electrónico		cuentaEUtilizaEquipoE equipoEEsUtilizadoPorCuentaE emiteDesde cCSoportaCuentaE correoEsEmitidoPorCuentaEmisor cuentaEmisorEmiteCorreo

CLASE	DESCRIPCIÓN	ATRIBUTOS DE INSTANCIA	RELACIONES
CuentaReceptor	Cuenta de correo del usuario que recibe un correo electrónico		cuentaRUtilizaEquipoR equipoREsUtilizadoPorCuentaR recibeEn cCSoportaCuentaR correoEsRecibidoPorCuentaReceptor cuentaReceptorRecibeCorreo
Cuerpo	Contenido del mensaje del correo electrónico	contenidoCuerpo	cuerpoPerteneceAOcurrencia ocurrenciaTieneCuerpo palabraClaveEstaEnCuerpo cuerpoContienePalabraClave
Equipo	Componente de hardware que almacena el correo electrónico durante el proceso de transmisión	descripcionEquipo macAddressEquipo	equipoTieneId idCorrespondeAEquipo equipoTieneOcurrencia ocurrenciaResideEnEquipo
EquipoEmisor	Equipo utilizado para emitir un correo. Puede ser una PC, Notebook, teléfono, etc.		equipoEEjecutaClienteC cCEsEjecutadoPorEquipoE equipoEEsUtilizadoPorCuentaE cuentaEUtilizaEquipoE
EquipoReceptor	Equipo utilizado para recibir el correo electrónico. Puede ser una PC, Notebook, Teléfono, etc.		equipoREjecutaClienteC cCEsEjecutadoPorEquipoR equipoREsUtilizadoPorCuentaR cuentaRUtilizaEquipoR
Expediente	Documento legal en donde consta la prueba digital de correo electrónico y los puntos de pericia	nroExpediente caratulaExpediente	expedienteContienePalabraClave palabraClaveEstaEnExpediente expedienteVinculaCuenta cuentaVinculadaAExpediente
Hilo	Agrupación de ocurrencias relacionadas a una cuenta del receptor	idHilo	hiloCorrespondeASecuencia secuenciaTieneHilo hiloTieneOcurrencia ocurrenciaCorrespondeAHilo
HostName	Identificación del Equipo mediante un nombre de dominio		
IdentificacionEquipo	Identificación única del hardware conectado a internet	identificadorEquipo geoLocalizacionIP propietarioIP referenteIP ispIP	idCorrespondeAEquipo equipoTieneId
IP	Identificación del Equipo mediante una dirección IP		

CLASE	DESCRIPCIÓN	ATRIBUTOS DE INSTANCIA	RELACIONES
Ocurrencia	Copia del correo electrónico que se almacena en cada dispositivo que participa en el proceso de transmisión	fechaHoraOcurrencia	esAnteriorA esSiguienteDe ocurrenciaTieneAdjunto adjuntoPerteneceAOcurrencia ocurrenciaTieneAsunto asuntoPerteneceAOcurrencia ocurrenciaTieneCabecera cabeceraPerteneceAOcurrencia ocurrenciaTieneCuerpo cuerpoPerteneceAOcurrencia ocurrenciaResideEnEquipo equipoTieneOcurrencia ocurrenciaCorrespondeAHilo hiloTieneOcurrencia
OcurrenciaDeEmision	Copia del correo electrónico residente en el equipo emisor		
OcurrenciaDeRecepcion	Copia del correo electrónico residente en el equipo receptor		
OcurrenciaDeTransmision	Copia del correo electrónico residente en el equipo servidor intermedio que participa en la transmisión del correo		
PalabraClave	Palabra utilizada para búsqueda de un tema de interés para la causa	idPalabraClave contenidoPalabraClave	palabraClaveEstaEnExpediente expedienteContienePalabraClave palabraClaveEstaEnCuerpo cuerpoContienePalabraClave palabraClaveEstaEnAdjunto adjuntoContienePalabraClave palabraClaveEstaEnAsunto asuntoContienePalabraClave
Secuencia	Serie de hilos de ocurrencias de correo electrónico asociadas a un mismo correo electrónico.	idSecuencia	secuenciaCorrespondeACorreo correoTieneSecuencia secuenciaTieneHilo hiloCorrespondeASecuencia
Servidor	Equipo utilizado para almacenar el correo electrónico, cada vez que se debe seleccionar un camino de distribución del correo durante la transmisión		

Tabla III-2: Relaciones y sus inversas

DOMINIO	RELACIÓN	RANGO	RELACIÓN INVERSA
Adjunto	adjuntoPerteneceAOcurrencia	Ocurrencia	ocurrenciaTieneAdjunto
Asunto	asuntoPerteneceAOcurrencia	Ocurrencia	ocurrenciaTieneAsunto
CabeceraCorreo	cabeceraPerteneceAOcurrencia	Ocurrencia	ocurrenciaTieneCabecera
ClienteCorreo	cCEsEjecutadoPorEquipoE	EquipoEmisor	equipoEEjecutaClienteC
ClienteCorreo	cCEsEjecutadoPorEquipoR	EquipoReceptor	equipoREjecutaClienteC
Correo	correoEsEmitidoPorCuentaEmisor	CuentaEmisor	cuentaEmisorEmiteCorreo
Correo	correoEsRecibidoPorCuentaReceptor	CuentaReceptor	cuentaReceptorRecibeCorreo
Correo	correoTieneSecuencia	Secuencia	secuenciaCorrespondeACorreo
Cuenta	cuentaVinculadaAExpediente	Expediente	expedienteVinculaCuenta
CuentaEmisor	cuentaEUtilizaEquipoE	EquipoEmisor	equipoEEsUtilizadoPorCuentaE
CuentaEmisor	emiteDesde	ClienteCorreo	cCSoportaCuentaE
CuentaReceptor	cuentaRUtilizaEquipoR	EquipoReceptor	equipoREsUtilizadoPorCuentaR
CuentaReceptor	recibeEn	ClienteCorreo	cCSoportaCuentaR
Cuerpo	cuerpoPerteneceAOcurrencia	Ocurrencia	ocurrenciaTieneCuerpo
Equipo	equipoTieneId	IdentificacionEquipo	ipCorrespondeAEquipo
Equipo	equipoTieneOcurrencia	Ocurrencia	ocurrenciaResideEnEquipo
Expediente	expedienteContienePalabraClave	PalabraClave	palabraClaveCorrespondeAExpediente
Hilo	hiloCorrespondeASecuencia	Secuencia	secuenciaTieneHilo
Hilo	hiloTieneOcurrencia	Ocurrencia	ocurrenciaCorrespondeAHilo
Ocurrencia	esAnteriorA	Ocurrencia	esSiguienteDe
PalabraClave	palabraClaveEstaEnCuerpo	Cuerpo	cuerpoContienePalabraClave
PalabraClave	palabraClaveEstaEnAdjunto	Adjunto	adjuntoContienePalabraClave
PalabraClave	palabraClaveEstaEnAsunto	Asunto	asuntoContienePalabraClave

ANEXO IV: AUTORIZACIONES DE USO DE DATOS

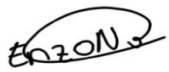
A continuación se incluyen las notas de Autorización otorgada por los usuarios de cuentas de correo electrónico que aceptaron formar parte de las pruebas realizadas sobre OntoFoCE y sobre ObE Forensics.

AUTORIZACIÓN DE USO DE DATOS

Por la presente, doy mi expresa autorización, para la utilización de mi nombre y la cuenta de correo denominada **enzo.notario@gmail.com** a fin de que se incluyan estos datos en el trabajo de Tesis denominado "Una Ontología del Correo Electrónico y su Trazabilidad como soporte para Forensia Digital", cuya autora es la MBA Ing. H. Beatriz P. de Gallo.

Lugar y fecha: Salta, 25 de Septiembre de 2019

Firma y Aclaración:



Enzo Rubén Notario

37.636.222

AUTORIZACIÓN DE USO DE DATOS

Por la presente, doy mi expresa autorización, para la utilización de mi nombre y la cuenta de correo denominada carlos.neil@uai.edu.ar a fin de que se incluyan estos datos en el trabajo de Tesis denominado "Una Ontología del Correo Electrónico y su Trazabilidad como soporte para Forensia Digital", cuya autora es la MBA Ing. H. Beatriz P. de Gallo.

Lugar y fecha: Buenos Aires, 25 de setiembre de 2019

Firma y Aclaración:

Carlos Neil

DNI: 14.031.103

AUTORIZACIÓN DE USO DE DATOS

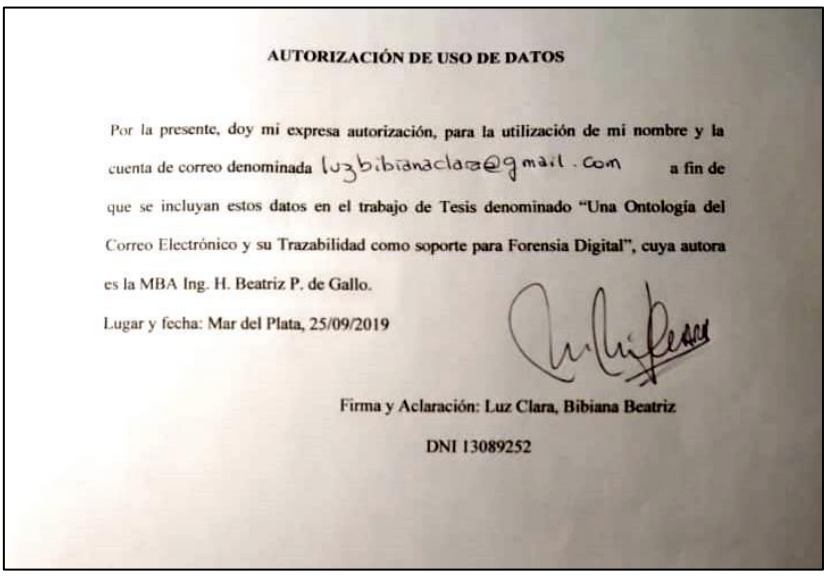
Por la presente, doy mi expresa autorización, para la utilización de mi nombre y la cuenta de correo denominada erivett.83@gmail.com a fin de que se incluyan estos datos en el trabajo de Tesis denominado "Una Ontología del Correo Electrónico y su Trazabilidad como soporte para Forensia Digital", cuya autora es la MBA Ing. H. Beatriz P. de Gallo.

Lugar y fecha: Salta, 24/09/19

Firma y Aclaración:

Rivetti Esteban

30222110



AUTORIZACIÓN DE USO DE DATOS

Por la presente, doy mi expresa autorización, para la utilización de mi nombre y la cuenta de correo denominada luzbibianaclara@gmail.com a fin de que se incluyan estos datos en el trabajo de Tesis denominado "Una Ontología del Correo Electrónico y su Trazabilidad como soporte para Forensia Digital", cuya autora es la MBA Ing. H. Beatriz P. de Gallo.

Lugar y fecha: Mar del Plata, 25/09/2019

Firma y Aclaración: Luz Clara, Bibiana Beatriz

DNI 13089252

ANEXO IV: INFORME DE ANÁLISIS FORENSE PARA ESCENARIO 1

En este apartado se incluye el informe emitido por la aplicación ObE Forensics, para el correo C1, tomado como ejemplo en el caso de estudio del Escenario 1, incluye los datos de identificación del correo (cuentas, asunto, entre otros), con el agregado de las preguntas de competencia PC01 a PC09 y sus correspondientes respuestas.

El informe consta de tres páginas, que se muestran a continuación.

ObE Forensics 1 E-mail

1° E-mail

Asunto: **Colaboracion en la investigacion**

ID: **<CAH18OQWHN0G8GUCadSqJnTN5q3ebjbkS0roX0ZMvD+tTQN+uLA@mail.gmail.com>**

Fecha: **2018-07-11T09:06:56-03:00**

Emisor	Receptor
Ing. H. Beatriz P. de Gallo bgallo@ucasal.edu.ar	**carlos.neil@uai.edu.ar** carlos.neil@uai.edu.ar

Preguntas de competencia

1) ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?

Fecha: 2018-07-11T09:07:06-03:00
Identificador: 2002:a2e:558c::

2) ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?

Fecha: 2018-07-11T09:07:17-03:00
Identificador: 10.1.100.14

3) ¿A qué cuentas se remitió el correo?

Alias usuario: carlos.neil@uai.edu.ar
Direccion de E-mail: carlos.neil@uai.edu.ar

4) ¿Cuál es el alias usuario y dirección de e-mail del Emisor?

Alias usuario: Ing. H. Beatriz P. de Gallo
Dirección de E-mail: bgallo@ucasal.edu.ar

5) ¿Cuál es el alias usuario y dirección de e-mail del Receptor?

Alias usuario: carlos.neil@uai.edu.ar
Dirección de E-mail: carlos.neil@uai.edu.ar

6) ¿Cuál fue el cliente de correo utilizado por cada usuario?

Alias usuario: carlos.neil@uai.edu.ar
Direccion de E-mail: carlos.neil@uai.edu.ar
ClienteLocal: Outlook

Alias usuario: Ing. H. Beatriz P. de Gallo
Direccion de E-mail: bgallo@ucasal.edu.ar
ClienteLocal: ThunderBird

7) ¿Cuál fue el equipo desde el cual se emitió el correo?

Equipo: Notebook Beatriz
MAC Address: F0:E1:D2:C3:B4:A5
Identificador: 2002:a2e:558c::

8) ¿Cuál fue el equipo en el que se recibió el correo?

Equipo: PC Carlos
MAC Address: F5:E2:D3:C4:B5:00
Identificador: 10.1.100.14

9) ¿Cuál fue la trazabilidad del mensaje desde su emisión hasta su recepción?

Ocurrencia de Emisión

Equipo Emisor: **2002:a2e:558c::**
Fecha: **2018-07-11 09:07:06**

Ocurrencia de Transmisión 1

Servidor: **mail-lj1-f180.google.co**
Fecha: **2018-07-11 09:07:09**

Ocurrencia de Transmisión 2

Servidor: **209.85.208.180**
Fecha: **2018-07-11 09:07:09**

Ocurrencia de Transmisión 3

Servidor: **mail.ucasal.edu.ar**
Fecha: **2018-07-11 09:07:09**

Ocurrencia de Transmisión 4

Servidor: **127.0.0.1**
Fecha: **2018-07-11 09:07:13**

Ocurrencia de Transmisión 5

Servidor: **127.0.0.1**
Fecha: **2018-07-11 09:07:13**

Ocurrencia de Transmisión 6

Servidor: **127.0.0.1**
Fecha: **2018-07-11 09:07:14**

Ocurrencia de Transmisión 7

Servidor: **mail.ucasal.edu.ar**
Fecha: **2018-07-11 09:07:14**

Ocurrencia de Transmisión 8

Servidor: **200.10.180.145**
Fecha: **2018-07-11 09:07:15**

Ocurrencia de Transmisión 9

Servidor: **FNDEXCHG01.adm.vaneduc.edu.ar**
Fecha: **2018-07-11 09:07:15**

Ocurrencia de Transmisión 10

Servidor: **10.1.100.15**
Fecha: **2018-07-11 09:07:17**

Ocurrencia de Recepción

Equipo Receptor: **10.1.100.14**
Fecha: **2018-07-11 09:07:17**

ANEXO V: INFORME DE ANÁLISIS FORENSE PARA ESCENARIO 2

Este es el apartado que contiene el informe emitido por la aplicación ObE Forensics, para el correo C1, pero considerando que el mismo se envía a tres receptores, tomado como ejemplo en el caso de estudio del Escenario 2, incluye los datos de identificación del correo (cuentas, asunto, entre otros), con el agregado de las preguntas de competencia PC10, PC14 y PC15 y sus correspondientes respuestas.

El informe consta de cinco páginas, que se muestran a continuación.

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 10

Correos electrónicos enviados por bgallo@ucasal.edu.ar

Total: **3 E-mails**

1° E-mail

Asunto: **Colaboracion en investigacion**

ID: **<CAH18OQUiLqm+iDyLJkrHg9kOto2PRDR4AG3mp1RAQDCwD2JccA@mail.gmail.com>**

Fecha: **2018-07-08T12:41:56-03:00**

Emisor	Receptor
Ing. H. Beatriz P. de Gallo bgallo@ucasal.edu.ar	**Enzo Notario** enzo.notario@gmail.com

2° E-mail

Asunto: **Colaboración en investigación**

ID: **<CAH18OQUiLqm+iDyLJkrHg9kOto2PRDR4AG3mp1RAQDCwD2JccA@mail.gmail.com>**

Fecha: **2018-07-08T12:41:56-03:00**

Emisor	Receptor
Ing. H. Beatriz P. de Gallo bgallo@ucasal.edu.ar	**Esteban Rivetti** erivetti83@gmail.com

3° E-mail

Asunto: **Colaboración en investigación**

ID: **<CAH18OQUiLqm+iDyLJkrHg9kOto2PRDR4AG3mp1RAQDCwD2JccA@mail.gmail.com>**

Fecha: **2018-07-08T12:41:56-03:00**

Emisor

Ing. H. Beatriz P. de Gallo
bgallo@ucasal.edu.ar

Receptor

luz bibiana Clara
luzbibianaclara@gmail.com

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 14

Geolocalización de la IP 209.85.208.170

IP **209.85.208.170**
Código ISO **US**
País **United States**
Ciudad **Ashburn**
Abrev. Provincia/Estado **VA**
Provincia/Estado **Virginia**
Código postal **20149**
Latitud **39.0438**
Longitud **-77.4874**
Zona horaria **America/New_York**
Continente **Unknown**
Moneda **USD**

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 15

Correos electrónicos que han pasado por la IP 209.85.208.170

Total: **3 E-mails**

1° E-mail

Asunto: **Colaboracion en investigacion**

ID: **<CAH18OQUiLqm+iDyLJkrHg9kOto2PRDR4AG3mp1RAQDCwD2JccA@mail.gmail.com>**

Fecha: **2018-07-08T12:41:56-03:00**

Emisor	Receptor
Ing. H. Beatriz P. de Gallo bgallo@ucasal.edu.ar	**Enzo Notario** enzo.notario@gmail.com

2° E-mail

Asunto: **Colaboración en investigación**

ID: **<CAH18OQUiLqm+iDyLJkrHg9kOto2PRDR4AG3mp1RAQDCwD2JccA@mail.gmail.com>**

Fecha: **2018-07-08T12:41:56-03:00**

Emisor	Receptor
Ing. H. Beatriz P. de Gallo bgallo@ucasal.edu.ar	**Esteban Rivetti** erivetti83@gmail.com

3° E-mail

Asunto: **Colaboración en investigación**

ID: **<CAH18OQUiLqm+iDyLJkrHg9kOto2PRDR4AG3mp1RAQDCwD2JccA@mail.gmail.com>**

Fecha: **2018-07-08T12:41:56-03:00**

Emisor	Receptor
Ing. H. Beatriz P. de Gallo bgallo@ucasal.edu.ar	**luz bibiana Clara** luzbibianaclara@gmail.com

ANEXO VI: INFORME DE ANÁLISIS FORENSE PARA ESCENARIO 3

Este apartado contiene el informe emitido por la aplicación ObE Forensics, para los siete correos considerados en el caso de estudio del Escenario 3, incluye los datos de identificación del correo (cuentas, asunto, entre otros), con el agregado de las preguntas de competencia PC11, PC12, PC13, PC16, PC17 y PC21 y sus correspondientes respuestas.

El informe consta de siete páginas, que se muestran a continuación.

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 11

Correos electrónicos recibidos por enzo.notario@gmail.com

Total: **2 E-mails**

1° E-mail

Asunto: **Colaboracion en investigacion**

ID: **<CAH18OQUiLqm+iDyLJkrHg9kOto2PRDR4AG3mp1RAQDCwD2JccA@mail.gmail.com>**

Fecha: **2018-07-08T12:41:56-03:00**

Emisor	Receptor
Ing. H. Beatriz P. de Gallo bgallo@ucasal.edu.ar	**Enzo Notario** enzo.notario@gmail.com

2° E-mail

Asunto: **Re: Trabajo para el CoNaIISI**

ID: **<CAD3ue2d7oCzSGkLeQbh7Bj4ApK4+57ki2Z1=qgwE+HU-UU-KPQ@mail.gmail.com>**

Fecha: **2018-10-02T00:01:09-03:00**

Emisor	Receptor
juan@empresa.com.ar juan@empresa.com.ar	**Enzo Notario** enzo.notario@gmail.com

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 12

Correos electrónicos emitidos por juan@empresa.com.ar y recibidos por Rivetti erivetti83@gmail.com

Total: **1 E-mail**

1° E-mail

Asunto: **Fwd: Forensia IoT**

ID: **<CAN=18jsTKN0YLqYr-Lr_LPYgpVOd2yGp96KDdTgMyygKtSRpGg@mail.gmail.com>**

Fecha: **2019-02-23T13:27:24-03:00**

Emisor	Receptor
Juan juan@empresa.com.ar	**Esteban Rivetti erivetti83@gmail.com** Rivetti erivetti83@gmail.com

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 13

Correos electrónicos emitidos por juan@empresa.com.ar y recibidos por luzbibianaclara@gmail.com

Total: **1 E-mail**

1° E-mail

Asunto: **Reenviar: ACM TIOT Call for Papers**

ID: **<euqbmf7paegoa8p8hcfbuudr.1528243015987@email.android.com>**

Fecha: **2018-06-05T20:56:55-03:00**

Emisor	Receptor
Juan juan@empresa.com.ar	**Luzbibianaclara** luzbibianaclara@gmail.com

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 16

Correos electrónicos enviados por juan@empresa.com.ar el día 02/10/2018

Total: **1 E-mail**

1° E-mail

Asunto: **Re: Trabajo para el CoNaIISI**

ID: **<CAD3ue2d7oCzSGkLeQbh7Bj4ApK4+57ki2Z1=qgwE+HU-UU-KPQ@mail.gmail.com>**

Fecha: **2018-10-02T00:01:09-03:00**

Emisor	Receptor
juan@empresa.com.ar juan@empresa.com.ar	**Enzo Notario** enzo.notario@gmail.com

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 17

Correos electrónicos recibidos por erivetti83@gmail.com el día 08/07/2018

Total: **1 E-mail**

1° E-mail

Asunto: **Colaboración en investigación**

ID: **<CAH18OQUiLqm+iDyLJkrHg9kOto2PRDR4AG3mp1RAQDCwD2JccA@mail.gmail.com>**

Fecha: **2018-07-08T12:41:56-03:00**

Emisor	Receptor
Ing. H. Beatriz P. de Gallo bgallo@ucasal.edu.ar	**Esteban Rivetti** erivetti83@gmail.com

ObE Forensics

RESULTADO DE LA PREGUNTA DE COMPETENCIA N° 21

Correos Electrónicos intercambiados entre juan@empresa.com.ar y enzo.notario@gmail.com desde 05/06/2018 hasta 23/02/2019

Total: **2 E-mails**

1° E-mail

Asunto: **Re: Trabajo para el CoNaIISI**

ID: **<CAD3ue2d7oCzSGkLeQbh7Bj4ApK4+57ki2Z1=qgwE+HU-UU-KPQ@mail.gmail.com>**

Fecha: **2018-10-02T00:01:09-03:00**

Emisor	Receptor
juan@empresa.com.ar juan@empresa.com.ar	**Enzo Notario** enzo.notario@gmail.com

2° E-mail

Asunto: **Trabajo de Ajuste**

ID: **<CAqgwE+HU-UU-KPQCAqgwE+HU-UU-KPQ@mail.gmail.com>**

Fecha: **2018-10-16T00:01:09-03:00**

Emisor	Receptor
Enzo Notario enzo.notario@gmail.com	**Juan** juan@empresa.com.ar

ANEXO VII: CÓDIGO OWL DE OntoFoCE

En la dirección https://digilab.ucasal.edu.ar/owl[72]/ se muestra el código OWL de la ontología, considerando 4 versiones:

- Código OWL de OntoFoCE sin instanciaciones
- Código OWL de OntoFoCE con las instanciaciones para el caso de estudio del Escenario 1 descriptos en el capítulo 3.
- Código OWL de OntoFoCE con las instanciaciones para el caso de estudio del Escenario 2 descriptos en el capítulo 3.
- Código OWL de OntoFoCE con las instanciaciones para el caso de estudio del Escenario 3 descriptos en el capítulo 3.

[72] Esta página se deshabilita a partir del 20/12/2019. Los interesados en conocer el código, sírvanse remitir su solicitud al correo bgallo@ucasal.edu.ar .

ANEXO VIII: INSTRUCTIVO DE USO DE ObE Forensics

Con el objetivo de que el lector pueda acceder a la aplicación en modo de prueba, se habilitó una primera pantalla de acceso para las siguientes credenciales:

Usuario: prueba@digilab.ucasal.edu.ar

Contraseña: prueba

Una vez que se ingresa, es posible correr la aplicación con los datos del caso de estudio descripto en el Escenario 3.

Y por otra parte, puede ingresar a la versión COMPLETA de la aplicación, sin restricciones de ningún tipo, mediante el procedimiento de registro que se indica a continuación.

Previo a la utilización de ObE Forensics, el Perito debe registrarse en la aplicación. A continuación, se presenta la pantalla para ingresar los datos de registro del nuevo usuario (Figura VIII-1).

Figura VIII-1: Pantalla de Registro de Datos de Nuevo Usuario

Una vez que el usuario registra sus datos (nombre, correo electrónico, contraseña y confirmación de contraseña), debe aceptar los Términos y Condiciones de Uso, cliqueando en la casilla correspondiente.

Para registrarse como usuario, es necesario ingresar a la aplicación por el link *Registrarse*, que habilita la pantalla de "Términos y Condiciones de Uso" (Figura VIII-2) en la que se establece el marco legal al que debe ajustarse el usuario para utilizar la aplicación.

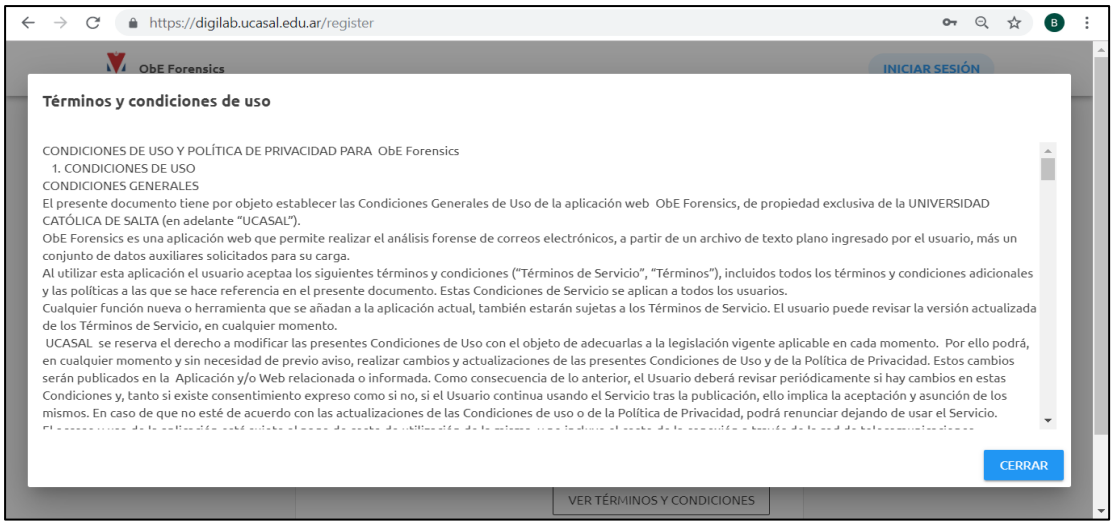


Figura VIII-2: Pantalla sobre Términos y Condiciones de Uso de ObE Forensics

Una vez que el usuario recibe sus credenciales de acceso, puede utilizar ObE Forensics sucesivas veces, y en distintas ocasiones, con diferentes correos o conjuntos de correos.

Las credenciales quedarán habilitadas de manera permanente, de acuerdo a las condiciones de registro y uso de la aplicación. Por el momento, al tratarse de una versión Beta, la aplicación no tiene restricciones de uso más que las indicadas para tramitar el formulario de registro.

Cabe mencionar que en los datos que se trabajan en cada sesión se mantienen vigentes durante la misma y para mantener las condiciones de reserva y privacidad de los datos que componen la evidencia digital que se está analizando, todos los datos ingresados y procesados se borran cuando que el usuario cierra su cesión de trabajo.

La pantalla de la Figura VIII-3 muestra los datos de ingreso a la aplicación para el usuario registrado.

Figura VIII-3: Pantalla de registro de datos de acceso para los usuarios registrados

REFERENCIAS BIBLIOGRÁFICAS

Akremi, A., Sallay, H., Rouached, M., Bouaziz, R., & Abid, M. (2015). Forensics-aware web services composition and ranking. *Proceedings of the 17th International Conference on Information Integration and Web-Based Applications &Services - IiWAS '15*, 1–10. https://doi.org/10.1145/2837185.2837226

Al-Zarouni, M. (2004). Tracing E-mail Headers. *2nd Australian Computer Network {\&} Information Forensics Conference, November 25th 2004, Perth, Western Australia, Forensic Computing - Evidence on the Move from Desktops to Networks, Conference Proceedings*, *0700*(May), 16–30.

Alberto, J., Luna, G., López, M. S. M., Ingrid, I., & Torres, D. (2012). Metodologías y métodos para la construcción de ontologías. *Scientia et Technica Año XVII, No 50, Universidad Tecnológica de Pereira. ISSN 0122-1701*, (50), 133–140.

Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. *Computers & Security*, *68*, 160–196. https://doi.org/https://doi.org/10.1016/j.cose.2017.04.006

Ali, M. (2016). Provenance-based Data Traceability Model and Policy Enforcement Framework for Cloud Services by. *UNIVERSITY OF SOUTHAMPTON*.

Alzaabi, M., Martin, T. A., Taha, K., & Jones, A. (2015). The use of ontologies in forencis analysis of smartphone content. *The Journal of Digital Forensics, Security and Law*, *10*(4), 105–114.

Alzaabi, M., Martin, T., Taha, K., & Jones, A. (2017). The Use of Ontologies in Forensic Analysis of Smartphone Content. *Journal of Digital Forensics, Security and Law*, *10*(4). https://doi.org/10.15394/jdfsl.2015.1215

Amato, F., Cozzolino, G., Mazzeo, A., & Moscato, F. (2018). An application of semantic techniques for forensic analysis. *2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, 380–385. https://doi.org/10.1109/WAINA.2018.00115

Amato, F., Cozzolino, G., & Mazzocca, N. (2017). Correlation of Digital evidences in forensic investigation through semantic technologies. In *31st International Conference on Advanced Information Networking and Applications Workshops* (pp. 415–424). https://doi.org/10.1007/978-3-319-49109-7

Amro, A., Almuhammadi, S., & Zhioua, S. (2017). NetInfoMiner: High-level information extraction from network traffic. *2017 IEEE International Conference on Big Data and Smart Computing, BigComp 2017*, 143–150. https://doi.org/10.1109/BIGCOMP.2017.7881730

Arcila-Calderón, C., Barbosa-Caro, E., & Cabezuelo-Lorenzo, F. (2016). Técnicas big data: análisis de textos a gran escala para la investigación científica y periodística. *El Profesional de La Información*, *25*(4), 623. https://doi.org/10.3145/epi.2016.jul.12

Armknecht, F., & Dewald, A. (2015). Privacy-preserving email forensics. *Digital Investigation*, *14*(S1), S127–S136. https://doi.org/10.1016/j.diin.2015.05.003

Baader, F., & Nutt, W. (2003). Basic Description Logics. In U. ©2003 Cambridge University Press New York, NY (Ed.), *The description logic handbook* (pp. 43–95).

Banday, M. T. (2011). Techniques and Tools for Forensic Investigation of E- Mail. *International Journal of Network Security & Its Applications (IJNSA)*, *3*(6), 227–241. https://doi.org/10.5121/ijnsa.2011.3617

Barchini, G. E., & Álvarez, M. M. (2010). Dimensiones e indicadores de la calidad de una ontología Dimensions and indicators of the ontology quality. *Revista Avances En Sistemas e Informática*, *7*(1).

Bender, A. (2017). La prueba digital. Jurisprudencia y normas del Código Civil y Comercial de la Nación. *ElDial.Com Biblioteca Jurídica OnLine*.

Bhardwaj, A., & Goundar, S. (2017). Security challenges for cloud-based email infrastructure. *Network Security*, *2017*(11), 8–15. https://doi.org/10.1016/S1353-4858(17)30094-6

Biolchini, J., Mian, P. G., Natali, A. C. C., & Travassos, G. H. (2005). *Systematic Review in Software Engineering*. *TECHNICAL REPORT ES 679/05 - Systems Engineering and Computer Science Department COPPE / UFRJ* (Vol. 107). https://doi.org/10.1007/978-3-540-70621-2

Bjelland, P. C., Franke, K., & Årnes, A. (2014). Practical use of Approximate Hash Based Matching in digital investigations. *Digital Investigation*, *11*(SUPPL. 1), S18–S26. https://doi.org/10.1016/j.diin.2014.03.003

Blandón Andrade, J. C. (2018). A State-of-the-art Review About Ontology Population. *Ingeniería y Desarrollo*, *36*(1), 259–284. https://doi.org/10.14482/inde.36.1.10949

Brank, J., Grobelnik, M., & Mladenic, D. (2005). A survey of ontology evaluation techniques. *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005)*, 166–170.

Breitinger, F., & Baggili, I. (2017). File Detection on Network Traffic Using Approximate Matching. *Journal of Digital Forensics, Security and Law*, *9*(2). https://doi.org/10.15394/jdfsl.2014.1168

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, *80*(4), 571–583. https://doi.org/10.1016/j.jss.2006.07.009

Brewster, C., Alani, H., Dasmahapatra, S., & Wilks, Y. (n.d.). BrewsterLREC.pdf.

Brown, C. S. D. (2015). Investigating and prosecuting cyber crime: Forensic dependencies and barriers to justice. *International Journal of Cyber Criminology*, *9*(1), 55–119. https://doi.org/10.5281/zenodo.22387

Burton-Jones, A., Storey, V. C., Sugumaran, V., & Ahluwalia, P. (2005). A semiotic metrics suite for assessing the quality of ontologies. *Data and Knowledge Engineering*, *55*(1), 84–102. https://doi.org/10.1016/j.datak.2004.11.010

Cafferata Nores, J. I., & García, G. (2003). *La prueba en el proceso penal*. (LexisNexis, Ed.) (5a Edición). Buenos Aires: Depalma.

Carvalho, R., & Carvalho, R. (n.d.). CDT Technical Paper 06/14 Online Banking Malware Ontology Rodrigo Carvalho.

Carvalho, R., Goldsmith, M., & Creese, S. (2016). Applying Semantic Technologies to Fight Online Banking Fraud. *Proceedings - 2015 European Intelligence and Security Informatics Conference, EISIC 2015*, 61–68. https://doi.org/10.1109/EISIC.2015.42

Casey, E. (2018). Using computed similarity of distinctive digital traces to evaluate non-obvious links and repetitions in cyber-investigations, *24*, 2–9. https://doi.org/10.1016/j.diin.2018.01.002

Casey, E., Back, G., & Barnum, S. (2015). Leveraging CybOXtm to standardize representation and exchange of digital forensic information. *Digital Investigation*, *12*(S1), S102–S110. https://doi.org/10.1016/j.diin.2015.01.014

Ce, D., Parlamento, D. E. L., & Del, E. Y. (2008). DIRECTIVA 2002/58/CE DEL PARLAMENTO EUROPEO Y DEL CONSEJO sobre PROTECCIÓN DE DATOS, *39*, 1–22.

Chabot, Y., Bertaux, A., Nicolle, C., & Kechadi, T. (2015). An ontology-based approach for the reconstruction and analysis of digital incidents timelines. *Digital Investigation*, *15*, 83–100. https://doi.org/10.1016/j.diin.2015.07.005

Chen, L., & Mao, Y. (2017). Forensic analysis of email on android volatile memory. *Proceedings - 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 10th IEEE International Conference on Big Data Science and Engineering and 14th IEEE International Symposium on Parallel and Distributed Proce*, 945–951. https://doi.org/10.1109/TrustCom.2016.0160

Chen, Z., Yang, Y., Chen, L., Wen, L., Wang, J., Yang, G., & Guo, M. (2017). Email Visualization Correlation Analysis Forensics Research. *2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud)*, 339–343. https://doi.org/10.1109/CSCloud.2017.28

Chhabra, G. S., & Bajwa, D. S. (2012). Review of E-mail System, Security Protocols and Email Forensics. *International Journal of Computer Science & Communication Networks*, *5*(3), 201–211.

Choi, Y., Lee, J., Choi, S., Kim, J., & Kim, I. (2016). Traffic Storing and Related Information Generation System for Cyber Attack Analysis, 1052–1057.

Clarke, N., Li, F., & Furnell, S. (2017). A novel privacy preserving user identification approach for network traffic. *Computers and Security*, *70*, 335–350. https://doi.org/10.1016/j.cose.2017.06.012

Darahuge, M. E., & Arellano González, L. E. (2016). *Manual de Informática Forense III*.

de Reuver, M., & Haaker, T. (2009). Designing viable business models for context-aware mobile services. *Telematics and Informatics*, *26*(3), 240–248. https://doi.org/10.1016/j.tele.2008.11.002

Devendran, V. K., Shahriar, H., & Clincy, V. (2015). A Comparative Study of Email Forensic Tools. *Journal of Information Security*, *06*(02), 111–117. https://doi.org/10.4236/jis.2015.62012

Di Iorio, A. H., Castellote, A. M., Constanzo, B., Curti, H., Waimann, J., Alberdi, J. I., … Lamperti, S. (2017). *El rastro digital del delito: aspectos técnicos, legales y estratégicos de la informática forense*. *Universidad FASTA*. Mar del Plata: Esitorial UFASTA. Retrieved from http://info-lab.org.ar/images/pdf/Libro.pdf

Duman, S., Kalkan-Cakmakci, K., Egele, M., Robertson, W., & Kirda, E. (2016). EmailProfiler: Spearphishing Filtering with Header and Stylometric Features of Emails. *Proceedings - International Computer Software and Applications Conference*, *1*, 408–416. https://doi.org/10.1109/COMPSAC.2016.105

Ellison, D., Venter, H., & Adeyemi, ikuesan R. (2017). *An Improved Ontology for Knowledge Management in Security and Digital Forensic*.

Ellison, D., Venter, H., & Adeyemi, I. (2016). An Improved Ontology for Knowledge Management in Security and Digital Forensics, 2016.

Eshete, B., & Venkatakrishnan, V. N. (2017). DynaMiner: Leveraging Offline Infection Analytics for On-the-Wire Malware Detection. *Proceedings - 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2017*, 463–474. https://doi.org/10.1109/DSN.2017.54

Fensel, D., Harmelen, F. Van, Klein, M., Akkermans, H., Schnurr, H., Studer, R., … Davies, J. (2000). On-To-Knowledge : Ontology-based Tools for Knowledge Management. In *Proceedings of the eBusiness and eWork* (pp. 18–20).

Fernández, M., Gómez Pérez, A., & Juristo, N. (1997). METHONTOLOGY : From Ontological Art Towards Ontological Engineering. *AAAI Technical Report SS-97-06*, 33–40.

Fleurbaaij, D., Scanlon, M., & Le-Khac, N. A. (2017). Privileged data within digital evidence. *Proceedings - 16th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 11th IEEE International Conference on Big Data Science and Engineering and 14th IEEE International Conference on Embedded Software and Systems*, 737–744. https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.307

Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J., & Gil, R. (2005). Ontology Evaluation and Validation. An integrated Formal Model for the Quality Diagnostic Task. Technical Report. *Media*, *3*, 1–53. Retrieved from http://www.loa-cnr.it/Files/OntoEval4OntoDev_Final.pdf

Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *Digital Investigation*, *7*(SUPPL.), S64–S73. https://doi.org/10.1016/j.diin.2010.05.009

Ghasem, Z., Frommholz, I., & Maple, C. (2015a). A hybrid approach to combat email-based cyberstalking. *2015 4th International Conference on Future Generation Communication Technology, FGCT 2015*, (Fgct). https://doi.org/10.1109/FGCT.2015.7300257

Ghasem, Z., Frommholz, I., & Maple, C. (2015b). A machine learning framework to detect and document text-based cyberstalking. In *CEUR Workshop Proceedings* (Vol. 1458, pp. 348–355).

Gilardi, M., & Unzaga Domínguez, G. (2007). La Prueba Pericial en el Proceso Penal de la Provincia de Buenos Aires. *Revista Buenos Aires La Ley*, *Año 14 Núm*, 709.

Gomez-Perez, A. (1995). Some ideas and examples to evaluate ontologies. *Proceedings the 11th Conference on Artificial Intelligence for Applications, CAIA 1995*, 299–305. https://doi.org/10.1109/CAIA.1995.378808

Gómez-Pérez, A., Fernández López, M., & Corcho, O. (2004). *Ontological Engineering: with examples from the areas of knowledge management, ecommerce and the Semantic Web*. (Springer, Ed.). https://doi.org/10.4018/978-1-59904-045-5.ch003

Grant, M. J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, *26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Gray, G. L., & Debreceny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems*, *15*(4), 357–380. https://doi.org/10.1016/j.accinf.2014.05.006

Gruber, T. R. (1993a). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220. https://doi.org/10.1006/knac.1993.1008

Gruber, T. R. (1993b). Toward Principles for the Design of Ontologies Used for Knowledge Sharing, 907–928.

Gruninger, M., & Fox, M. S. (1995). Methodology for the Design and Evaluation of Ontologies 1 Introduction 2 Motivating Scenarios, 1–10.

Guarino, N. (1997). Understanding, Building, and Using Ontologies. *International Journal of Human-Computer Studies*, *46(2-3)*(May 1998), 293-310.

Guizzardi, G. (2005). *Foundations for Structural Conceptual*. *PhD thesis* (Vol. 015). Retrieved from http://doc.utwente.nl/50826

Gupta, A., Dasgupta, S., & Bagchi, A. (2017). PROFORMA: Proactive Forensics with Message Analytics. *IEEE Security and Privacy*, *15*(6), 33–41. https://doi.org/10.1109/MSP.2017.4251112

Gupta, S., Pilli, E. S., Mishra, P., Pundir, S., & Joshi, R. C. (2014). Forensic analysis of E-mail address spoofing. *Proceedings of the 5th International Conference on Confluence 2014: The Next Generation Information Technology Summit*, 898–904. https://doi.org/10.1109/CONFLUENCE.2014.6949302

Hadjidj, R., Debbabi, M., Lounis, H., Iqbal, F., Szporer, A., & Benredjem, D. (2009). Towards an integrated e-mail forensic analysis framework. *Digital Investigation*, *5*(3–4), 124–137. https://doi.org/10.1016/j.diin.2009.01.004

Harichandran, V. S., Breitinger, F., Baggili, I., & Marrington, A. (2016). A cyber forensics needs analysis survey: Revisiting the domain's needs a decade later. *Computers and Security*, *57*, 1–13. https://doi.org/10.1016/j.cose.2015.10.007

Horridge, M., Drummond, N., Goodwin, J., Rector, A., Stevens, R., & Wang, H. H. (2006). The manchester OWL syntax. *CEUR Workshop Proceedings*, *216*.

Iqbal, S., Kiah, M. L. M., Dhaghighi, B., Hussain, M., Khan, S., Khan, M. K., & Choo, K.-K. R. (2016). On cloud security attacks: A taxonomy and intrusion detection and prevention as a service. *Journal of Network and Computer Applications*, *74*, 98–120. https://doi.org/https://doi.org/10.1016/j.jnca.2016.08.016

Iyer, R. P., Atrey, P. K., Varshney, G., & Misra, M. (2017). Email spoofing detection using volatile memory forensics. *2017 IEEE Conference on Communications and Network Security, CNS 2017*, *2017-Janua*, 619–625. https://doi.org/10.1109/CNS.2017.8228692

Jahanirad, M., Wahab, A. W. A., & Anuar, N. B. (2016). An evolution of image source camera attribution approaches. *Forensic Science International*, *262*, 242–275. https://doi.org/10.1016/j.forsciint.2016.03.035

Jayan, A., & Dija, S. (2015). Detection of spoofed mails. In *2015 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2015* (pp. 1–4). https://doi.org/10.1109/ICCIC.2015.7435764

Jeong, D., Kang, H., & Lee, S. (2017). Towards Syntactic Approximate Matching - A Pre-Processing Experiment. *Journal of Digital Forensics, Security and Law*, *11*(2). https://doi.org/10.15394/jdfsl.2016.1381

Jo, S., Kim, J., & Choi, D. (2015). The study of document filter for smart device. *17th Asia-Pacific*

Network Operations and Management Symposium: Managing a Very Connected World, APNOMS 2015, 515–518. https://doi.org/10.1109/APNOMS.2015.7275388

Jusas, V., Birvinskas, D., & Gahramanov, E. (2017). Methods and tools of digital triage in forensic context: Survey and future directions. *Symmetry*, *9*(4). https://doi.org/10.3390/sym9040049

Kalemi, E., & Yildirim-Yayilgan, S. (2016). Ontologies for Social Media Digital Evidence. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *10*(2), 335–340. Retrieved from http://waset.org/Publications?p=110

Karie, N. M., & Venter, H. S. (2014). Toward a general ontology for digital forensic disciplines. *Journal of Forensic Sciences*, *59*(5), 1231–1241. https://doi.org/10.1111/1556-4029.12511

Kaur, R., & Singh, M. (2014). A survey on zero-day polymorphic worm detection techniques. *IEEE Communications Surveys and Tutorials*, *16*(3), 1520–1549. https://doi.org/10.1109/SURV.2014.022714.00160

Khan, R., Mizan, M., Hasan, R., & Sprague, A. (2017). Hot Zone Identification: Analyzing Effects of Data Sampling On Spam Clustering. *Journal of Digital Forensics, Security and Law*, (c). https://doi.org/10.15394/jdfsl.2014.1164

Kota, V. K. (2012). An Ontological Approach for Digital Evidence Search. *International Journal of Scientific and Research Publications*, *2*(12), 409–414. https://doi.org/10.1.1.642.3055

Koven, J., Bertini, E., Dubois, L., & Memon, N. (2016). InVEST: Intelligent visual email search and triage. *Digital Investigation*, *18*, S138–S148. https://doi.org/10.1016/j.diin.2016.04.008

Lenat, D. B., Guha, R. V., Pittman, K., Prat, D., & Shepherd, M. (1990). OF THE ACM CYC : TOWARD WITH COMMON SENSE. *Comunications of ACM*, *33*.

Lozano Tello, A. (2002). *Métrica de Idoneidad de Ontologías*.

Maake, L. M., Kebande, V. R., & Karie, N. M. (2017). Onto-Engineering : A Conceptual framework for Integrating Requirement Engineering Process with scientifically tuned Digital Forensics Ontologies, (June). https://doi.org/10.17781/P002271

Matic, S., Kotzias, P., & Caballero, J. (2015). CARONTE: Detecting Location Leaks for Deanonymizing Tor Hidden Services. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS '15*, 1455–1466. https://doi.org/10.1145/2810103.2813667

Mavroeidis, V. (2018). A Framework for Data-Driven Physical Security and Insider Threat Detection. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1108–1115. https://doi.org/10.1109/ASONAM.2018.8508599

Mazurczyk, W., & Caviglione, L. (2015). Steganography in Modern Smartphones and Mitigation Techniques. *IEEE Communications Surveys and Tutorials*. https://doi.org/10.1109/COMST.2014.2350994

Medina Lopez, C., Marin Garcia, J. A., & Alfalla Luque, R. (2010). Una propuesta metodológica para la realización de búsquedas sistemáticas de bibliografía. *WPOM-Working Papers on Operations Management*, *1*(2), 13–30. https://doi.org/10.4995/wpom.v1i2.786

Mehta, R. (2017). SEMANTIC E-MAIL ADDRESSING USING.

Mohammed, H., Clarke, N., & Li, F. (2017). An Automated Approach for Digital Forensic Analysis of Heterogeneous Big Data. *Journal of Digital Forensics, Security and Law*, *11*(2). https://doi.org/10.15394/jdfsl.2016.1384

Morgan, R. M. (2017a). Science and Justice Conceptualising forensic science and forensic reconstruction . Part I : A conceptual model. *Science & Justice*, *57*(6), 455–459. https://doi.org/10.1016/j.scijus.2017.06.002

Morgan, R. M. (2017b). Science and Justice Conceptualising forensic science and forensic reconstruction . Part II : The critical interaction between research , policy / law and practice. *Science & Justice*, *57*(6), 460–467. https://doi.org/10.1016/j.scijus.2017.06.003

Msongaleli, D. L. (2018). Electronic Mail Forensic Algorithm for Crime Investigation and Dispute Settlement. *In Digital Forensic and Security (ISDFS), 2018 6th International Symposium*, 1–5.

Musen, M. (2015). The Protege Project: A Look Back and a Look Forward. *AI Matters*, *1*(4), 4–12. https://doi.org/10.1126/science.1249098.Sleep

Nampoothiri, A. P. (2015). Email Forensic Analysis Based on k- means clustering, 814–817. https://doi.org/10.1109/ICACCI.2015.7275710

Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. R. (1991). Enabling Technology for Knowledge Sharing. *AI Magazine*, *12*(3).

Nikkel, B. (2017). Registration Data Access Protocol (RDAP) for digital forensic investigators. *Digital Investigation*, *22*, 133–141. https://doi.org/10.1016/j.diin.2017.07.002

Nurse, J., Erola, A., Goldsmith, M., & Creese, S. (2015). Investigating the leakage of sensitive personal and organisational information in email headers. *Journal of Internet Services and Information Security*, *5*(1), 70–84.

Obrst, L., Ceusters, W., Mani, I., Ray, S., & Smith, B. (2007). The evaluation of ontologies toward improved semantic interoperability. *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, *9780387484*, 139–158. https://doi.org/10.1007/978-0-387-48438-9_8

Ouyang, L., Zou, B., Qu, M., & Zhang, C. (2011). A method of ontology evaluation based on coverage, cohesion and coupling. *Proceedings - 2011 8th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2011*, *4*, 2451–2455. https://doi.org/10.1109/FSKD.2011.6020046

Ovens, K. M., & Morison, G. (2016). Identification and analysis of email and contacts artefacts on iOS and OSX. *Proceedings - 2016 11th International Conference on Availability, Reliability and Security, ARES 2016*, (October 2017), 321–327. https://doi.org/10.1109/ARES.2016.56

Park, J. (2018). TREDE and VMPOP: Cultivating multi-purpose datasets for digital forensics – A Windows registry corpus as an example. *Digital Investigation*, *26*, 3–18. https://doi.org/10.1016/j.diin.2018.04.025

Patil, D., & Meshram, B. (2018). Network Packet Analysis for Detecting Malicious Insider. *2018 3rd International Conference for Convergence in Technology, I2CT 2018*, 1–8. https://doi.org/10.1109/I2CT.2018.8529451

Pieterse, H., Olivier, M. S., & Van Heerden, R. P. (2016). Reference architecture for android applications to support the detection of manipulated evidence. *SAIEE Africa Research Journal*, *107*(2), 92–103.

Pieterse, Heloise, Olivier, M. S., & Van Heerden, R. P. (2015). Playing hide-and-seek: Detecting the manipulation of Android Timestamps. *2015 Information Security for South Africa - Proceedings of the ISSA 2015 Conference*. https://doi.org/10.1109/ISSA.2015.7335065

Pluskal, J., Matoušek, P., Ryšavý, O., Kmeť, M., Veselý, V., & Karpíšek, F. (n.d.). Netfox Detective : A Tool for Advanced Network Forensics Analysis, (i).

Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, *97*, 205–227. https://doi.org/10.1016/j.eswa.2017.12.020

Porzel, R., & Malaka, R. (2004). A Task-based Approach for Ontology Evaluation. *ECAI Workshop on Ontology Learning and Population*.

Poveda Villalón, M. (2016). Ontology Evaluation: a pitfall-based approach to ontology diagnosis, 236. https://doi.org/10.20868/UPM.thesis.39448

Ramisch, F., & Rieger, M. (2015). Recovery of SQLite Data Using Expired Indexes. *Proceedings - 9th International Conference on IT Security Incident Management and IT Forensics, IMF 2015*, 19–25. https://doi.org/10.1109/IMF.2015.11

Ramos, E., Nuñez, H., & Casañas, R. (2009). Esquema para evaluar ontologías únicas para un dominio de conocimiento. *Enlace*, *6*(1), 1–11.

Riaz, H., & Tahir, M. A. (2018). Analysis of VMware virtual machine in forensics and anti-forensics paradigm. *6th International Symposium on Digital Forensic and Security, ISDFS 2018 - Proceeding*, *2018-Janua*(Isdfs), 1–6. https://doi.org/10.1109/ISDFS.2018.8355375

Rivetti, E., Araoz Fleming, J., Parra de Gallo, B., & Leone, H. (2016). Análisis de los Documentos Oficiales sobre Obtención , Tratamiento y Preservación de la Evidencia Digital Aportes para el Tratamiento del Correo Electrónico como Evidencia Digital. In *CONAIISI 2016*.

Rivetti, E., & Parra de Gallo, B. (2017). Estudio comparativo de desempeño de herramientas para el Análisis Forense de Correos Electrónicos. In *CONAIISI 2017* (pp. 46–51).

Romaios, B., Nikolaos, K., George, K., & Andreas, A. (2016). Email forensic tools: A roadmap to email header analysis through a cybercrime use case. *Journal of Polish Safety and Reliability Association Summer Safety and Reliability Seminars*, *7*(1), 21–28.

Rudd, E. M., Harang, R., & Saxe, J. (2018). MEADE: Towards a Malicious Email Attachment Detection Engine. *2018 IEEE International Symposium on Technologies for Homeland Security, HST 2018*, 1–7. https://doi.org/10.1109/THS.2018.8574202

Samuel, S., Graham, J., & Hinds, C. (2018). Hunting Malware: An Example Using Gh0st. *Proceedings - 2017 International Conference on Computational Science and Computational Intelligence, CSCI 2017*, 97–102. https://doi.org/10.1109/CSCI.2017.16

Scanlon, M., Farina, J., & Kechadi, M. T. (2015). Network investigation methodology for BitTorrent Sync: A Peer-to-Peer based file synchronisation service. *Computers and Security*, *54*, 27–43. https://doi.org/10.1016/j.cose.2015.05.003

Schmid, M. R., Iqbal, F., & Fung, B. C. M. (2015). E-mail authorship attribution using customized associative classification. *Digital Investigation*, *14*(S1), S116–S126. https://doi.org/10.1016/j.diin.2015.05.012

Schreiber, G., Amsterdam, V. U., Wielinga, B. J., & Jansweijer, W. N. H. (1995). The KACTUS View on the ' O ' Word. *IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing*, 159–168.

Selamat, S. R., Shahrin, S., Hafeizah, N., Yusof, R., & Abdollah, M. F. (2013). A Forensic Traceability Index in Digital Forensic Investigation. *Journal of Information Security*, *04*(01), 19–32. https://doi.org/10.4236/jis.2013.41004

Senthivel, S., Ahmed, I., & Roussev, V. (2017). SCADA network forensics of the PCCC protocol. *Digital Investigation*, *22*, S57–S65. https://doi.org/10.1016/j.diin.2017.06.012

Settanni, G., Skopik, F., Shovgenya, Y., Fiedler, R., Carolan, M., Conroy, D., … Olli, P. (2017). A collaborative cyber incident management system for European interconnected critical infrastructures. *Journal of Information Security and Applications*, *34*, 166–182. https://doi.org/https://doi.org/10.1016/j.jisa.2016.05.005

Shashidhar, K. K., & Manjaiah, D. H. (2014). Forensic Investigation Processes for Cyber Crime and Cyber Space. *Advances in Intelligent Systems and Computing*, *216*, 169–178. https://doi.org/10.1007/978-81-322-1299-7

Souali, K., Rahmaoui, O., & Ouzzif, M. (2017). An overview of traceability: Definitions and techniques. *Colloquium in Information Science and Technology, CIST*, 789–793. https://doi.org/10.1109/CIST.2016.7804995

Stadlinger, J., & Dewald, A. (2017). A Forensic Email Analysis Tool Using Dynamic Visualization. *The Journal of Digital Forensics, Security and Law (JDFSL)*, *12*(1). https://doi.org/https://doi.org/10.15394/jdfsl.2017.1413

Studer, R., Benjamins, V. R., & Fensel, D. (1998). KNOWLEDGE ENGINEERING: Principles and Methods. *Data & Knowledge Engineering*, *25*, 161–197. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.8406&rep=rep1&type=pdf

Suárez-figueroa, M. C., Gómez-pérez, A., & Villazón-terrazas, B. (2009). How to Write and Use the Ontology Requirements Specification Document, 966–982.

Suárez-Figueroa, M. del C. (2010). *NeOn Methodology for Building Ontology Networks : Specification , Scheduling and Reuse*. Universidad Politécnica de Madrid (España).

Subedi, K. P., Budhathoki, D. R., & Dasgupta, D. (2018). Forensic analysis of ransomware families using static and dynamic analysis. *Proceedings - 2018 IEEE Symposium on Security and Privacy Workshops, SPW 2018*, (June), 180–185. https://doi.org/10.1109/SPW.2018.00033

Swartout, B., Knight, K., Russ, T., & Rey, M. (1997). t Toward Distributed Use of Large-Scale Ontologies. *AAAI Technical Report SS-97-06*, 138–148.

Szulman, S., & Biébow, B. (2002). Structuration de terminologies à l'aide d'outils de TAL avec TERMINAE. *Revue Traitement Automatique Des Langues*, *43*, 103–128.

Tanwar, G. S., & Poonia, A. S. (2014). Live forensics analysis: Violations of business security policy. *Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014*, 971–976. https://doi.org/10.1109/IC3I.2014.7019695

Teing, Y.-Y., Ali, D., Choo, K., Abdullah, M. T., & Muda, Z. (2017). Greening Cloud-Enabled Big Data Storage Forensics: Syncany as a Case Study. *IEEE Transactions on Sustainable Computing*, 1–1. https://doi.org/10.1109/tsusc.2017.2687103

Teing, Y. Y., Dehghantanha, A., Choo, K. K. R., & Yang, L. T. (2017). Forensic investigation of P2P cloud storage services and backbone for IoT networks: BitTorrent Sync as a case study. *Computers and Electrical Engineering*, *58*(2016), 350–363. https://doi.org/10.1016/j.compeleceng.2016.08.020

Turner, R. C. (2017). Proposed Model for Natural Language ABAC Authoring, 61–72.

Ullah, F., Edwards, M., Ramdhany, R., Chitchyan, R., Babar, M. A., & Rashid, A. (2018). Data exfiltration: A review of external attack vectors and countermeasures. *Journal of Network and Computer Applications*, *101*, 18–54. https://doi.org/10.1016/j.jnca.2017.10.016

Umar, R., Riadi, I., & Muthohirin, B. F. (2018). Acquisition of Email Service Based Android Using NIST. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, *3*(3), 263. https://doi.org/10.22219/kinetik.v3i4.637

Uschold, M., & King, M. (1995). Towards a Methodology for Building Ontologies conjunction with IJCAI-95 Abstract, (July).

Vale, T., de Almeida, E. S., Alves, V., Kulesza, U., Niu, N., & de Lima, R. (2017). Software product lines traceability: A systematic mapping study. *Information and Software Technology*, *84*, 1–18. https://doi.org/10.1016/j.infsof.2016.12.004

Varma, S., Walls, R. J., Lynn, B., & Levine, B. N. (2014). Efficient Smart Phone Forensics Based on Relevance Feedback, 81–91. https://doi.org/10.1145/2666620.2666628

Velásquez, J. D. (2015). Una Guía Corta para Escribir Revisiones Sistemáticas de Literatura Parte 3. *Dyna*, *82*(189), 9–12. https://doi.org/10.15446/dyna.v82n189.48931

Wimmer, H., Chen, L., Narock, T., & Chen, L. (2018). Ontologies and the Semantic Web for Digital Investigation Tool Selection. *The Journal of Digital Forensics, Security and Law (JDFSL)*, *13*(3), 21–46.

Wu, Y., Ye, D., Wei, Z., Wang, Q., Tan, W., & Deng, R. H. (2018). Situation-aware Authenticated Video Broadcasting over Train-trackside WiFi Networks. *IEEE Internet of Things Journal*. https://doi.org/10.1109/JIOT.2018.2859185

Xie, L., Liu, Y., & Chen, G. (2016). A forensic analysis solution of the email network based on email contents. *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015*, 1613–1619. https://doi.org/10.1109/FSKD.2015.7382186

Yang, T. Y., Dehghantanha, A., Choo, K. K. R., & Muda, Z. (2016). Windows Instant Messaging App Forensics: Facebook and Skype as Case Studies. *PloS One*, *11*(3), e0150300. https://doi.org/10.1371/journal.pone.0150300

Youn, S. (2014). SPONGY (SPam ONtoloGY): Email classification using two-level dynamic ontology. *Scientific World Journal*, *2014*. https://doi.org/10.1155/2014/414583

Yu, J. (2008). *Requirements-Oriented Methodology for Evaluating Ontologies Doctor of Philosophy*. *School of Computer Science and Information Technology*,. RMIT University, Australia.

Yu, S. (2015). Covert communication by means of email spam: A challenge for digital investigation. *Digital Investigation*, *13*, 72–79. https://doi.org/10.1016/j.diin.2015.04.003

Zhang, X., Baggili, I., & Breitinger, F. (2017). Breaking into the vault: Privacy, security and forensic analysis of Android vault applications. *Computers and Security*, *70*, 516–531.

https://doi.org/10.1016/j.cose.2017.07.011

Zhang, Y., Liu, Y., & Chen, G. (2015). A Solution of Anonymous Email Identification Based on Writing Structural Pattern, 1559–1565.

Zuccardi, G., Gutiérrez, J. D., Cano, J. J., Plazas, A. R., Cano, J. J., Cano Martinez, J. J., … Felipe, C. J. A. (2006). Evidencia Digital: Contexto, Situación, e Implicaciones Nacionales, (Evidencia Digital), 1–17.