risti
Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

Recebido/Submission: 16/02/2019
Aceitação/Acceptance: 20/04/2019

OntoFoCE: Herramienta para el Análisis Forense de Correos Electrónicos

Enzo Notario[1], Beatriz Parra de Gallo[1], Marcela Vegetti[2], Horacio Leone[2]

ernotario@ucasal.edu.ar,bgallo@ucasal.edu.ar, mvegetti@santafe-conicet.gov.ar, hleone@santafe-conicet.gov.ar

[1] IEsIIng – Facultad de Ingeniería, Universidad Católica de Salta, Salta, Argentina,

[2] INGAR – Instituto de Desarrollo y Diseño (Conicet/UTN), Santa Fe, Santa Fe, Argentina

DOI: 10.17013/risti.32.17–32

Resumen: Este trabajo tiene como objetivo presentar una herramienta para el análisis forense de correos electrónicos a partir de la cabecera de los mismos y la utiliza para instanciar una ontología definida para responder a los puntos de pericia solicitados sobre el correo electrónico. La herramienta consta de cuatro componentes que permiten la obtención de las cabeceras de los correos a peritar, la instanciación de la ontología con los datos obtenidos de las cabeceras y la obtención de respuestas a los puntos de pericia a partir de las preguntas de competencia definidas para la ontología. Se describe cada componente y se ejemplifica el uso de la herramienta mediante un caso de estudio sobre análisis forense de un correo electrónico.

Palabras-clave: Ontología; cabecera de correo; forensia digital.

OntoFoCE: A Framework for the Forensic Analysis of Electronic Mail based on an ontology

***Abstract*:** The purpose of this work is to present a tool for the forensic analysis of emails from their headers. The tool is used to generate a defined ontology, in order to respond to the points of expertise requested about the email. It consists of four components which allow us to obtain the headers of the mails to be analyzed, to generate the ontology with the data acquired from those headers, and to give an answer to the points of expertise produced by the competency questions defined for the ontology. Each component is described and the use of the tool is exemplified through a case study on forensic analysis of an email.

Keywords: Ontology; mail header; digital forensia

1. Introducción

El análisis forense de correos electrónicos se realiza a partir de los datos que residen en su *cabecera*. Allí constan las cuentas, direcciones IP, fechas y demás elementos que permiten validar un correo electrónico cuando es propuesto como prueba en un juicio.

Un correo electrónico se compone de una cabecera y un cuerpo presentados según el formato definido por la norma RFC 822[1]. Los datos de la cabecera necesarios para el análisis forense deben ser identificados y tratados individualmente durante la pericia.

La pericia de correos electrónicos puede suponer dos escenarios de trabajo distintos: a) el análisis de un correo en particular, cuando se trata de un único documento; y b) el análisis de un conjunto de correos que deben considerarse en su totalidad. Ambos casos suponen métodos y herramientas diferentes. En el primero, los datos se obtienen directamente de la revisión visual de la cabecera, pero en el segundo caso, se requiere procesos automáticos de extracción por el volumen de datos a considerar.

Teniendo en cuenta esta necesidad de extraer de manera automática datos de la cabecera de varios correos, este trabajo propone el framework OntoFoCE. Este framework, que tiene como componente principal una ontología, asiste al perito informático en el análisis forense de un conjunto de correos electrónicos, brindando soporte desde la preparación de los datos e instanciación de la ontología mediante un proceso automático, hasta la obtención de las respuestas a los puntos de pericia.

Estos trabajos han contribuido a la investigación en curso con un aporte particular en cada caso. Los trabajos (Flores & Hadfeg, 2017), (Tolaba, Caliusco, & Galli, 2015) y (Sarli, Leone, & Gutierrez, 2019) muestran distintos enfoques de aplicación de las ontologías útiles para definir cómo aplicar las ontología en el análisis forense. Los trabajos (Kota, 2012), (Brady, Overill, & Keppens, 2015), (Ovens & Morison, 2016), (Xie, Liu, & Chen, 2016) y (Stadlinger & Dewald, 2017) ayudaron en la definición del procesamiento interno de los datos, atendiendo al volumen y diversidad de formatos que se utilizan en la forensia de correos electrónicos. Las propuestas de (Rueda-rueda, Rico-bautista, & Guerrero, 2018) y de (Di Ioro et al., 2017) se tuvieron en cuenta para definir el método más adecuado para el análisis forense de correos electrónicos. Los criterios de calidad para que las herramientas forenses permitan la reconstrucción de la prueba digital, bajo cánones de reproducibilidad, integridad y credibilidad propuestos por (Chabot, Bertaux, Nicolle, & Kechadi, 2015) y los criterios de evaluación de herramientas propuestos por (David, Parra, Rico-bautista, Medina-cárdenas, & Sanchez-ortiz, 2018) resultaron de interés para aplicar a la herramienta de soporte aquí presentada. El trabajo (Chhabra & Bajwa, 2012) que revisa integralmente el tema del análisis forense de correos electrónicos, se tomó como estructura formal para la descripción ordenada del objeto de estudio, mientras que los trabajos (Devendran, Shahriar, & Clincy, 2015) y (Youn, 2014) permitieron ajustar la ontología al modelo requerido para representar la trazabilidad y las preguntas de competencia. Los trabajos sobre trazabilidad de (Selamat, Shahrin, Hafeizah, Yusof, & Abdollah, 2013), (Morgan, 2017a) y (Morgan, 2017b), (Ya'u, Nordin, & Salleh, 2017) y (Noll & Ribeiro, 2007) sustentan la aplicación de la trazabilidad como camino válido para modelar el proceso de transmisión del correo electrónico, y particularmente el trabajo de (Yubao, 2015) en el que propone el uso de la trazabilidad como componente de confianza para generar la evidencia digital. Los aportes de (Blandón Andrade, 2018), (Arcila-Calderón, Barbosa-

[1] Se puede consultar en: https://www.ietf.org/rfc/rfc822.txt

Caro, & Cabezuelo-Lorenzo, 2016) y (Faria, Serra, & Girardi, 2014) sirvieron como punto de inicio en el proceso de instanciación incluido en OntoFoCE.

La estructura del trabajo es la siguiente: la sección 2 describe el proceso de transmisión y trazabilidad del correo electrónico y su representación mediante una ontología. La sección 3 describe OntoFoCE mientras que en la sección 4 se muestra un ejemplo de aplicación de la misma. En la sección 5 se compara OntoFoCE con otras herramientas forenses de correos electrónicos y. la sección 6 describe las conclusiones del trabajo.

2. Representación ontológica de la trazabilidad del proceso de transmisión de un correo electrónico

La pericia de correos electrónicos se efectúa sobre la cabecera del mismo, pues allí están los datos de la emisión, transmisión y recepción del mismo y siempre sobre *cabeceras de correos recibidos*, ya que los correos emitidos solo contienen datos referidos al envío del mismo, mientras que las cabeceras de correos recibidos tienen todos los datos requeridos para establecer la trazabilidad de la transmisión.

Considerando el proceso de transmisión de un correo electrónico desde su emisión hasta la recepción, ocurren diferentes procesos que se van ejecutando. De éstos, interesan en particular aquellos que pueden impactar en la *modificación del paquete de datos* que circula. Un correo electrónico es manejado por un mínimo de 4 equipos: el emisor, el servidor de correo del remitente, el servidor de correo del receptor y el equipo receptor. En todos ellos, se añade a la *cabecera del correo* una etiqueta de identificación cada vez que pasa por un servidor, y finalmente se agregan los datos del equipo y cuenta receptora una vez que el correo electrónico llega a su destino.

Así, se puede establecer la *trazabilidad* de un correo electrónico a partir de la cabecera. La trazabilidad permite conocer con certeza la procedencia y la historia de un objeto. La necesidad de reconstruir el camino inverso de un correo electrónico recibido se sustenta en que así se puede probar su existencia, avalando el carácter probatorio de esa evidencia.

Durante la transmisión el correo va residiendo en diferentes dispositivos, así, es posible tomar cada archivo almacenado en éstos y verificar si el paquete de datos *original* fue modificado en algún punto de la transmisión. La comprobación de la inalterabilidad del correo se realiza probando que éste es el mismo en cada equipo en que se almacena durante la transmisión. Esto es en cuanto al *cuerpo del mensaje*¸ que debería estar sin cambios mientras la *cabecera* se va extendiendo a medida que va pasando por los dispositivos durante la transmisión. Así, se puede establecer la trazabilidad de la transmisión, a partir de estos criterios: a) el correo se representa mediante *ocurrencias*, tantas como veces se almacena éste en los distintos dispositivos durante la transmisión. Una ocurrencia es una copia exacta del correo (en cuanto al contenido), con el agregado en la cabecera de la identificación del dispositivo actuante; b) existen tres tipos de ocurrencias: de emisión, de transmisión y de recepción; mientras que las ocurrencias de emisión y recepción son únicas, las de transmisión serán tantas como servidores intervienen en el proceso de envío; d) las ocurrencias se asocian en hilos que señalan el recorrido del correo desde el equipo emisor hasta el equipo receptor, así, habrá tantos hilos como receptores tenga el correo; y e) una secuencia relaciona los distintos hilos que

conforman el correo. La Figura 1 ejemplifica las ocurrencias, hilos y secuencia del envío de un correo desde una cuenta dirigida a dos remitentes.

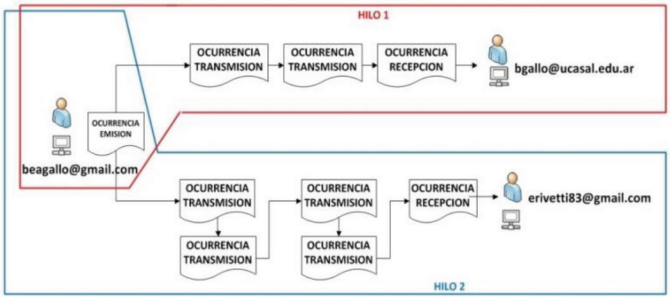

Figura 1 – Ocurrencias del correo electrónico

El envío del correo se realiza en una *SECUENCIA* que contiene dos *HILOS*. Existe una única *OCURRENCIA_DE_EMISION,* correspondiente a la cuenta beagallo@gmail .com, que a través del *HILO_1*, contiene dos *OCURRENCIAS_ DE_TRANSMISION*, que llegan a la *OCURRENCIA_DE_RECEPCIÓN* correspondiente a la cuenta receptora bgallo@ucasal.edu.ar. De igual modo, desde esa única *OCURRENCIA_DE_ EMISIÓN* se genera el *HILO_2* con 4 *OCURRENCIAS_DE_TRANSMISIÓN* hasta llegar a la *OCURRENCIA_DE_RECEPCION* de la cuenta erivetti@gmail.com.

La ontología definida en (Gallo & Leone, 2016) representa el correo electrónico y la trazabilidad del proceso de transmisión, y responde a los puntos de pericia a partir de las preguntas de competencia. Durante la especificación de la ontología se conformó un banco de puntos de pericias, expresados en lenguaje natural, del cual se extrajeron las preguntas de competencia. Para esto, se encuestó a peritos informáticos sobre pericias realizadas sobre correos electrónicos, obteniéndose 151 puntos de pericia que se clasificaron, ordenaron y asociaron por similitud de datos requeridos, y se usaron para definir las 21 preguntas de competencia que responde OntoFoCE:

1. ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?
2. ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?
3. ¿Cuántos receptores tiene el correo?
4. ¿Cuál es el nombre de usuario y dirección de e-mail del Emisor?
5. ¿Cuál es el nombre de usuario y dirección de e-mail del Receptor?
6. ¿Cuál fue el cliente de correo utilizado por cada usuario?
7. ¿Cuál fue el equipo desde el cual se emitió el correo?
8. ¿Cuál fue el equipo en el que se recibió el correo?
9. Dado una cuenta ¿cuáles son los correos que emitió?
10. Dado una cuenta ¿cuáles son los correos que recibió?
11. Dado un correo, un emisor y un receptor ¿cuáles son los servidores de paso del correo?
12. Dado una cuenta C1 ¿se ha emitido un correo hacia la cuenta C2?
13. Dado una cuenta C1 ¿se ha recibido un correo desde la cuenta C2?
14. Dada una dirección IP ¿cuál sería la localización geográfica del mismo?

15. ¿Cuáles son los correos que han pasado por el dispositivo que posee una IP dada?
16. ¿Cuáles son los mails enviados desde una determinada cuenta en una fecha dada?
17. ¿Cuáles son los mails recibidos por una determinada cuenta en una fecha dada?
18. Dada una palabra clave ¿Figura en el asunto de un correo?
19. Dada una palabra clave ¿Figura en el cuerpo de un correo?
20. Dada una palabra clave ¿Figura en el adjunto de un correo?
21. ¿Cuáles son los correos intercambiados entre las cuentas C1 y C2 en un rango de fechas dado?

A partir de las preguntas de competencia, se derivaron los conceptos, objetos, sus clases y relaciones, así como los axiomas que permiten modelar el correo electrónico desde varios enfoques: la estructura interna, la trazabilidad del proceso de transmisión, y los criterios de validez del documento digital como prueba no repudiable.

A continuación, se detallan algunos aspectos de la ontología que interesan para el presente trabajo. La taxonomía define los conceptos necesarios para representar la trazabilidad del correo electrónico sobre la cual se basa el análisis pericial.

Correo representa el correo electrónico que es objeto de la pericia, e incluye la subclase *CorreoValido* que representa un correo que cumple con los criterios de mínima para que pueda ser analizado por la herramienta. *CabeceraDeCorreo* representa la parte del correo que se toma para el análisis. *Cuenta* es la clase que representa las cuentas de correo que intervienen en el envío/recepción del mismo. Se descompone en dos subclases: *CuentaEmisor* que representa la cuenta desde la cual se emite el correo, y *CuentaReceptor* que representa la cuenta destinataria. *Ocurrencia* representa la copia del correo que reside en los dispositivos utilizados durante el proceso de transmisión. Esta clase se descompone en tres subclases: *OcurrenciaDeEmision, OcurrenciaDe Transmision y OcurrenciaDeRecepcion* que representan respectivamente la copia del correo emitido, las copias generadas durante la transmisión y la copia del correo recibido. *Hilo* representa la asociación de las ocurrencias relacionadas a cada cuenta receptora, o sea, vincula una *OcurrenciaDeEmisión* con una o más *OcurrenciaDe Transmisión* y una *OcurrenciaDeRecepcion*; existiendo tantas instancias de *Hilo* como cuentas receptoras haya. *Secuencia* representa el conjunto de *Hilos* que se generan para las cuentas receptoras del correo. También se han definido otras clases para representar los conceptos necesarios para que la ontología modele cabalmente el análisis forense: *Equipo; IdentificaciónEquipo (*con las subclases *IP y HostName), ClienteCorreo, Adjunto, Asunto, CuerpoCorreo, Expediente y PalabraClave*.

La clase *Correo* se asocia a las clases *CuentaEmisor* y *CuentaReceptor* mediante las relaciones de *cuentaEmisorEmiteCorreo* y *cuentaReceptorRecibeCorreo* respectivamente. La trazabilidad del proceso de transmisión se representa con las relaciones que vincula las clases *Secuencia, Hilo y Ocurrencias*. En particular, interesa las relaciones *esAnteriorA* que representan el orden de aparición de las ocurrencias en cada *Hilo*.

La ontología propuesta incluye además un axioma que sustenta la validez del análisis forense de correos electrónicos, referido a la *existencia* de un correo electrónico, que

se puede verificar mediante 3 elementos: a) los datos del remitente (cuenta de correo y dirección IP); b) la trazabilidad del mismo (datos de los dispositivos que intervienen en la transmisión); y c) los datos del destinatario (cuenta de correo y dirección IP).

En (Gallo & Leone, 2016), (Gallo, Vegetti, & Leone, 2015) y (Gallo, Vegetti, & Leone, 2017) se pueden encontrar más detalles acerca de la ontología propuesta.

La implementación de la ontología es el componente destacable del framework *OntoFoCE,* que toma como entrada las cabeceras de los correos a analizar, instancia los datos en la ontología y genera como resultado las respuestas a las preguntas de competencia, que a su vez, permiten responder a los puntos de pericia solicitados.

3. Framework para el Análisis Forense de Correos Electrónicos

Si el análisis forense debe hacerse sobre un único correo, es sencillo revisar visualmente la cabecera para responder a los puntos de pericia. Cuando se trata de un conjunto de correos, el análisis se complica por el volumen de datos que se deben trabajar, a lo que se suma la dificultad para obtener las cabeceras de correos mediante procesos automáticos. Si bien hay herramientas para forensia de correos que resuelven en parte los problemas derivados del volumen de datos, usualmente no cuentan con procesos automáticos para extraerlos. Y principalmente, no responden de manera directa a los puntos de pericia. Por ello, se propone un framework que permita obtener la cabecera del correo de manera automática, instancie los datos en la ontología y permita responder los puntos de pericia. El Framework OntoFoCE está integrado por cuatro componentes (Figura 2): Extractor de Cabeceras, Gestor de Instancias de la Ontología, Analizador de Puntos de Pericia y la ontología, que trabajan sobre un servicio SPARQL Endpoint que las consultas sobre la ontología.

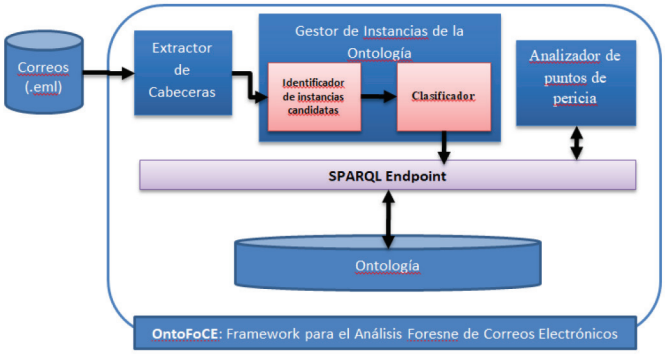

Figura 2 – OntoFoCE Framework para el Análisis Forense de Correos Electrónicos

En la sección 2 se describió la ontología que sirve de base para la herramienta. Seguidamente, se describen los objetivos y funcionalidad de los restantes componentes.

3.1. Extractor de las Cabeceras de Correos

Este componente obtiene las cabeceras de los correos, mediante un proceso automático que evite una tarea manual engorrosa y sujeta a errores. Para ello se accede al cliente de correo de la cuenta y se exportan los correos en un archivo con formato de texto plano. Si el cliente de correo no cuente con funciones de exportación, se debe recurrir a un cliente de correo multiplataforma, con capacidad de importación/exportación de cuentas de otros clientes de correo, para obtener los correos en formato de texto plano.

3.2. Gestor de Instancias de la Ontología

Este proceso toma el archivo en texto plano de la cabecera del correo y descompone los datos en elementos mínimos adecuados para la instanciación de los conceptos y relaciones de la ontología. Este proceso incluye los pasos para cargar la ontología con los datos de la cabecera: a) seleccionar los datos provenientes de la cabecera; b) realizar un proceso ETL (extracción, transformación y carga) para compatibilizarlos con la ontología; c) identificar cuáles son los conceptos y relaciones de la ontología que se corresponden con esos datos, y d) instanciarlos en la ontología.

Se propone un método propio para el proceso ETL de las cabeceras, considerando algunos de los criterios vistos en las diversas técnicas de instanciación estudiadas. Así, se requiere del análisis morfoléxico de la cabecera del correo, aunque no enfocado en el *significado* de las palabras, sino más bien de la *ubicación* en el texto respecto de la estructura señalada por la norma RFC 822. Y se debe dotar al procedimiento de la máxima automatización, en virtud del volumen de datos que puedan considerarse durante el análisis forense de una cuenta de correo electrónico.

En esta fase se identifican las instancias de las clases de la ontología (*Correo, CuentaEmisor, Ocurrencias, etc.)* y las relaciones correspondientes, que se insertarán en la base de tripletas TCB a la cual se accede luego mediante *Apache Fuseki*[2]. Este proceso consta de dos sub-componentes: el *Identificador de Instancias Candidatas*, que detecta los datos relevantes en el archivo de texto plano y el *Clasificador de Instancias* que lleva adelante la instanciación en la ontología. Se describe a continuación cada una de éstas.

Identificador de Instancias Candidatas

En esta etapa la aplicación informática recibe un archivo en texto plano con la cabecera y el cuerpo del correo electrónico. Según la norma RFC 822, la cabecera y el cuerpo están separados por una línea en blanco, así, se lee el contenido línea por línea hasta hallar la primera línea en blanco, almacenando todo lo anterior como la cabecera del correo y el resto como el cuerpo. Luego, siguiendo la estructura de esa norma, se procesa la

[2] 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/).

cabecera seleccionando los datos para armar la base de conocimiento con los atributos y valores hallados. Previo, se verifica que la cabecera cumpla con los criterios mínimos exigidos para la pericia: a) el correo tiene un único emisor; b) se identifica la dirección IP del emisor, y c) se identifica la dirección IP del receptor.

Clasificador de Instancias

El Clasificador es un algoritmo que analiza la cabecera de correo en búsqueda de los atributos relevantes, los cuales están representados según el formato definido por la RFC 822. El algoritmo extrae los atributos relevantes de la cabecera y establece las relaciones para instanciar las clases necesarias para conformar una instancia de correo, por último, inserta las tripletas en la base que almacena la ontología.

A continuación, se describe el algoritmo utilizado y las detalles a considerar para analizar la cabecera de un correo electrónico. Cabe mencionar que durante las pruebas se encontraron casos en donde la cabecera no sigue el formato preestablecido por el RFC 822. Además, se ha tenido en cuenta lo descrito por (Darahuge & Arellano González, 2016) acerca del proceso de análisis forense de correos electrónicos.

Lectura de la cabecera

Palabra clave	Descripción	Clase	Atributo
Subject	Resumen del contenido del mensaje.	Asunto	contenidoAsunto
Message-ID	Identificador único del correo.	CorreoValido	idCorreo
Date	Fecha y hora de envío del mensaje.	OcurrenciaDeEmision	fechaHoraOcurrencia
From	Cuenta de correo electrónico del emisor y dirección IP del equipo emisor	CuentaEmisor	cuentaCorreo
To	Cuenta de correo electrónico del Receptor	CuentaReceptor	cuentaCorreo
Delivered-To			
X-Original-To			
Received o X-Received	Atributo insertado por cada equipo que interviene en la transmisión (incluyendo las máquinas del remitente, servidores y el destinatario), se obtiene la fecha y hora de la ocurrencia y la dirección IP del equipo (servidor o receptor).	OcurrenciaDeEmision	fechaHoraOcurrencia
		IP / HostName	IdentificadorEquipo
		OcurrenciaDeTransmision	fechaHoraOcurrencia
		IP / HostName	IdentificadorEquipo
		OcurrenciaDeEmisión	fechaHoraOcurrencia
		IP / HostName	IdentificadorEquipo

Tabla 1 – Palabras claves de una cabecera y clases que representan

La cabecera del correo se lee de abajo hacia arriba, ya que a medida que ésta va circulando, el proceso de transmisión añade datos al inicio de la cabecera. Cada línea define un atributo junto con su valor mediante el formato *{Palabra clave}: {Valor}*. A veces el valor ocupa varias líneas, y es necesario concatenar las líneas que en realidad son continuación de una anterior. Durante las pruebas realizadas se observó que éstas comienzan con al menos un espacio en blanco, así, aquellas que cumplen esa condición

son concatenadas al valor de la línea anterior (recursivamente). Luego que se obtienen todas las líneas concatenadas, cada una es leída en búsqueda del patrón {Palabra clave}: {Valor}. Las palabras claves que se buscan en la cabecera se resumen en la Tabla 1, indicando además el atributo de la clase que la representa en la ontología.

3.3. Analizador de los Puntos de Pericia

Este es el tercer componente de OntoFoCE, y su objetivo es responder 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 han sido descriptas en la sección 2, y se formalizan mediante consultas SPARQL, que son implementadas en la herramienta propuesta.

4. Caso Ejemplo

Para mostrar el uso de OntoFoCE como soporte para la forensia de correos electrónicos, se propone un caso ejemplo. Supongamos una causa en la que el juez formule al perito este punto de pericia: "*... Informe si durante la relación contractual (01/05/2017 al (01/01/2018) las partes se enviaron mensajes de correo electrónico desde la cuenta beatriz@empresa.com hacia la cuenta jose@empresa.com ...*", ordenando la pericia sobre la cuenta beatriz@empresa.com.

El perito accede a dicha cuenta y normalmente desde el equipo desde el cual se gestiona dicha cuenta. Es en este momento cuando se obtiene los datos auxiliares para el informe pericial: dirección física del equipo, cliente de correo utilizado y otra información relevante para el informe pericial.

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 requeridos. Ubicado ya en la cuenta de correo para realizar el análisis, se observa que la misma tiene un total de 31784 correos, de éstos, se busca y selecciona los 774 correos enviados/recibidos desde la cuenta *beatriz@empresa.com* a la cuenta *jose@empresa.com*, se asocian a una carpeta particular y se exportan en un archivo de texto plano. Se debe mantener la cadena de custodia de la prueba digital obtenida, para eso el perito aplica una función *hash* al archivo antes de grabarlo en el dispositivo externo protegido contra escritura, dejando constancia del número de encriptación del hash en el acta que registra todo el procedimiento pericial. Con ese conjunto de 774 correos, el perito responderá los puntos de pericia solicitados. Necesita recurrir a una herramienta que automatice la tarea de analizar esa cantidad de correos permitiéndole seleccionar los correos intercambiados en un rango de fechas. Puede realizar esto de varias formas, accediendo a distintas herramientas que le permitan procesar los correos, trabajando con cada una por separado según la funcionalidad ofrecida y obtener los datos que luego deberá interpretar para responder a los puntos de pericia.

OntoFoCe integra en un mismo entorno todas las funcionalidades requeridas, incluidas las respuestas a los puntos de pericia mediante las 21 preguntas de competencia contenidas en la ontología.

Una vez cargados los 774 correos en la aplicación, se inicia el proceso de identificación de los pares atributo valor para instanciarlos y representar la trazabilidad de cada correo. Tomando la cabecera de un correo de este ejemplo, en la Figura 3 se señalan las ocurrencias con sus correspondientes valores de fecha y dirección IP de los dispositivos en las que se almacenaron a cada paso de la transmisión.

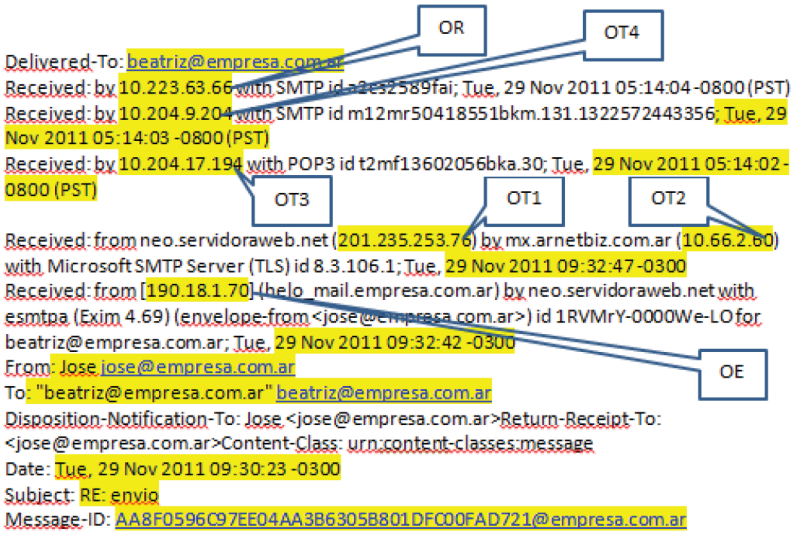

Figura 3 – Ejemplo de correo analizado con identificación de las ocurrencias

Y para el ejemplo de la Figura 3, en la Figura 4 se ilustra cómo muestra OntoFoCE las ocurrencias obtenidas a partir de la instanciación de la cabecera. Se observa en dicha figura que para el correo cuya cabecera se ilustra en la Figura 4, hay 1 ocurrencia de emisión y 1 de recepción que muestran el paso del correo por los equipos cuyas direcciones IP son 190.18.1.70 y 10.223.63.66, respectivamente. Asimismo, se observa los cinco equipos por los que pasó el correo electrónico en su camino desde el equipo emisor hasta el equipo receptor. Cada uno de estos equipos son identificados con un nombre, como en el caso de la ocurrencia de transmisión 1 o con su dirección IP, como ocurre con las otras 4 ocurrencias de transmisión. Para cada ocurrencia, se muestra también la fecha y hora en que el correo pasó por el equipo.

Considerando el requerimiento del punto de pericia, se puede responder al mismo mediante las preguntas de competencia 12, 13 y 21 (ver sección 2), cuyos resultados se muestran en la Figura 5. La respuesta a la pregunta 12 muestra los correos emitidos desde la cuenta beatriz@empresa.com a la cuenta jose@empresa.com; mientras que la respuesta a la pregunta 13 muestra los correos recibidos en la cuenta beatriz@empresa.com, constatando así que efectivamente hubo intercambio epistolar entre ambas cuentas. Por otra parte, la pregunta 21 muestra todos los correos intercambiados entre ambas cuentas en el período 01/05/2017 al 01/01/2018. Así, el perito puede adjuntar un impreso de los resultados de estas preguntas de competencia como base del informe pericial que debe entregar.

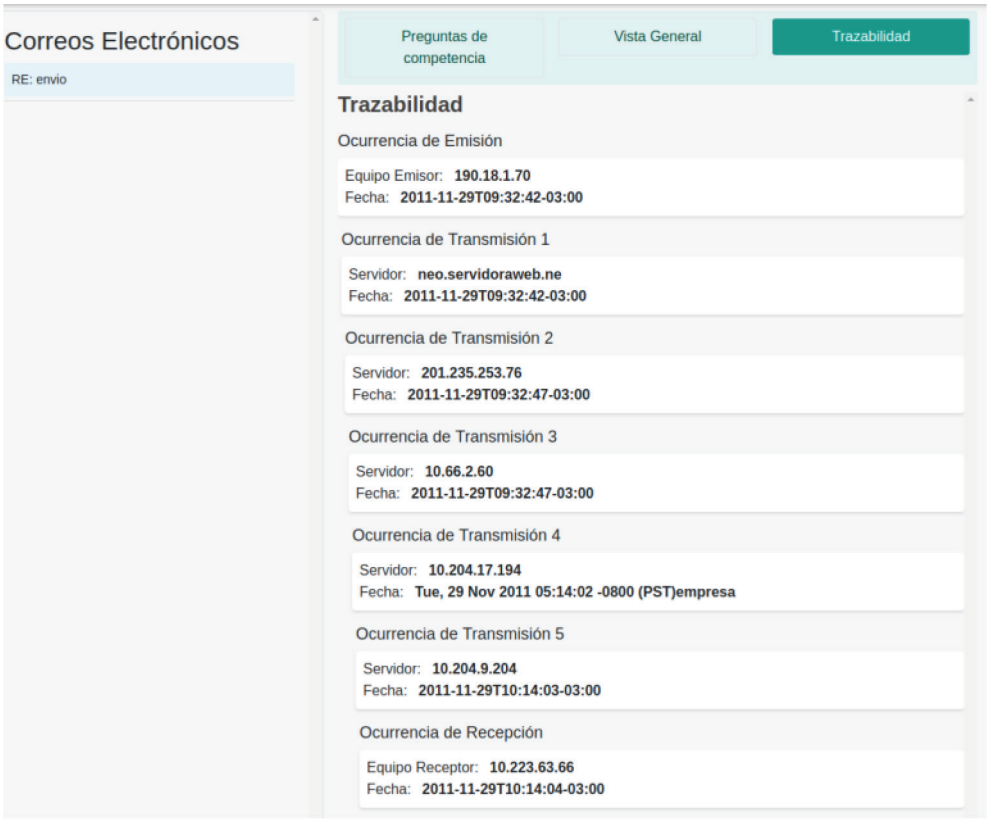

Figura 4 – Trazabilidad de un correo analizado

5. Discusión

A continuación, se describe el marco teórico considerado para definir el framework OntoFoCE propuesto para el análisis forense de correos electrónicos a partir de la ontología desarrollada.

Respecto de las herramientas disponibles para analizar un correo electrónico, se citan aquellas de más interés para la presente investigación.

Aid4Mail, utilizado para la migración y conversión de correos de/a diversos formatos, 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. Respecto de esta herramienta, OntoFoCE es superadora en cuanto que acepta archivos con formato plano, el modelo ontológico utilizado como soporte para el análisis forense brinda información más completa y se enfoca en dar una respuesta a los puntos de pericia.

EmailTrackerPro no sólo ofrece la capacidad de rastrear un correo electrónico usando su encabezado, sino que también permite filtrar spam, escaneando cada correo a medida

Figura 5 – Resultados de la pregunta de competencia N° 12, 13 y 21

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 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. Si se compara esta herramienta con OntoFoCE, se observa la coincidencia en el enfoque de la trazabilidad

del proceso de transmisión, a lo que OntoFoCE agrega por su parte las preguntas de competencia para responder a los puntos de pericia.

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 interfase gráfica con 5 visores (detalles del e-mail, ubicación geográfica de las IP, estadísticas; red social que vincula las IP con cuentas de correo; y minería de datos). No fue posible obtener una versión de esta herramienta para realizar un estudio comparativo sobre su funcionalidad versus las que ofrece OntoFoCE, pero de lo investigado se observa que, si bien el análisis apoyado en la minería de datos es lo más poderoso en esta herramienta, no se deduce que genere resultados directamente orientados a responder los puntos de pericia.

EnCase Forensic 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.

En el trabajo (Rivetti & Gallo, 2017) se realizó un estudio comparativo de algunas herramientas de uso libre, disponibles en la web para el análisis de estos casos, encontrándose que *MailXaminer* es la más eficiente cuando se trata de un conjunto masivo de cabeceras de correos electrónicos. Con la aplicación informática que aquí se presenta se realizó el análisis forense de 774 correos con un tiempo de procesamiento total de 22 segundos, incluyendo la extracción de los datos, clasificación e instanciación correspondiente en la ontología. Este mismo conjunto de datos, analizados con las herramientas *MailXaminer* insumió un tiempo de 83 segundos, lo que muestra la eficiencia del algoritmo para procesar grandes volúmenes de datos.

Respecto del Framework OntoFoCE, se puede decir que 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 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 es la posibilidad de incrementar el banco de preguntas de competencia, en caso de encontrar un punto de pericia que no pueda responderse desde las preguntas existentes. Esto es posible debido a que una nueva pregunta de competencia se formaliza e integra en la aplicación mediante una consulta SPARQL.

6. Conclusiones

Considerando el conjunto de herramientas analizadas, ninguna de ellas cumple integralmente con el cometido final de dar respuesta a los puntos de pericia de manera directa e inmediata, en todos los casos, corre a cargo del perito la identificación de aquellos datos que le son de utilidad para responder a los puntos de pericia. Ello es así porque no se consideran a estos últimos, los puntos de pericia, como elementos expresados en lenguaje natural que luego se representan mediante preguntas de competencia de una ontología. Se considera que ésta es la principal ventaja de OntoFoCE frente a otras herramientas.

Otra cuestión propia de la herramienta desarrollada, es la generación de un marco integral para el procesamiento automático de las cabeceras, cuando la cantidad de correos a analizar significa un volumen importante de datos, con el consiguiente esfuerzo en tiempo y dedicación que implicaría revisar una a una las cabeceras de correo o trabajando separadamente con herramientas de forensia e integrar luego los resultados obtenidos. El Framework OntoFoCE contempla el proceso ETL completo de las cabeceras de correos electrónicos, así como los procesos para instanciar la ontología, y realizar el análisis forense permitiendo contestar los puntos de pericia.

Cabe destacar que la aplicación se construyó utilizando tecnologías web, lo que permite disponer de una herramienta en línea accesible desde cualquier dispositivo con conexión a internet a través de un navegador web, considerando además los criterios y políticas de seguridad necesarias para el análisis forense.

Referencias

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

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

Brady, O., Overill, R., & Keppens, J. (2015). DESO: Addressing volume and variety in large-scale criminal cases. Digital Investigation, 15, 72–82. https://doi.org/10.1016/j.diin.2015.10.002

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

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.

Darahuge, M. E., & Arellano González, L. E. (2016). Manual de Informática Forense III.

David, L., Parra, L., Rico-bautista, D., Medina-cárdenas, Y., & Sanchez-Ortiz, A. (2018). Definición de una metodología práctica para la adquisición y análisis de evidencia digital en el contexto de una investigación post mortem, 369–384.

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 Ioro, A. H., Castellote, M. A., Constanzo, B., Curti, H., Waimann, J., Lamperti, S. B., ... & Nuñez, L. (2017). El rastro digital del delito Aspectos técnicos , legales y estratégicos de la Informática Forense. Mar del Plata: Esitorial UFASTA.

Faria, C., Serra, I., & Girardi, R. (2014). A domain-independent process for automatic ontology population from text. Science of Computer Programming, 95(P1), 26–43. https://doi.org/10.1016/j.scico.2013.12.005

Flores, V., & Hadfeg, Y. (2017). Un método para generar explicaciones de resultados de un Sistema Experto, usando Patrones de discurso y Ontología. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao, (21), 99–114. https://doi.org/10.17013/risti.21.99-114

Gallo, B. P. de, & Leone, H. (2016). Aplicación de la Ingeniería Ontológica para representar la trazabilidad de un Correo Electrónico. In: 2º Simposio Argentino de Ontologías y Sus Aplicaciones (SAOA 2016), 108–121.

Gallo, B. P. de, Vegetti, M., & Leone, H. (2015). Población de ontologías con datos no estructurados utilizando herramientas de minería de datos. In Congreso Nacional de Ingeniería Informática / Sistemas de Información - CONAIISI 2015, Buenos Aires, Argentina.

Gallo, B. P. de, Vegetti, M., & Leone, H. (2017). Hacía una Ontología para el soporte de la trazabilidad del correo electrónico en la Forensia Digital. In V Congreso Iberoamericano de Docentes e Investigadores de Derecho e Informática (CIIDDI 2017), Habana, Cuba.

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

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

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

Noll, R. P., & Ribeiro, M. B. (2007). Enhancing traceability using ontologies. In Proceedings of the 2007 ACM Symposium on Applied Computing, 1496–1497. https://doi.org/10.1145/1244002.1244322

Ovens, K. M., & Morison, G. (2016). Identification and analysis of email and contacts artefacts on iOS and OSX. In Proceedings of 2016 11th International Conference on Availability, Reliability and Security, ARES 2016, (October 2017), 321–327. https://doi.org/10.1109/ARES.2016.56

Rivetti, E. A., & Gallo, B. P. de. (2017). Estudio comparativo de desempeño de herramientas para el Análisis Forense de Correos Electrónicos. In Congreso Nacional de Ingeniería Informática / Sistemas de Información - CONAIISI 2017 (pp. 46–51).

Rueda-Rueda, J. S., Rico-Bautista, D., & Guerrero, C. D. (2018). Guía práctica abierta para el análisis forense digital en dispositivos Android, 442–457.

Sarli, J. L., Leone, H., & Gutierrez, M. (2019). SCFHLA: Un Modelo de Interoperabilidad Semántica para Simulación Distribuida de Cadenas de Suministro. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, (30), 34–50. https://doi.org/10.17013/risti.30.34-50

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

Stadlinger, J., & Dewald, A. (2017). A Forensic Email Analysis Tool Using Dynamic Visualization. The Journal of Digital Forensics, Security and Law, 12(1), 3-31. https://doi.org/https://doi.org/10.15394/jdfsl.2017.1413

Tolaba, A. C., Caliusco, M. L., & Galli, M. R. (2015). Representación del Conocimiento de la Información Geográfica siguiendo un Enfoque basado en Ontologías. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao, (14), 101–116. https://doi.org/10.17013/risti.14.101-116

Xie, L., Liu, Y., & Chen, G. (2016). A forensic analysis solution of the email network based on email contents. In 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015, 1613–1619. https://doi.org/10.1109/FSKD.2015.7382186

Ya'u, B. I., Nordin, A., & Salleh, N. (2017). Software requirements patterns and meta model: A strategy for enhancing requirements reuse (RR). In Proceedings of 6th International Conference on Information and Communication Technology for the Muslim World, ICT4M 2016, 188–193. https://doi.org/10.1109/ICT4M.2016.42

Youn, S. (2014). SPONGY (SPam ONtoloGY): Email classification using two-level dynamic ontology. Scientific World Journal, 2014. https://doi.org/10.1155/2014/414583

Yubao, W. U. (2015). Research on Electronic Data Tracing Method Based on Trusted Forensics. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, (16a), 204–213. https://doi.org/10.17013/risti.16a.204-213