Ontología para el Análisis Forense de Correo Electrónico

Beatriz P. de Gallo[a], Marcela Vegetti[b], Horacio Leone[b]
[a]*I.Es.I.Ing. /Facultad de Ingeniería, Universidad Católica de Salta*
Campo Castañares S/N, Salta, Argentina
[b]*INGAR/ Facultad Regional Santa Fe UTN*
Avellaneda 3657, Santa Fe, Argentina
[a]*bgallo@ucasal.net,* [b]*{mvegetti, hleone}@santafe-conicet.gov.ar*

Abstract

La gran cantidad de información técnica resultante del análisis forense de un correo electrónico debe insertarse en el conjunto de pruebas documentales de la causa judicial que lo aborda, 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. 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 de uso pluridisciplinar para facilitar el análisis de la prueba documental, por parte de todo los actores (abogados, jueces, investigadores y peritos). Este trabajo tiene como objetivo sentar las bases para el desarrollo de una ontología que colabore en la interpretación de los datos producto del análisis forense de un correo electrónico.

1. Introducción

Cada día toma mayor importancia el uso del correo electrónico, en su carácter de ***registro formal de una comunicación entre dos partes***. En el contexto legal, durante los últimos años ha crecido exponencialmente la presentación de correos electrónicos como prueba documental en las causas judiciales[1].

En su mayoría, el análisis forense de correos electrónicos consiste en responder dos cuestiones sustanciales para el carácter de prueba documental, que este elemento tiene:

- la ***autenticación del correo***, es decir, la validación de que este documento digital presentado como prueba legal es verdadero y no ha sido alterado ni manipulado; y
- la ***identificación del autor/receptor*** del mismo.

Queda por resolver la discusión planteada respecto de si un análisis forense puede aseverar con plena certeza que el *contenido* del correo electrónico no ha sido adulterado y validarlo en su carácter de prueba legal[2].

Si bien 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***.

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 (Internet Protocol) o la trazabilidad 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 dato 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 todo los actores (abogados, jueces, investigadores y peritos).

Existe abundante bibliografía a la que se puede recurrir para tomar los conceptos iniciales sobre ontologías. La definición de Gruber[1] parte del concepto de *conceptualización* como una abstracción o visión simplificada del universo que queremos representar con algún propósito. Las ontologías son una representación explícita o formal de esa conceptualización, recurriendo a la representación de los diferentes elementos que conforman el universo de discusión (objetos, conceptos y otras entidades), así como las relaciones que los vinculan.

1 De cada 10 causas judiciales en los que se solicitó la participación de un perito informático, 26 % tratan acerca de la autenticación y autoría de correos electrónicos. (Fuentes propias).

[2] En el ámbito del Derecho existen posturas contrapuestas acerca de si la verificación de autenticidad de un correo electrónico a partir del análisis forense, abarca (o nó) a su contenido, por ello, no se incluye esta cuestión en la problemática del presente trabajo.

Así, este trabajo tiene como objetivo sentar las bases para el desarrollo de una ontología que colabore en la interpretación de los datos producto del análisis forense de un correo electrónico.

La organización de este trabajo es la siguiente: la sección 2 describe el objeto de estudio (correo electrónico) señalando los aspectos de interés para el análisis forense y refiere brevemente el marco teórico de las ontologías. En la sección 3 se desarrollan las fases de especificación y conceptualización de una Ontología para el Análisis Forense de Correos Electrónicos. Finalmente, en la sección 4 se presentan las conclusiones y trabajos futuros.

2. Trabajos Relacionados

2.1. El Correo Electrónico

2.1.1. Definición y características propias

La literatura específica contiene diferentes definiciones del término correo electrónico. A los fines del presente trabajo, puede tomarse como válida la señalada por las Directrices de la Unión Europea 2002/58/CE [2] relativas a la protección de datos, en las que se definen el Correo Electrónico como "*Todo mensaje de texto, voz, sonido o imagen enviado a través de una red de comunicación pública que puede almacenarse en la red o en el equipo terminal del receptor hasta que éste acceda al mismo*".

El correo electrónico (o e-mail) se ha transformado en el medio de comunicación más utilizado en el tráfico de red facilitando grandemente la comunicación entre las personas. Además de acortar tiempos y distancia, permite el intercambio de múltiples tipos de datos (video, imagen, audio) y se encuentra accesible en prácticamente todos los medios de comunicación tecnológicos, habiendo avanzado rápidamente en la telefonía celular. Esta última característica de "portabilidad" abre instancias de comunicación que antes no estaban presentes, reforzando la inmediatez de la comunicación interpersonal, con el agregado de que ahora existe ***un registro de esta comunicación***. Si bien en el correo electrónico se adopta lenguaje coloquial, y es muy utilizado para la comunicación informal, es importante reconocer que es posible recuperar la conversación y utilizarla como prueba de que tal comunicación existió.

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.

Es importante señalar que la característica de volatilidad de los datos digitales, impacta negativamente en el reconocimiento de un correo electrónico (o cualquier otro componente digital) como prueba documental.

Los profesionales del foro judicial quieren "ver" la prueba, darle forma, buscar el origen, su historia, imaginar todo lo que hay alrededor de este componente, cual si fuera un "arma homicida". Recién en esta instancia pueden reconocer la validez de la prueba digital, i.e., una vez que logran ver la "sustancia" y "consistencia" de elemento probatorio.

2.1.2. Naturaleza Jurídica del correo electrónico

Relacionado con el contexto jurídico, Castro Bonilla [3] menciona tres enfoques o miradas diferentes que se pueden considerar en un correo electrónico:

a) Como correspondencia o comunicación: con idéntica naturaleza que el correo postal tradicional, se encuentra 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. En nuestro país, la Ley 26.388 incluye el término "correo electrónico" en el tipo de violación de correspondencia privada establecida por los arts. 153 y 154 del Código Penal.
b) Como conjunto de datos: al ser datos personales, su manipulación se encuentra supeditada a las normas relativas a la protección de datos personales. A partir del análisis forense de un correo electrónico es posible identificar una serie de datos del individuo (receptor/emisor del correo) que vulnera el derecho a la autodeterminación informativa de una persona. En Argentina, la Ley 25.326 *"…tiene por objeto la protección integral de los datos personales asentados en archivos, registros, bancos de datos, u otros medios técnicos de tratamiento de datos, sean éstos públicos, o privados destinados a dar informes, para garantizar el derecho al honor y a la intimidad de las personas, así como también el acceso a la información que sobre las mismas se registre,…".*
c) Como transmisor de material protegido por derechos de autor: al ser un medio de transmisión de datos de gran diversidad de formatos, hace posible la difusión de material protegido por derechos de autor, con un impacto notorio en la vulnerabilidad de la propiedad intelectual del material contenido o adjunto en un correo electrónico.

2.1.3. Estructura y componentes

A los fines del presente trabajo, se abordará brevemente aquellos aspectos del correo electrónico que resulten de interés para un análisis forense.

Tomando como base la tipificación propuesta por Banday [4] 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.

En referencia a los ***actores participantes*** se puede decir que, 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 actúan internamente durante la transmisión.

La ***arquitectura lógica*** de funcionamiento del envío/recepción de mail consta de varios componentes de hardware y software. Además de los dispositivos utilizados por el *Autor/Receptor*, sean éstos una computadora, un celular o cualquier otro equipo utilizado para el acceso al sistema de correo electrónico, se requiere de objetos transparentes al usuario principal 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, modem) y de la capa de aplicación de TCP/IP (nodos de e-mail) así como los diversos protocolos y paquetes de software utilizados (SMTP, HTTP, clientes de correo, etc.).

En cuanto a la ***estructura interna,*** un correo electrónico contiene un conjunto de *campos* asociados en *layers* que cumplen una función específica en la transmisión de datos. De ellos interesa identificar en particular: la cabecera del mensaje, el cuerpo o contenido, los archivos adjuntos asociados al mensaje, y las direcciones IP involucradas en la transmisión.

2.1.4. Análisis Forense de un Correo Electrónico

Una pericia [5] es un conjunto de operaciones técnicas científicas puestas en práctica para el esclarecimiento de un posible hecho ilícito y ordenadas por el Tribunal interviniente. 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[3].

Usualmente los puntos de pericia referidos a correos electrónicos abordan las siguientes cuestiones:

- Verificar la autenticidad, dirección, fecha y hora de emisión y recepción, autoría si fuera posible de los correos electrónicos cuyos impresos se adjuntan, así como cualquier otra información que estime relevante.
- Determinar la validez de origen de los e-mail enviados por A
- Expedirse acerca de la existencia de los envíos de los correos electrónicos emitidos entre A y B
- Informar si las cuentas de correo electrónico a@dominio.com y b@dominio.com están o estuvieron activas en los períodos consignados
- Informar si A tiene como correo electrónico la cuenta a@dominio.com
- Informar si en fecha dd/mm/aaaa desde la cuenta a@dominio.com se envió un correo a la cuenta b@dominio.com y en ese caso transcriba el texto del mensaje.
- Llevar a cabo la exploración y análisis de todos los soportes informáticos idóneos para el almacenamiento de información (notebooks, netbooks, pc's, discos rígidos, diskettes, cd's, dvd's, pen drives), instalados o que se encuentren en determinado lugar, ello con el objeto de determinar el soporte informático desde donde se hayan confeccionado documentos y/o remitido e-mails que dieran origen a las actuaciones.
- Realizar una correlación de eventos que relacione los datos de los correos electrónicos recibidos (dirección IP, emisor, destinatario/s, asunto, fecha, hora y observaciones varias) con los datos informados por la empresa X (nombre del usuario, titular del dominio, documento del titular, domicilio y localidad).
- Determinar si los correos que aparecen como enviados o recibidos son de las fechas que indica la parte en su demanda y si son coincidentes con los que mencionan como prueba la parte actora.
- Determinar 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
- Identificar cuáles fueron los equipos de origen y de destino del mensaje, si 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 realizado.

En particular interesa definir el ***carácter probatorio*** de un correo electrónico, mediante las características de autenticidad y existencia.

[3] A los fines del presente trabajo, se tomará en consideración el contexto más usual de la Forensia Digital, es decir, el ámbito de la justicia, sin perjuicio de que lo dicho aquí se pueda aplicar en los restantes ámbitos en que actúa esta disciplina (investigaciones privadas, fraudes tecnológicos internos, etc.).

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[4] 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.

En cuanto a las *técnicas*, la investigación forense de un correo electrónico se puede abordar desde varias ópticas: el análisis de los datos de cabecera, el análisis de los equipos emisores/receptores del correo, el análisis de los servidores ISP, el análisis de los metadatos ocultos en las aplicaciones utilizadas para la escritura del correo. Todas estas técnicas se utilizan habitualmente en conjunto para la confirmación redundante sobre la autenticidad del correo electrónico.

Respecto de las *herramientas* disponibles para analizar un correo electrónico, se puede citar como las más usuales las siguientes: eMailTrackerPro (analiza los encabezados de correo), Email Tracer (identifica el trazado o camino recorrido por un correo electrónico desde la emisión hasta la recepción), Adcomplain (informe los abusos de mailing), Aid4Mail Forense (utilizado para la migración y conversión de correos de/a diversos formatos), FTK (utilizada para el descifrado de claves), Encase (para la obtención de un archivo imagen del correo con el objeto de preservar la prueba original libre de manipulación), FINALeMAIL (para restaurar y recuperar mails borrados o perdidos).

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 alt. [6], 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 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.

Para el caso particular de los dispositivos móviles que se encuentran en el mercado, se pueden utilizar las siguientes herramientas de análisis forense: TULP2G, PARABEN y MOBILedit! FORENSIC[7]. Incluso existen herramientas diseñadas específicamente para el análisis forense de determinada tecnología celular, tal como el WOLF de Sixth Legion diseñada específicamente para el celular iPhone 3G [8].

Si bien las técnicas y herramientas mencionadas constituyen el marco formal y científico que califican la profesionalidad y rigor metodológico que se requiere en un análisis forense, los resultados que se obtienen no siempre cumplen su cometido: brindar información fundada sobre los puntos en litigio, o mejor dicho, responder los puntos de pericia de manera clara y contundente. El principal inconveniente radica en las dificultades que tienen los partícipes no informáticos de la causa (jueces, fiscales, abogados, investigadores forenses de otras disciplinas) para *interpretar los datos técnicos* a la luz de la causa judicial, y en el contexto del resto de las pruebas documentales presentes en el litigio.

De allí que se requiera de un sistema de representación que haga posible mostrar los datos en función del objetivo que se persigue (expresado en los puntos de pericia) y vinculados semánticamente en base a la relación que los mismos mantienen entre sí.

2.2. Las Ontologías

2.2.1. Definiciones básicas y características

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

[4] ISP = Internet Service Provider, proveedor del servicio de internet.

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.

2.2.2. Metodologías y herramientas para la construcción de ontologías

En la Ontology Summit 2007[9] se discutió sobre la gran variedad de metodologías de diseño, algunas de las cuales enfatizan una que otra propiedad en la construcción de la ontología: la ingeniería de requerimientos o la evaluación y validación de la aplicación informática resultante. Incluso se proponen metodologías sin diseño como en el caso de las folsonomías que parten del comportamiento local de miles de individuos.

No debe perderse 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. Y en el caso particular de las ontologías semánticas, en las que la definición del dominio es una parte sustancial, se debe orientar la construcción de la ontología según sea la característica distintiva del tema.

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".

En particular, interesa el trabajo de Corcho et al. [10] en el que presentan una adaptación al dominio legal español de una taxonomía de clases sobre entidades legales propuesta por Breuker[11], aplicando la metodología METHONTOLOGY y la herramienta WebODE.

Esta metodología ha sido desarrollada por el Grupo de Ingeniería Ontológica de la Universidad Politécnica de Madrid, permite construir ontologías en el nivel de conocimientos, y tiene sus raíces en las actividades identificadas por el proceso de desarrollo de software propuesto por la IEEE y en otras metodologías de ingeniería de conocimientos. ODE y WebODE se construyeron para dar soporte tecnológico a METHONTOLOGY.

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.

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. El resultado de esta actividad es el modelo conceptual de la ontología.
- 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 mayor parte de las herramientas de ontologías permiten llevar a cabo esta actividad de manera automática. Por ejemplo, WebODE puede importar y exportar ontologías desde y a los siguientes lenguajes: XML, RDF(S), OIL, DAML+OIL, OWL, CARIN, FLogic, Jess y Prolog.
- 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).

2.3. Trabajos de aplicación de ontologías en el análisis forense

Existen varios trabajos de investigación que relacionan ambos temas: las ontologías y el análisis forense.

Se recurre a las ontologías para representar la multiplicidad de dominios expertos que se requieren en el análisis forense (desde conocimiento de redes hasta conocimientos de sistemas contables) [12], o bien para el

diseño de un sistema inteligente en red aplicado a la forensia [13]. Para el caso particular de los correos electrónicos, es de interés el trabajo de Balakumar et al. [14] sobre la definición de una ontología para la clasificación y categorización de e-mail con el objetivo de conformar un filtro para la detección de spam mediante la conformación de una *whitelists* de remitentes conocidos, así como el trabajo de clasificación de e-mails propuesto por Taghva [15] en referencia a la exigencia legal de resguardar ciertos registros de datos residentes o anexados a correos electrónicos.

3. Ontología para el Análisis Forense de Correo Electrónico

En esta sección se presenta una ontología que permite definir la semántica de los conceptos relacionados con el análisis forense de un correo electrónico a fin de formalizar los mismos, así como sus relaciones y restricciones.

En este trabajo solo se abordarán las dos primeras fases propuestas por METHONTOLOGY, llegando a una primera conceptualización que servirá de insumo para el proceso iterativo que marca esta metodología.

3.1. Fase de Especificación

La primera actividad a desarrollar consiste en determinar el objetivo de la ontología, su funcionalidad y destinatarios finales.

El objetivo de una ontología para el análisis forense de correos electrónicos es el de construir un marco de referencia formal y científico, que sustente el análisis semántico de un correo electrónico en su carácter de prueba documental, basando esa interpretación en la relación que esos datos tienen en el contexto de la causa.

La función de esta herramienta, será la de servir de apoyo a los profesionales forenses no informáticos, facilitando la interpretación de los datos obtenidos con foco en la relación que los mismos tienen con el contexto de análisis y otras pruebas documentales, mediante la identificación de los conceptos, atributos, valores y relaciones que mantienen estos datos, más allá de las características técnico-informáticas que se obtienen como resultado del análisis forense de los correos electrónicos.

Los destinatarios finales son aquellos usuarios no informáticos, o con escasa experiencia con la tecnología (abogados, jueces, fiscales, otros investigadores forenses), que necesitan interpretar los resultados del análisis forense de un correo electrónico en el marco de la causa judicial, y a quienes –usualmente- los datos técnicos "crudos" que se obtienen como resultado de un análisis forense digital les resultan complejos o difíciles de comprender.

La definición del dominio parte de un conjunto de preguntas que guían la delimitación o acotación del universo de discusión, y que ayudan a decidir qué objetos son relevantes y cuales no son representativos para el análisis.

Estas preguntas tienen como cometido los siguientes:

- establecer las funcionalidades a las que debe responder la aplicación informática resultante de la ontología;
- delimitar el dominio tal como lo conciben los usuarios finales; y
- considerar los puntos de vista de los usuarios a la hora de modelizar los conceptos de la realidad.

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 las siguientes:
   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 las siguientes:
   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 el seguimiento del mensaje desde que se envía hasta que se recibe?
   4.4. ¿Cuáles son los diferentes actores/servicios que participaron de la transmisión?

3.2. Fase de Conceptualización

En esta segunda fase, corresponde trabajar sobre la definición del dominio, a partir de la percepción informal inicial que se tiene del mismo, estructurándolo en un modelo conceptual que luego sea posible formalizarlo.

Se comienza por definir el vocabulario de términos que se utilizará como base para responder a las preguntas de competencia. La Tabla 1 muestra el ***Glosario de Términos*** que integra la ontología propuesta.

Tabla 1: Glosario de Términos

Nombre	Sinónimos	Acrónimos	Descripción	Tipo
Correo Electrónico	e-mail	-	Todo mensaje de texto, voz, sonido o imagen enviado a través de una red de comunicación pública que puede almacenarse en la red o en el equipo terminal del receptor hasta que éste acceda al mismo	Concepto
Usuario	-	-	Persona que envía/recibe el correo electrónico	Concepto
Hardware	-	-	Componente informático que interviene en el emisión/transmisión/recepción del correo electrónico	Concepto
Equipo emisor/receptor	-	-	Hardware utilizado por el usuario para la escritura/lectura del correo electrónico	Concepto
Dispositivo de Transmisión	-	-	Hardware que interviene en el proceso de transmisión del correo electrónico	Concepto
Servidor de correo	-	-	Servidor que almacena el correo electrónico enviado/recibido	Concepto
Gestor de servicio	Actor	-	Componente de hardware o software que participa en la emisión/transmisión/recepción del correo	Concepto
Software	-	-	Aplicación informática que interviene en la emisión/transmisión/recepción del correo electrónico	Concepto
Cliente de correo	-	-	Aplicación informática que interviene en la escritura-emisión y lectura-recepción del correo electrónico	Concepto
Emisor	Remitente	-	Usuario que envía un correo electrónico	Concepto
Receptor	Destinatario	-	Usuario que recibe un correo electrónico	Concepto
Cuenta de correo	Dirección de mail	-	cuenta electrónica del usuario emisor o receptor	Atributo de Instancia
Fecha y hora	Date	-	Fecha y hora del correo electrónico emitido/recibido	Atributo de instancia
Dirección IP	-	IP	Dirección IP del correo electrónico a la cual se envío o desde la cual se remito	Atributo de instancia
Asunto	Subject	-	Tema de referencia del contenido del correo electrónico	Atributo de Instancia
Cuerpo	Contenido	-	Cuerpo del mensaje del correo electrónico	Atributo de Instancia
Archivo adjunto	Adjunto	-	Archivos que se remiten conjuntamente con el mensaje	Atributo de Instancia
Rol del usuario	-	-	Identificación del usuario respecto de si actúa como emisor o receptor del correo electrónico	Atributo de Instancia
Alias	-	-	Sigla o apodo con que se identifica el usuario. Usualmente este nombre está asociado al correo electrónico. Ej: pepe<juan.perez@dominio.com>	Atributo de Instancia
Nombre del usuario	-	-	Apellido y nombre real del usuario	Atributo de Instancia
Firma	-	-	Datos de identificación del usuario (nombre completo, cargo, teléfono, domicilio, etc.). Opcionalmente figuran estos datos en el cuerpo del mensaje	Atributo de Instancia
Tipo de Hardware	-	-	Referencia del hardware en cuanto a si actúa como equipo emisor/recetpor del correo electrónico o si participa como dispositivo intermedio en la transmisión.	Atributo de Instancia
Identificación única del equipo	ID equipo	-	Número de identificación única del equipo emisor/receptor (Serial Number, MAC address, etc.)	Atributo de Instancia
Dirección IP del equipo	-	-	Dirección IP del equipo emisor/receptor del correo electrónico	Atributo de Instancia
Nombre del equipo	-	-	Nombre básico del equipo emisor/receptor del correo electrónico. (PC,celular, notebook, tablet, etc.)	Atributo de Instancia
Descripción del equipo	-	-	Datos de descripción del equipo emisor/receptor del correo electrónico (tipo, marca, características volumétricas, configuración física, software de base, etc)	Atributo de Instancia
Ubicación geográfica del equipo	-	-	Ubicación geográfica del equipo emisor/receptor del correo electrónico	Atributo de Instancia
Identificación única del dispositivo	ID dispositivo	-	Número de identificación única del dispositivo que interviene en la transmisión del correo electrónico (Serial Number, MAC address, etc.)	Atributo de Instancia
Dirección IP del dispositivo	-	-	Dirección IP del dispositivo que interviene en la transmisión del correo electrónico	Atributo de Instancia
Nombre del dispositivo	-	-	Nombre básico del dispositivo que interviene en la transmisión del correo electrónico. (router, etc.)	Atributo de Instancia
Descripción del dispositivo	-	-	Datos de descripción del dispositivo que interviene en la transmisión del correo electrónico (tipo, marca, características volumétricas, configuración física, software de base, etc)	Atributo de Instancia

Nombre	Sinónimos	Acrónimos	Descripción	Tipo
Ubicación geográfica del dispositivo	-	-	Ubicación geográfica del dispositivo que interviene en la transmisión del correo electrónico	Atributo de Instancia
Identificación única del servidor	-	-	Número de identificación única del servidor de correo (Serial Number, MAC address, etc.)	Atributo de Instancia
Dirección IP del servidor	-	-	Dirección IP del servidor de correo	Atributo de Instancia
Dominio del servidor	-	-	Nombre de dominio del servidor que almacena el correo electrónico emitido/recibido	Atributo de Instancia
Descripción del servidor	-	-	Datos de descripción del servidor de correo (tipo, marca, características volumétricas, configuración física, software de base, etc)	Atributo de Instancia
Ubicación geográfica del servidor	-	-	Ubicación geográfica del servidor de correo	Atributo de Instancia
Tipo de servicio	-	-	Referencia de la actividad que desarrollan los componentes gestores de servicio, tales como: gestión de direcciones, gestión de reenvíos, gestión de distribución, etc	Atributo de Instancia
Identificación única del software	-	-	Número de identificación única del software (Serial Number, versión, etc.)	Atributo de Instancia
Nombre del software	-	-	Nombre comercial del software	Atributo de Instancia
Descripción del software	-	-	Datos de descripción del software (tipo, autoría, configuración actual, funcionalidades más notorias, etc)	Atributo de Instancia
Identificación única del cliente de correo	-	-	Número de identificación única del software cliente de correo (Serial Number, versión, etc.)	Atributo de Instancia
Nombre del cliente de correo	-	-	Nombre comercial del software cliente de correo	Atributo de Instancia
Descripción del cliente de correo	-	-	Datos de descripción del software cliente de correo (tipo, autoría, configuración actual, funcionalidades más notorias, etc)	Atributo de Instancia
Emisión (usuario, equipo, cabecera)			Relación que muestra el envío de un correo electrónico	Relación
Recepción (usuario, equipo, cabecera)	-	-	Relación que muestra la recepción de un correo electrónico	Relación
Seguimiento (hardware, software, datos)	Received	-	Relación que muestra la información de seguimiento de un correo electrónico durante la transmisión	Relación
Lista de distribución (Cuenta de correo 1, Cuenta de correo 2,…, Cuenta de correo N)	List	-	Conjunto de correos electrónicos que participan en la recepción simultánea de un correo	Relación

Se recomienda buscar vocabularios controlados existentes, es decir, recurrir a la reutilización de ontologías, para refinar y extender recursos ya establecidos para el dominio en discusión. Esto es muy útil cuando se requiere la interacción con otras aplicaciones que ya cuentan con representaciones ontológicas[5]. Un ejemplo se puede encontrar en la *LKIF core legal ontology* [16], consta de 15 módulos, cada uno de los cuales describe un conjunto de conceptos estrechamente relacionados a dominios tanto legales como de sentido común.

Tomando los componentes descriptos en el apartado 2 de este trabajo, los términos de interés deben surgir de los tres conjuntos definidos: los actores participantes en la transmisión, la arquitectura lógica y la estructura interna de un correo electrónico.

Siempre teniendo presente que las preguntas de competencia apuntan al carácter probatorio del correo electrónico, en función de las características de autentificación y existencia citadas anteriormente, se definen los conceptos, atributos de instancia y relaciones más notables

Una vez identificados los términos, se seleccionan aquellos que actúan como *conceptos*, para armar la ***Taxonomía de Conceptos***, que se muestra en la Figura 1.

METHONTOLOGY propone cuatro relaciones taxonómicas: Subclase-de, Descomposición-Disjunta, Descomposición-Exhaustiva, y Partición. A los fines del presente trabajo se han identificado las relaciones de Subclase-de que se señalan en la Figura 1. La identificación de las restantes relaciones taxonómicas se realizará en las siguientes instancias de revisión del modelo dentro del proceso iterativo señalado por la metodología.

La taxonomía desarrollada señala la relación de subclase de entre los usuarios más destacados (el emisor y el receptor del correo), que se distinguen a la vez del resto de los gestores del servicio porque son aquellos sobre quienes finalmente impacta el carácter probatorio del correo electrónico.

[5] Existen bibliotecas de ontologías disponibles en la web, tales como DAML, Ontolingua, SUMO o DOLCE.

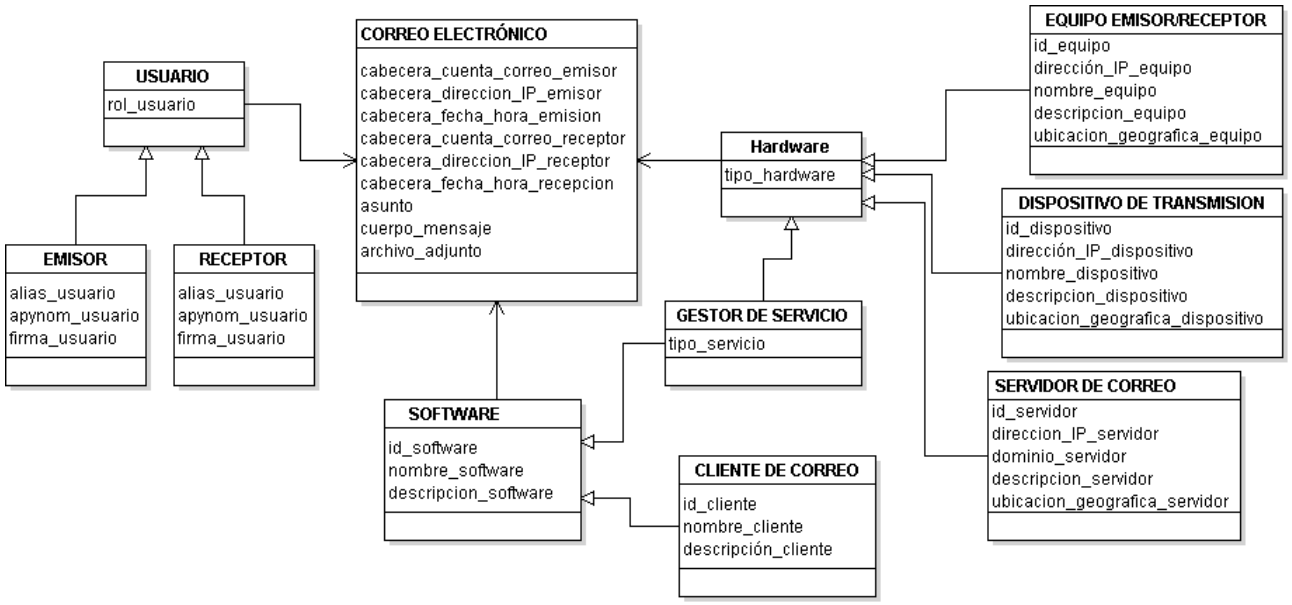

Figura 1: Taxonomía de Conceptos

La clase "software" representa los diversos programas que intervienen en la emisión / transmisión / recepción de un correo electrónico, de ellos resulta de interés la subclase "clientes de correo" ya que permiten *vincular* el correo electrónico con una persona (usuario emisor/receptor).

En cuanto a la especialización de la clase "hardware", en particular se identifica como sub-clase a los equipos utilizados para la emisión/recepción del correo, los servidores de correo y los dispositivos que participan de la transmisión del correo electrónico. Por su parte los equipos de emisión/recepción colaboran en la identificación de la vinculación usuario-correo electrónico, en los servidores de correo se encuentran almacenados los correos electrónicos, mientras que los dispositivos de transmisión muestran la *trazabilidad* del mensaje.

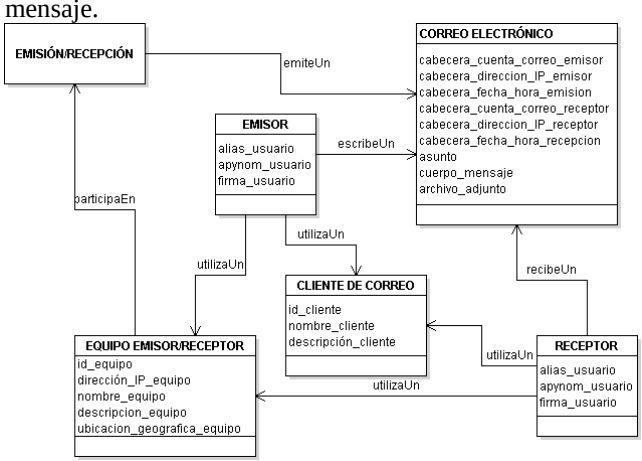

Figura 2: Relaciones ad-hoc para emisión/recepción del correo electrónico

Partiendo de la narrativa escrita en lenguaje natural – los puntos de pericia- se establecen las ***relaciones ad hoc*** existentes entre los conceptos definidos en la taxonomía. Las relaciones deben establecer con exactitud y precisión, indicando el origen y destino de cada una, evitando imprecisiones o sobreespecificación de esos puntos.

La Figura 2 muestra el proceso principal en el análisis forense, señalando los conceptos que participan y la vinculación que existe entre ellos.

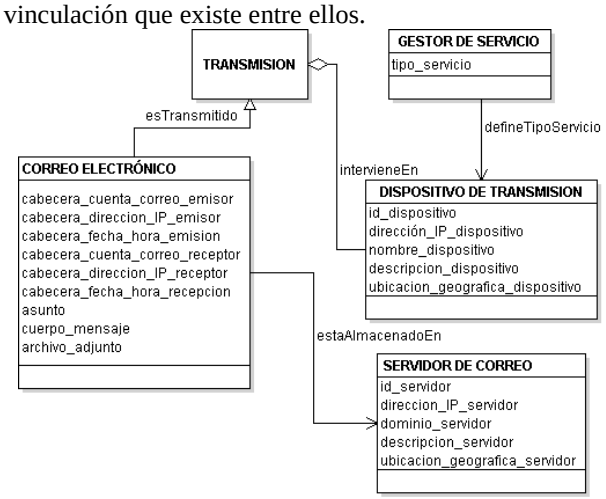

Figura 3: Relaciones ad-hoc para transmisión del correo electrónico

Una pregunta reiterada a los peritos es acerca de cómo es posible aseverar con plena certeza que el correo escrito por una persona es el que efectivamente recibió la otra (y viceversa). La relación ad-hoc que permite observar la vinculación de los componentes intermedios durante el proceso de transmisión se muestra en la Figura 3.

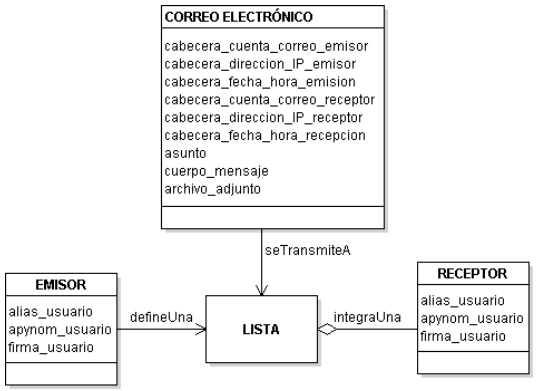

Figura 4: Relaciones ad-hoc para la Lista de Distribución del correo electrónico

Mientras que la Figura 4 muestra la vinculación entre los distintos partícipes de un correo electrónico, cuando se realiza un envío masivo, a través de una lista de distribución de correos.

Luego de identificadas las relaciones ad-hoc, METHONTOLOGY propone como siguiente paso la generación de una serie de tablas que definen el conjunto de metadatos de la ontología: el Diccionario de Conceptos, la Tabla que describe las Relaciones Binarias ad-hoc, los atributos de instancia y de clase, y las constantes.

Por razones de espacio, solo se incluye en este trabajo el Diccionario de Conceptos que se presenta en la Tabla siguiente:

Tabla 2: Diccionario de Conceptos

Concepto	Instancia	Atributos de Clase	Atributos de Instancia	Relaciones
Correo Electrónico	Prueba Digital Prueba documental	-	Cuenta de correo, Fecha y hora, Dirección IP, Asunto, Cuerpo, Archivo adjunto	emiteUn recibeUn esTransmitido seTransmteA
Usuario	-	Atributo de Clase	Rol del usuario	-
Hardware	-	Atributo de Clase	Tipo de Hardware	-
Equipo emisor/receptor	PC Celular Tablet	-	Identificación única del equipo, Dirección IP del equipo, Nombre del equipo, Descripción del equipo, Ubicación geográfica del equipo	participaEn utilizaUn
Dispositivo de Transmisión	Switch Router Servidor DNS	-	Identificación única del dispositivo, Dirección IP del dispositivo, Nombre del dispositivo, Descripción del dispositivo, Ubicación geográfica del dispositivo	intervienEn
Servidor de correo	Servidor del ISP	-	Identificación única del servidor, Dirección IP del servidor, Dominio del servidor, Descripción del servidor, Ubicación geográfica del servidor	estaAlmacenadoEn
Gestor de servicio	-	Atributo de Clase	Tipo de servicio	defineTipoActividad
Software	Protocolos de transmisión Firmware	Atributo de Clase	Identificación única del software, Nombre del software, Descripción del software	-
Cliente de correo	Outlook WebMail Mozilla Thunderbird	-	Identificación única del cliente de correo, Nombre del cliente de correo, Descripción del cliente de correo	utilizaUn
Emisor	Actor Demandado	-	Alias, Nombre del usuario, Firma	escribeUn defineUna
Receptor	Actor Demandado	-	Alias, Nombre del usuario, Firma	recibeUn integraUna

Si bien la metodología seguida propone especificar los ***axiomas formales*** con todo rigor (nombre, descripción en lenguaje natural, expresión lógica que define de manera formal el axioma usando lógica de primer orden, y los conceptos, atributos y relaciones ad hoc utilizadas en el axioma, así como las variables utilizadas), a los fines del presente trabajo solo se presentan dos de los axiomas destacados, y que en cierta forma, conforman la base del análisis forense de un correo electrónico. Éstos se presentan a continuación, y escritos en lenguaje natural:

- ***Axioma 1: sobre la autenticidad de un correo electrónico***

Un correo electrónico es auténtico cuando se identifican: 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).

- ***Axioma 2: sobre la existencia de un correo electrónico***

Un correo electrónico existe cuando se comprueba la presencia del archivo digital del mismo tanto en el

dispositivo emisor (ó en el servidor del ISP del emisor) como en el dispositivo receptor del correo (ó en el servidor ISP del receptor); y ambos archivos digitales son idénticos.

Hasta aquí se presentó la primera iteración de la fase de conceptualización de la ontología, resta realizar las iteraciones necesarias hasta ajustar el proceso y avanzar luego en las fases de formalización, implementación y mantenimiento propuestas por METHONTOLOGY.

4. Conclusiones

Si bien se avanzó en una primera formulación de la ontología, que seguramente será actualizada y mejor representada en cada proceso iterativo que conlleva la metodología elegida, lo escrito hasta aquí permite mostrar los beneficios de la propuesta. Esta primera formulación del modelo conceptual deberá ajustarse tanto en la profundidad del análisis de cada paso, como en la actualización de los conceptos, atributos, relaciones y axiomas, cuidando de agregar términos o descartar aquellos que fueran subsumidos en otros como resultado del trabajo iterativo.

Las ontologías generan un marco de trabajo adecuado y suficiente para el análisis e interpretación de datos en contextos de cierta complejidad, al proporcionar un sistema de representación basado en un vocabulario común y consensuado entre todos los partícipes del tema, facilitando el intercambio y enriqueciendo la base de conocimiento a partir de la interpretación que cada quién realiza sobre el mismo conjunto de datos.

Así, es posible establecer un contexto específico de comunicación y reutilizar el conocimiento existente haciendo que resulte más fácil incrementar ese conocimiento y difundirlo entre la comunidad interesada en la temática de estudio.

Estos beneficios –propios de cualquier tipo de ontología- se ven incrementados cuando facilita la comunicación y vinculación entre dos áreas muy separadas entre sí: por un lado el Derecho (que se apoya en la formalidad de la norma escrita) y por otro la Informática (que se apoya en el contexto virtual).

5. Agradecimientos

Este trabajo ha sido financiado en forma conjunta por CONICET, la UTN (PID 25-O156) y el Consejo de Investigaciones de la Universidad Católica de Salta. Se agradece el apoyo brindado por estas instituciones.

6. Referencias

[1] Gruber, Thomas R., "A Translation Approach to Portable Ontology Specifications", *Appeared in Knowledge Acquisition*, 5(2):199-220, 1993.

[2] DIRECTIVA 2002/58/CE DEL PARLAMENTO EUROPEO Y DEL CONSEJO SOBRE LA PRIVACIDAD Y LAS COMUNICACIONES ELECTRÓNICAS, 12 de julio de 2002, p.37

[3] Castro Bonilla, A."El uso legítimo del correo electrónico", *II Congreso Mundial de Derecho Informático*, 2002

[4] 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.

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

[6] Hadjidj, R., Debbabi, M., Lounis, H., Iqbal, F., Szporer, A., & Benredjem, D. "Towards an integrated e-mail forensic analysis framework". *Digital Investigation*, *5*(3), 124-137.2009

[7] Agualimpia, C., & Hernández, R. Análisis forense en dispositivos móviles con Symbian OS. *Documento de maestría, Dept. Ingeniería electrónica, Pontifica Universidad Javeriana, http://www. criptored. upm. es/guiateoria/gt_m142e1. htm*.

[8] Ariza, A., Ruíz, J., & Cano, J. iPhone 3G: Un Nuevo Reto para la Informática Forense. *Universidad Pontifica Javeriana, Bogotá-Colombia*.

[9] Ontology Summit 2007 Communiqué, 2007, version 1.0.0 / 2007.04.24

[10] 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

[11] Breuker, J., Elhag, A., Petkov, E., & Winkels, R." Ontologies for legal information serving and knowledge management", *Legal Knowledge and Information Systems, Jurix 2002: The Fifteenth Annual Conference* (pp. 1-10).

[12] Schatz, B., Mohay, G. M., & Clark, A. (2004). Generalising event forensics across multiple domains. *School of Computer Networks Information and Forensics Conference*, Edith Cowan University. .

[13] Saad, S., & Traore, I. (2010, August). Method ontology for intelligent network forensics analysis. In *Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on* (pp. 7-14). IEEE.

[14] Balakumar, M., & Vaidehi, V. (2008, January). Ontology based classification and categorization of email. In *Signal Processing, Communications and Networking, 2008. ICSCN'08. International Conference on* (pp. 199-202). IEEE.

[15] Taghva, K., Borsack, J., Coombs, J., Condit, A., Lumos, S., & Nartker, T. (2003, April). Ontology-based classification of email. In *Information Technology: Coding and Computing, International Conference on* (pp. 194-194). IEEE Computer Society.

[16] LKIF-Core Ontology: A core ontology of basic legal concepts. http://www.estrellaproject.org/lkif-core/ vigente al 2-ago-2014.