Hacía una Ontología para el soporte de la trazabilidad del correo electrónico en la Forensia Digital

Towards an ontology supporting electronic mail traceability in Digital Forensic

Beatriz P. de Gallo (expositor)[1], Marcela Vegetti[2], Horacio Leone[3]

[1]I.Es.I.Ing. /Facultad de Ingeniería, Universidad Católica de Salta
Campo Castañares S/N, Salta, Argentina
bgallo@ucasal.edu.ar
http://www.ucasal.edu.ar
[2] INGAR/ Facultad Regional Santa Fe UTN
Avellaneda 3657, Santa Fe, Argentina
mvegetti@santafe-conicet.gov.ar
http://www.ingar.santafe-conicet.gov.ar/
[3]INGAR/ Facultad Regional Santa Fe UTN
Avellaneda 3657, Santa Fe, Argentina
hleone@santafe-conicet.gov.ar
http://www.ingar.santafe-conicet.gov.ar/

Resumen. Si bien la Forensia Digital avanzó en concordancia con la tecnología, aún se debe trabajar para que los resultados periciales se presenten adecuadamente, en función de los partícipes no informáticos que requieren del auxilio de esta disciplina. El análisis forense no debe presentarse como un reporte técnico sino como información sistemática y con sentido semántico en el marco de la causa judicial. Resulta conveniente contar con un marco de referencia común para todos los actores judiciales. Este trabajo propone, recurrir a tecnologías semánticas como soporte para ese marco de referencia, focalizándose en la representación de la *trazabilidad de un correo electrónico*, es decir, representando el camino recorrido desde la emisión del mail hasta la recepción por parte de sus destinatarios e identificando todas las instancias en las que el correo podría ser vulnerado. En particular, se introduce una ontología que define los principales conceptos y relaciones que representan este camino. Para la construcción de dicha ontología se recurre a METHONTOLOGY, que propone un proceso de ajuste iterativo para la especificación, conceptualización, formalización, implementación y mantenimiento de la ontología. Actualmente, la investigación avanzó hasta la definición de dos ciclos de iteración, se generaron instancias y procedimientos propios para el pretratamiento de datos para poblar la ontología, así como la formalización del modelo de datos estableciendo las relaciones, clases y objetos necesarios para la representación del problema a resolver.

1 Introducción

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 [1] presenta varios desafíos, involucrando no solo los modelos de "visilibilidad 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.).

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 [2] "una ontología es una especificación explícita de una conceptualización", vale detallar un poco más el concepto, tomando lo dicho por Reuver et al. [3] "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 grandemente en los resultados que se obtienen, i.e., la interpretación de los resultados. Harichandran et al. [4] señalan la importancia de mejorar las instancias de comunicación entre los técnicos y los profesionales del derecho, mejorando la accesibilidad y usabilidad de las herramientas de análisis forense para facilitar su interpretación por parte de los no técnicos. El volumen de datos que se obtiene al realizar el análisis forense debe ser interpretado a la luz de la pesquisa. La enorme cantidad de información técnica resultante del análisis de un correo electrónico debe insertarse 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. 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, no como información técnica, sino como elemento documental.

En el contexto de este requerimiento "no técnico", se encuentra la motivación de este trabajo. Resulta conveniente contar con un marco de referencia basado en la conceptualización formal del universo de discusión. Y 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).

Se propone considerar el correo electrónico como objeto de estudio, y en particular, aplicar las tecnologías semánticas para formular una ontología que permita incorporar el análisis forense de un correo electrónico como evidencia digital no repudiable.

Este artículo se organiza de la siguiente manera. La sección 2 define el marco de trabajo y la metodología utilizada como base para la construcción de la ontología. La sección 3 describe la ontología propuesta poniendo énfasis en la representación de los conceptos relacionados con la trazabilidad de un correo, mientras que la sección 4 presenta los resultados logrados hasta el momento. Finalmente, la sección 5 presenta las conclusiones y los trabajos a futuro.

2 Definición del Marco de Trabajo y Aspectos Metodológicos

El objetivo de esta sección es presentar una breve descripción del marco de trabajo en el cual se lleva adelante este proyecto. En particular se introducen las características principales asociadas al análisis forense de un correo electrónico, sus componentes y puntos de pericia. También, se muestra como las ontologías pueden dar soporte a las actividades dentro de la Forensia Digital. Finalmente, se describe brevemente la metodología utilizada para desarrollar la ontología propuesta.

2.1 Análisis Forense del Correo Electrónico

Tomando como base la tipificación propuesta por Banday [5] para el análisis forense de un correo electrónico se puede identificar tres componentes principales: los actores participantes en la transmisión, la arquitectura lógica y la estructura interna de un correo electrónico. De todos ellos, solo es de interés para una pericia la identificación de los actores principales, o sea, de aquellas personas que actuaron como probables emisores y/o receptores del correo electrónico bajo análisis. En referencia a esto se puede decir que, si bien la comunicación de un correo electrónico requiere de personas que actúan como emisor y receptor del mensaje, no son éstos los únicos partícipes de la transmisión. Banday [5] establece un mapa de relaciones y caminos posibles que puede recorrer un correo durante el proceso de transmisión e identifica los procesos responsables de sostener el servicio –denominados actores- que actúan internamente durante la transmisión.

Básicamente, 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, añadiéndole una etiqueta de identificación cada vez que el mail ingresa a un servidor. En base a estos elementos, es posible realizar la ***trazabilidad*** de un correo electrónico. La norma ISO 9000:2015 [6] define trazabilidad como la "capacidad para seguir el 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. Existen investigaciones que relacionan la trazabilidad y las ontologías, un detalle de estos estudios se incluye en [7]. Aplicado a un correo electrónico, la trazabilidad permitiría establecer el proceso desarrollado durante la comunicación que lleva del receptor al emisor, siendo éste –en última instancia- el objeto esencial del análisis forense de un correo electrónico.

Una pericia es un conjunto de operaciones técnicas científicas puestas en práctica para el esclarecimiento de un posible hecho ilícito y ordenadas por el Tribunal interviniente [8].

En cuanto a 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. Los puntos periciales se proponen en un pliego que señala las cuestiones técnicas, de manera clara y precisa, siempre referidas al tema que se dilucida en la litis y que técnicamente puedan ser respondidas por el Perito. Usualmente los puntos de pericia referidos a correos electrónicos abordan cuestiones relacionadas con la verificación de la autenticidad y existencia de un correo electrónico.

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

- la identificación de los datos del remitente (nombre de usuario, cuenta de correo y dirección IP),
- la trazabilidad del mismo (diferentes servicios o agentes que intervienen en la transmisión), y
- los datos del destinatario (nombre de usuario, cuenta de correo y dirección IP.

En cuanto a la existencia de un correo electrónico, ésta se puede probar fehacientemente cuando se comprueba la presencia del archivo digital del mismo tanto en el dispositivo emisor (o en el servidor del ISP del emisor) como en el dispositivo receptor del correo (o en el servidor ISP del receptor); y ambos archivos digitales son idénticos.

Desde el punto de vista de la forensia digital, existen muchas técnicas y herramientas que ayudan al Perito Informático en el análisis de un correo electrónico, algunas ya fueron analizadas en [9], a los citados allí se agrega el trabajo de Devendran, Shahriar y Clincy [10] acerca de un estudio comparativo de varios software open source para el análisis de correos electrónicos. Sobre este particular, se está trabajando en el análisis de un conjunto de herramientas propias de la forensia de correos electrónicos, a fin de identificar los datos que se pueden obtener con cada una (encabezado, IP, id del mensaje, casillas de correo intervinientes, fechas, hora, etc.) y que puedan ser utilizados para poblar la ontología que se propone en este trabajo.

Otra de las líneas de trabajo del presente proyecto de investigación, es la definición de procedimientos formales y normados para la realización de pericias sobre correos electrónicos. Al respecto en [11] se aborda el análisis de los documentos emitidos por autoridades judiciales respecto de procedimientos fijados para la obtención, presentación y tratamiento de la evidencia digital, considerando el caso particular de los correos electrónicos. Se observa que –si bien se definen desde el contexto procedimental del derecho- la más de las veces los documentos propuestos desde ese ámbito lo hacen con escaso rigor técnico en cuanto al procedimiento técnico-informático de tratamiento de los documentos digitales como evidencia. Por ello, se concluye con un aporte que podría considerarse en estas normas oficiales, en todo aquello que implique la obtención, presentación y tratamiento de los correos electrónicos como evidencia digital. Siempre sujetas al contexto jurídico vigente en Argentina. A continuación se resume brevemente el procedimiento citado:

1. Identificar la cuenta de correo a analizar, y determinar el proveedor de correo (dominio público, dominio privado).
2. Identificar el dispositivo (PC, Celular, Tablet, servidor de correo, etc) en el cual reside el correo electrónico aportado como prueba.
3. Identificar si el correo en análisis corresponde a un correo Emitido por el usuario o a un correo Recibido por el usuario. Si es un correo emitido, el encabezado solo dará certeza de que el correo salió de la cuenta de correo en análisis, mientras que si se trata de un correo recibido, el encabezado nos permite trabajar con la trazabilidad del correo hasta su origen.
4. Indicar al juez la necesidad de acceder a la cuenta y/o al dispositivo en el cual se encuentra residente el correo, considerando que puede tratarse de gestores de correo en la nube (webmail) o de clientes de correo local.
5. Una vez identificado el correo electrónico se debe extraer el encabezado completo, accediendo al mismo a través de los metadatos.
6. Se debe realizar una copia forense de la evidencia con su correspondiente valor hash para el resguardo correspondiente.
7. Analizar los campos de interés en los metadatos (X-Originating-IP, Received, direcciones IP, Message ID).
8. Todo el procedimiento debe ser efectuado en presencia de un escribano el cual validará las tareas realizadas.

En el trabajo citado en [11] se aborda además otras problemáticas relacionadas con la actividad pericial, como ser: herramientas forenses para la obtención de la prueba; necesidad de mantener un resguardo formal de la cadena de custodia; si corresponde, designación del hardware en donde se realizará la pericia; entre otros.

2.2 Ontologías para la Forensia Digital

Las ontologías proponen un marco de referencia basado en el conocimiento, mediante un vocabulario de representación que describe cada elemento según una definición declarativa y axiomas formales que acotan la interpretación y permiten una aplicación correcta de esos términos.

En el conjunto de sistemas de representación del conocimiento, las ontologías se definen según sus características distintivas:

- Permiten consensuar el significado de los elementos y relaciones de un universo de discusión, de manera que es posible desarrollar un software para modelar los procesos de toma de decisiones por parte de los gestores del conocimiento.

- Abordan siempre un dominio acotado del conocimiento. Si bien uno de los principales problemas al definir una ontología es identificar donde está el límite de lo que queremos representar, esa misma acotación sustenta y valida la representatividad de la ontología. Es decir, una vez

definido el dominio, las reglas de representación de una ontología permiten modelar acabadamente ese ámbito restringido denominado universo de discusión.

- Se recurre a la lógica formal para representar los componentes mediante los conceptos tradicionales de objetos, clases, instancias, restricciones y propiedades. Al utilizar modelos formales para la representación, es posible la aplicación de lenguajes compatibles con entornos abiertos y comprensibles para una máquina, tales como OWL, RDF, XML.

Se recurre a la semántica como hilo conductor para definir los componentes que se representarán, así como las relaciones de vinculación entre ellos, permitiendo expresar el dominio en base al significado que tienen sus componentes en el marco de referencia en el que actúan.

Existe una gran variedad de metodologías de diseño y construcción de ontologías. Algunas enfatizan una etapa sobre las otras, como la ingeniería de requerimientos o la evaluación y validación de la aplicación informática resultante por sobre la construcción del modelo formal.

Sin perder de vista el concepto en sí de una metodología (como herramienta solo es útil en la medida en que su uso acompaña el logro del objetivo propuesto), es importante destacar que en el caso particular de las ontologías, la definición del dominio es una parte sustancial, por lo que será necesario orientar la construcción de la ontología según sea la característica distintiva del universo de discusión que se considera. Así, las ontologías que tratan sobre vocabularios o taxonomías deben reforzar las instancias de significación de las palabras en el dominio que están abarcando; en otros casos, como en ontologías de integración de datos, es de interés profundizar la etapa de validación de los metamodelos de datos; o, en contextos específicos como la Forensia Digital, cobra vital importancia la validación de las instancias de captura de la prueba digital y su correspondiente "cadena de custodia".

Durante la investigación se avanzó en la lectura de trabajos de investigación que vinculen las ontologías y la forensia digital y los correos electrónicos. Valga como ejemplo el trabajo de Zhu [12] propone una ontología para compartir información sobre patrones de ciberataques y realizar un análisis sistemático y más eficiente de la información, así como el trabajo de Allen [13] sobre imágenes forenses de redes de correos electrónicos, y los trabajos de análisis de emails en otros soportes como teléfonos inteligentes [14] y [15]. De los últimos avances sobre el tema se puede considerar de interés el trabajo de Alzaabi et al. [16] que propone F-DOS un conjunto de ontologías que modelan formalmente el contenido de un teléfono inteligente. En particular, interesa el trabajo de Corcho et al. [17] en el que presentan una adaptación al dominio legal español de una taxonomía de clases sobre entidades legales, aplicando la metodología METHONTOLOGY y la herramienta WebODE.

Se observa en estos trabajos que se recurre a las ontologías como herramienta para modelizar el contexto de análisis, encontrándose en este punto la coincidencia de estas investigaciones con la que aquí se propone. Pero por otra parte se observa que son pocas las investigaciones conocidas acerca de la aplicación de ontologías a la forensia digital de correos electrónicos, con lo cual se entiende que la presente investigación puede generar aportes de interés para la problemática que se estudia.

2.3 Metodología para la Ontología Propuesta

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 sintetizan estas fases:

- La actividad de ***especificación*** permite determinar por qué se construye la ontología, cuál será su uso, y quiénes serán sus usuarios finales.
- La actividad de ***conceptualización*** se encarga de organizar y convertir una percepción informal del dominio en una especificación semi-formal, para lo cual utiliza un conjunto de representaciones intermedias (RRII), basadas en notaciones tabulares y gráficas, que pueden ser fácilmente comprendidas por los expertos de dominio y los desarrolladores de ontologías.
- La actividad de ***formalización*** se encarga de la transformación de dicho modelo conceptual en un modelo formal o semicomputable.

- La actividad de ***implementación*** construye modelos computables en un lenguaje de ontologías (Ontolingua, RDF Schema, OWL, etc.).
- La actividad de ***mantenimiento*** se encarga de la actualización y/o corrección de la ontología, en caso necesario.

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). Estas actividades consideradas accesorias a la construcción de la ontología, resultan de cierto peso durante el proceso de la investigación, principalmente por la necesidad de integrar datos y herramientas en el contexto semántico necesario.

3 Ontología Propuesta

La ontología que se aborda en este trabajo se formuló inicialmente en [9], sentando las bases de los principales componentes: preguntas de competencias, conceptos, relaciones, etc. En [18] se avanzó en la construcción de la ontología mediante el refinamiento de estos primeros conceptos incorporando todo lo relacionado con ocurrencias de correos, su trazabilidad y se ejemplificó con un caso de estudio.

Los interrogantes base de esta ontología se pueden buscar en los puntos de pericia que usualmente se proponen al solicitar un análisis forense de un correo electrónico y que se han enunciado en el apartado anterior, de allí se extractan las ***preguntas de competencia*** en lenguaje natural:

1. ¿Cuáles son las partes de un correo electrónico que resultan de interés para un análisis forense?
2. ¿Cuáles son los componentes informáticos a través de los cuales se escribe y se lee un correo electrónico?
3. ¿Cuáles son los datos o componentes que permiten validar la existencia de un correo electrónico? Esta pregunta puede descomponerse en:
   3.1. ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?
   3.2. ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?
4. Dado un correo electrónico ¿Cuáles son los datos que permiten identificar la autoría y recepción del mismo? Esta pregunta puede descomponerse en:
   4.1. ¿Cuál es el nombre de usuario y dirección de e-mail del Autor del mismo?
   4.2. ¿Cuál es el nombre de usuario y dirección de e-mail del Receptor del mismo?
   4.3. ¿Es posible establecer la trazabilidad del mensaje desde que se envía hasta que se recibe?
   4.4. ¿Cuáles son los diferentes actores/servicios que participaron de la transmisión?

El desarrollo de las diferentes etapas de la ontología se muestra en los trabajos [9], [18] y [19] en los que se describen los principales conceptos de la ontología, así como el diccionario de conceptos y la tabla de definición de objetos y relaciones del modelo. A continuación se introducen los conceptos y relaciones de la ontología propuestos para dar respuesta a estas preguntas de competencia a través del concepto de trazabilidad aplicado a un correo electrónico.

Considerando el "camino" que realiza un correo electrónico desde su emisión hasta la recepción por parte del destinatario, intervienen diferentes actores y procesos que se van desarrollando durante la transmisión. De esta secuencia de acciones, interesan en particular aquellas que pueden impactar en la *modificación del paquete de datos* que circula. Entonces, considerando que durante la transmisión el correo electrónico va residiendo en diferentes dispositivos de almacenamiento (equipo emisor, servidor de correo, servidores de paso, equipo receptor) es posible tomar cada archivo almacenado en estos dispositivos y verificar si el paquete de datos "original" fue modificado en algún punto durante la transmisión hasta llegar a destino. La verificación de una posible alteración del correo se realiza chequeando que el correo es el *mismo* en cada dispositivo en donde se va almacenando durante la transmisión. Esto es en cuanto al *cuerpo del mensaje¸* que debería mantenerse inalterable mientras que la *cabecera* del correo se va extendiendo a medida que se va almacenando en los servidores durante todo el camino de la transmisión.

Con independencia de las diversas definiciones de *correo electrónico* que ya se han formulado, en el marco de este trabajo se define al mismo en función de los elementos necesarios para la realización del análisis forense, i.e., *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.*

A partir de esta definición, se propone utilizar un enfoque ontológico para representar la trazabilidad de un correo electrónico basado en los siguientes aspectos de interés: a) el correo *original* se descompone en *ocurrencias*, tantas como veces se almacena en los distintos dispositivos durante el proceso de transmisión; b) existen tres tipos de ocurrencias: de emisión, de transmisión y de recepción; c) mientras que las ocurrencias de emisión y recepción son únicas, las de transmisión serán tantas, como veces en que el correo en transmisión quedó almacenado en un dispositivo durante su camino; d) las ocurrencias se asocian mediante una *secuencia* que establece el orden de aparición de cada ocurrencia, y con tantos *hilos* como receptores tenga el correo; y e) por otra parte, desde el punto de vista pericial, se requiere un conjunto de datos que permitan identificar plenamente al hardware (equipo emisor, servidor de correo, servidor de paso, equipo receptor) en donde el correo estuvo almacenado. Estos datos se incluyen en la clase localización que además recoge información derivada de los datos originales (geolocalización de la IP, ISP, entre otros).

El modelo conceptual especificado debe representarse por medio de un lenguaje formal, dando lugar a los componentes de la ontología: clases, atributos, conceptos, relaciones, funciones, axiomas e instancias. Las Fig. 1 y 2 muestran sendas vista parciales de la ontología propuesta[1]. En tanto, en el anexo A se presenta una vista parcial de la formalización de la ontología propuesta, que incluye los principales axiomas relacionados con los conceptos presentados a continuación.

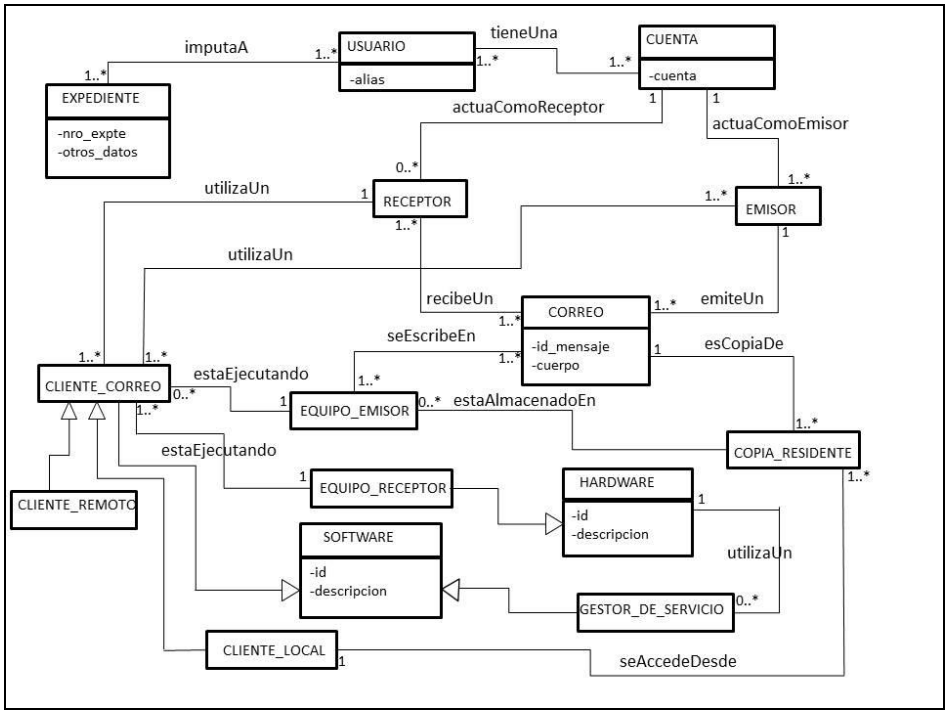

Figura 1: Vista Parcial de la ontología propuesta

Las principales clases identificadas son *CORREO, EMISOR y RECEPTOR.* El concepto *CORREO* se definió en los párrafos precedentes, *EMISOR* es la cuenta de correo desde la cual se emite el correo electrónico *y RECEPTOR* es la cuenta de correo destinataria del mismo. *CORREO* es una clase que se asocia con las clases *EMISOR* y *RECEPTOR* mediante las relaciones de *emitUN* y *recibeUn* respectivamente. A su vez las clases *EMISOR* y *RECEPTOR* se vinculan con la clase *CUENTA* mediante la relación *actuaComoEmisor* y *actuaComoReceptor* pues una misma cuenta puede actuar como emisor en una oportunidad y como receptor en otra. Un correo tiene un único emisor y podría tener uno o más receptores.

En el momento de la emisión del correo interactúan varios elementos, representados en el modelo conceptual por las clases *EQUIPO_EMISOR* y *CLIENTE_CORREO* mediante la asociación señalada como *estaEjecutando.* Cuando se escribe el correo se puede utilizar un *CLIENTE_CORREO* que sea *CLIENTE_REMOTO* en caso de utilizar un sistema web mail o que sea un

[1] Por razones de espacio no se incorporan los resultados de las primeras etapas de la metodología (Glosario de Términos, Taxonomía de Conceptos y definición de las relaciones binarias ad-hoc) que sustentan el modelo que aquí se describe.

CLIENTE_LOCAL en caso de un software de correo instalado en el equipo y con una cuenta de mail permanente, de allí que *CLIENTE_REMOTO* y *CLIENTE_LOCAL* se describen como subclases de *CLIENTE_CORREO*. Si se utiliza un *CLIENTE_LOCAL*, entonces existirá una *COPIA_RESIDENTE* del correo en el *EQUIPO_EMISOR,* asociado al *CORREO* mediante la relación *esCopiaDe* y al *EQUIPO_EMISOR* con la relación *estaAlmacenadaEn.*

Como se mencionara anteriormente, la ontología propuesta incorpora conceptos que permiten realizar la trazabilidad de un correo entre el emisor y un receptor. Los mismos se describe a continuación y se ilustran en la Fig. 2.

Si bien durante el proceso de transmisión el contenido del correo no se altera, cada vez que éste pasa por un servidor se va actualizando los datos de cabecera. A fin de lograr la trazabilidad será necesario identificar estas sucesivas versiones del mismo correo describiéndolas en la ontología propuesta mediante el concepto de OCURRENCIA. Así, un correo es una secuencia de ocurrencias representada mediante la relación *tieneUna* entre *CORREO y SECUENCIA.* Por otra parte, al representar el "camino recorrido" desde el emisor a un receptor, las ocurrencias conforman un *HILO.* Para representar esto, la ontología propone dos relaciones: i) *contiene* entre *HILO* y *OCURRENCIA* que especifica las ocurrencias que conforman un hilo y ii) la relación recursiva *siguiente* que establece el orden en que se fueron generando cada una de las ocurrencias en cada hilo. Para un correo electrónico, existirán tantos hilos como receptores tenga dicho correo ya que para ubicar a cada receptor puede ser necesario recorrer distintas vías de comunicación. Es así que, una secuencia agrupa uno o más hilos, esto se describe en la ontología mediante la relación de composición *agrupa* entre *SECUENCIA e HILO.* Todo correo tiene una y sólo una secuencia, la cual tiene al menos un hilo de ocurrencias.

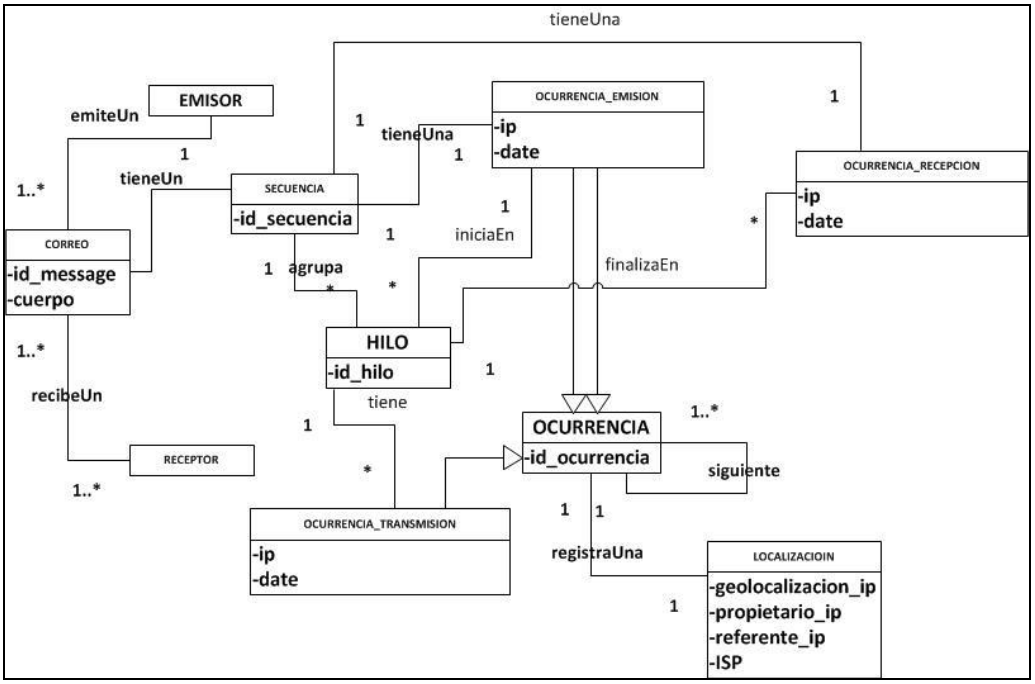

Figura 2: Representación de las ocurrencias de un correo electrónico

Por otra parte, en un HILO es posible identificar diferentes tipos de ocurrencias: *OCURRENCIA_EMISION, OCURRENCIA_TRANSMISION y OCURRENCIA_RECEPCION,* los cuales se definen a continuación. La *OCURRENCIA_EMISION* es la primera ocurrencia que se genera para el correo al momento de emitirlo, y es siempre única. Es la primera ocurrencia de la *SECUENCIA* y es compartida por todos los hilos que se generan en una misma SECUENCIA. La *OCURRENCIA_RECEPCIÓN* se genera al momento de la recepción del correo y es la última de cada *HILO.* Si bien es única para cada *HILO*, existen tantas *OCURRENCIA_RECEPCION* como hilos integren la secuencia del correo. La *OCURRENCIA_TRANSMISION* es el conjunto de ocurrencias que se generan durante los pasos internos del correo de servidor a servidor. Se distinguen porque siempre tienen una ocurrencia anterior y una ocurrencia siguiente y además porque no pueden estar en dos hilos diferentes. Se ha incorporado una clase *LOCALIZACION* asociada por medio de la relación

RegistraUna a la clase *OCURRENCIA.* Cada instancia de la clase *LOCALIZACION* permite representar los datos que ayudan a la identificación del servidor en donde se almacena la ocurrencia asociada a dicha instancia. En el caso particular de la localización asociadas a ocurrencias de emisión y recepción, se podrá registrar también otra información de interés para la causa judicial, como ser la localización geográfica del servidor, el proveedor del servicio de internet, el propietario registrado para la dirección IP, entre otros.

En [7] se describe exhaustivamente la trazabilidad mencionada, a partir de un caso de estudio y se incluyen además las formalizaciones expresadas en términos de predicados de lógica de primer orden.

4 Resultados Logrados

Siguiendo la metodología propuesta, se lograron resultados parciales que se fueron detallando en este trabajo y se describen a modo de resumen a continuación:

- Se formularon los modelos básicos de la ontología (conceptualización, formalización, instanciación) en una primera iteración y se expusieron en [9].
- La definición del dominio a partir de un conjunto de interrogantes base de esta ontología se pueden buscar en los puntos de pericia que usualmente se proponen al solicitar un análisis forense de un correo electrónico, de allí se extractan las preguntas de competencia en lenguaje natural que se describen en [9].
- En [19] se muestra el estudio realizado sobre las tecnologías para búsqueda y recuperación de información disponibles para la preparación de datos no estructurados, provenientes de correos electrónicos, como paso previo para la preparación del reservorio que se utilizará para poblar una ontología para el análisis forense de correo electrónico. Se formula un diagrama de procesos que representa la extracción semiautomática de datos, que posteriormente nutrirán ese reservorio, utilizando herramientas de minería de datos. En este trabajo se realiza una aplicación con una base de 1200 correos para realizar una validación preliminar de la estrategia propuesta.
- Durante la investigación surgió la necesidad de analizar las distintas fases del proceso de comunicación y transmisión por el que transcurre un correo electrónico, enfocando los aspectos de invariabilidad y protección del correo y sus metadatos, observando que se podía modelizar desde la característica de trazabilidad. Así en [20] se verifica el proceso de transmisión, los actores y componentes del mismo propuestos por Banday [5] con un caso de ejemplo.
- Se introduce el concepto de trazabilidad en la ontología, considerando las múltiples ocurrencias de un correo electrónico en el proceso de transmisión, identificándose así la ocurrencia de emisión, de transmisión y de recepción que se formalizan con lógica de primer orden en [7].

5 Conclusiones

En este trabajo se presentó el avance logrado en el proyecto de investigación con los resultados logrados a la fecha.

La necesidad de reconstruir el camino de inverso de un correo electrónico recibido se sustenta en que de este modo se puede probar la *existencia y autenticidad* del correo, avalando el carácter probatorio de este documento digital y reforzando la capacidad de *no repudio* del documento digital. Por otra parte, representar esta característica de trazabilidad del correo electrónico mediante una ontología, permite establecer un marco referencial científico y metodológico validado, que sirve de entorno de comunicación entre los partícipes informáticos y no informáticos del proceso judicial, al incorporar al proceso de análisis forense los beneficios de la ingeniería ontológica.

Entre las líneas previstas para la continuación de este trabajo se señala:

- Continuación del ciclo de construcción de la ontología avanzando en la formalización de las restantes reglas que permitan restringir el modelo e inferir nuevo conocimiento, instanciación y validación de la ontología.

- Verificación de los axiomas de autenticidad y existencia mediante una contrastación de puntos de pericia relevados entre los profesionales de la informática forense.
- Desarrollo del marco jurídico a partir de la incorporación de ontologías existentes sobre la temática, como por ejemplo SC-ONT propuesta por Kalemi et al. [21] que representa el dominio criminal a partir de las redes sociales en línea, en las que el correo electrónico es profusamente utilizado.

Referencias

[1] Garfinkel, Simson L.2010. Digital forensics research: The next 10 years, http://dfrws.org/2010/proceedings/2010-308.pdf página vigente al 15/11/2013

[2] Gruber, Thomas R. 1993. A Translation Approach to Portable Ontology Specifications. Knowledge Systems Laboratory. Technical Report KSL 92-71

[3] Reuver, Mark de. y Haaker Timber, 2009, Designing viable business models for context-aware mobile services. Elsevier, Volume 26, Issue 3, Telematics and Informatics, Pages 240–248 (August 2009)

[4] 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 & Security, 57, 1-13.

[5] Banday, M. Tarik, "TECHNIQUES AND TOOLS FOR FORENSIC INVESTIGATION OF E-MAIL", International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.6, November 2011

[6] ISO 9000-2015 (Traducción Oficial), Instituto Argentino de Normalización, Argentina, 2015

[7] 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

[8] Fernández, Eduardo Enrique: "Aspectos legales del peritaje". Revista INDICIOS, Año 2. Vol. 2. La Rioja (Argentina) 2011. pp. 24-33.

[9] Gallo Beatriz P. de, Vegetti Marcela, Leone Horacio, "Ontología para el Análisis Forense de Correo Electrónico", CoNaIISI 2014 Actas del 2° Congreso Nacional de Ingeniería Informática/Sistemas de Información, San Luis, Argentina, ISSN: 2346-9927, 2014

[10] Devendran, Vamshee Krishna, Hossain Shahriar, and Victor Clincy. "A Comparative Study of Email Forensic Tools." Journal of Information Security6.2 (2015): 111

[11] Rivetti Esteban, Aràoz 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

[12] Zhu, Y. (2015). Attack pattern ontology: A common language for attack information sharing between organizations (Doctoral dissertation, TU Delft, Delft University of Technology).

[13] Allen, G. (2016). Constructing and classifying email networks from raw forensic images (Doctoral dissertation, Monterey, California: Naval Postgraduate School).

[14] Ovens, K. M., & Morison, G. (2016, August). Identification and Analysis of Email and Contacts Artefacts on iOS and OS X. In Availability, Reliability and Security (ARES), 2016 11th International Conference on (pp. 321-327). IEEE.

[15] Chen, L., & Mao, Y. (2016, August). Forensic Analysis of Email on Android Volatile Memory. In Trustcom/BigDataSE/I SPA, 2016 IEEE (pp. 945-951). IEEE.

[16] Alzaabi, Mohammed. "The Use of Ontologies in Forensic Analysis of Smartphone Content." Journal of Digital Forensics, Security and Law 10.4 (2015): 105-114.

[17] Corcho, Óscar y Fernández-López, M. y Gómez-Pérez, A. y López-Cima, A., "Construcción de ontologías legales con la metodología METHONTOLOGY y la herramienta WebODE", Law and the Semantic Web. Legal Ontologies, Methodologies, Legal Information Retrieval, and Applications. Springer-Verlag, pp. 142-157. ISBN 0302-9743, 2005

[18] Gallo Beatriz P. de, Vegetti Marcela, Leone Horacio, "Avances en la Construcción de una Ontología para el Análisis Forense de Correo Electrónico", VI CIIDDI, Actas del 6° Congreso Ibeoramericano de Investigadores y Docentes de Derecho e Informática, Santa Fe, Argentina, 2016.

[19] Gallo Beatriz P. de, 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, 2015

[20] Rivetti E., Parra Beatriz, "Verificación de la trazabilidad de un correo electrónico mediante un caso ejemplo", Cuadernos de Ingeniería 2015, Número 9 del 2015. ISSN 2422-6572 (On line), ISSN 2422-6564, in press (febrero de 2016)

[21] Kalemi, E., & Yildirim-Yayilgan, S. (2016). Ontologies for Social Media Digital Evidence. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 10(2), 335-340.