Aplicación de tecnologías semánticas a la Forensia Digital

ETAPA 1: Estudio y Diseño de una Ontología Semántica

Beatriz P. de Gallo[a], 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.edu.ar, [b]{ hleone}@santafe-conicet.gov.ar

Resumen

Considerando la vinculación interdisciplinaria de la Informática y las Ciencias Jurídicas, se propone trabajar en el área de la Forensia Digital, particularmente mediante el empleo de la ingeniería ontológica para estudiar un conjunto de problemas básicos que puedan ser tratados y expresados eventualmente a través de una ontología.

El trabajo de investigación propone generar un contexto de análisis de la evidencia digital que permita una visión integrada, sistemática y orientada al sujeto. Se propone circunscribir el contexto de aplicación y experimentación del tema en estudio, según criterios de alcance, profundidad, oportunidad y acceso a problemáticas reales de la Forensia Digital, identificando los problemas básicos al momento de realizar el análisis forense de la información digital. Esto permitió acotar el universo de estudio al análisis forense de correos electrónicos y en la propuesta de una ontología dirigida a generar un espacio de comunicación común entre todos los actores del proceso judicial (jueces, abogados, peritos, etc.), con la intención de que la información técnica proveniente del análisis forense de correos electrónicos se incorpore a la causa con idénticas consideraciones sobre el valor probatoria que otras pruebas documentales.

Palabras clave:

Ontologías Semánticas, Forensia Digital

Contexto

En respuesta a la política institucional de definición de planes de desarrollo de la investigación en todas las unidades académicas con especificación de prioridades, la Facultad de Ingeniería de la UCASAL definió sus líneas de investigación prioritarias, atendiendo a la evolución de las actividades de investigación desarrolladas en la facultad, a la conformación de equipos docentes y la demanda del medio productivo local. Entre ellas, interesa la referida a "*Tecnologías Informáticas Aplicadas"* que sirven de marco de referencia para el GRUPO DE INVESTIGACIÓN EN INFORMÁTICA FORENSE, integrado por docentes de la carrera de Ingeniería en Informática de esa Casa de Altos Estudios.

Este proyecto se presentó al Consejo de Investigaciones de la UCASAL, y pasó las instancias de admisibilidad y evaluación por pares externos según lo

estipula la RR N° 1083/14 en el procedimiento de presentación de proyectos de investigación. Por RR N° 656/15 se aprueba el proyecto y su financiamiento por parte de la UCASAL.

En este contexto se desarrolla el proyecto que aquí se presenta.

Introducción

Desde hace más de una década que los diferentes estamentos de seguridad –tanto militares como judiciales y políticos- se preocupan por encarar la lucha contra el crimen desde la óptica tecnológica, es decir, con una mirada cada vez más preocupante sobre el uso de la tecnología para delinquir. En 2001 la Digital Forensic Research Conference (DFRWS) definió la "Forensia Digital" como "El uso de métodos científicamente derivados y provados a la preservación, recolección, validación, identificación, análisis, interpretación, documentación y presentación de la evidencia digital derivada de fuentes digitales para el propósito de facilitar o favorecer la reconstrucción de los hechos criminales o para la prevención de acciones no autorizadas que se estima como perjudiciales para operaciones planificadas"[1].

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 [2] 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 la ingeniería reversa o 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 [3] "una ontología es una especificación explícita de una conceptualización", vale detallar un poco más el concepto, tomando lo dicho por Reuver et alt.[4] "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".

Líneas de Investigación, Desarrollo e Innovación

El Plan de trabajo consiste en profundizar el análisis del dominio de la forensia digital, sus requerimientos de información y estudiar los casos vinculados al análisis forense de correos electrónicos. Se identificarán delitos de interés social tomando como organismo público de investigación de delitos de referencia al Ministerio Público de la Provincia de Salta y se tratará la naturaleza jurídica de la prueba digital –y en particular- del correo electrónico.

A partir de los requerimientos identificados, se formalizarán las debilidades de las arquitecturas y modelos desarrollados hasta el presente y las razones de las mismas, y se comenzará en el diseño de la ontología que represente el conocimiento requerido en el caso seleccionado.

Se seleccionó METHONTOLOGY como metodología de trabajo para la definición de la ontología, y se prevé el desarrollo de las etapas básicas (especificación de requerimientos, conceptualización, implementación y evaluación). A fin de evaluar la ontología en una aplicación se procederá a la implementación de una herramienta a nivel de prototipo.

En resumen, para el desarrollo de este plan se proponen las siguientes actividades: Búsqueda bibliográfica, estudio del estado del arte; definición del contexto de aplicación y experimentación; estudio de factibilidad técnica y operativa de aplicación; adquisición del conocimiento; especificación de usos y usuarios de la ontología; conceptualización y estructuración del conocimiento en modelos significativos; formalización del modelo; implementación mediante modelos computables y validación y ajuste de la ontología.

Resultados y Objetivos

La Figura 1 resume el grado de avance en la conceptualización de la ontología. Se describen los principales conceptos, atributos de instancias y las relaciones más notorias.

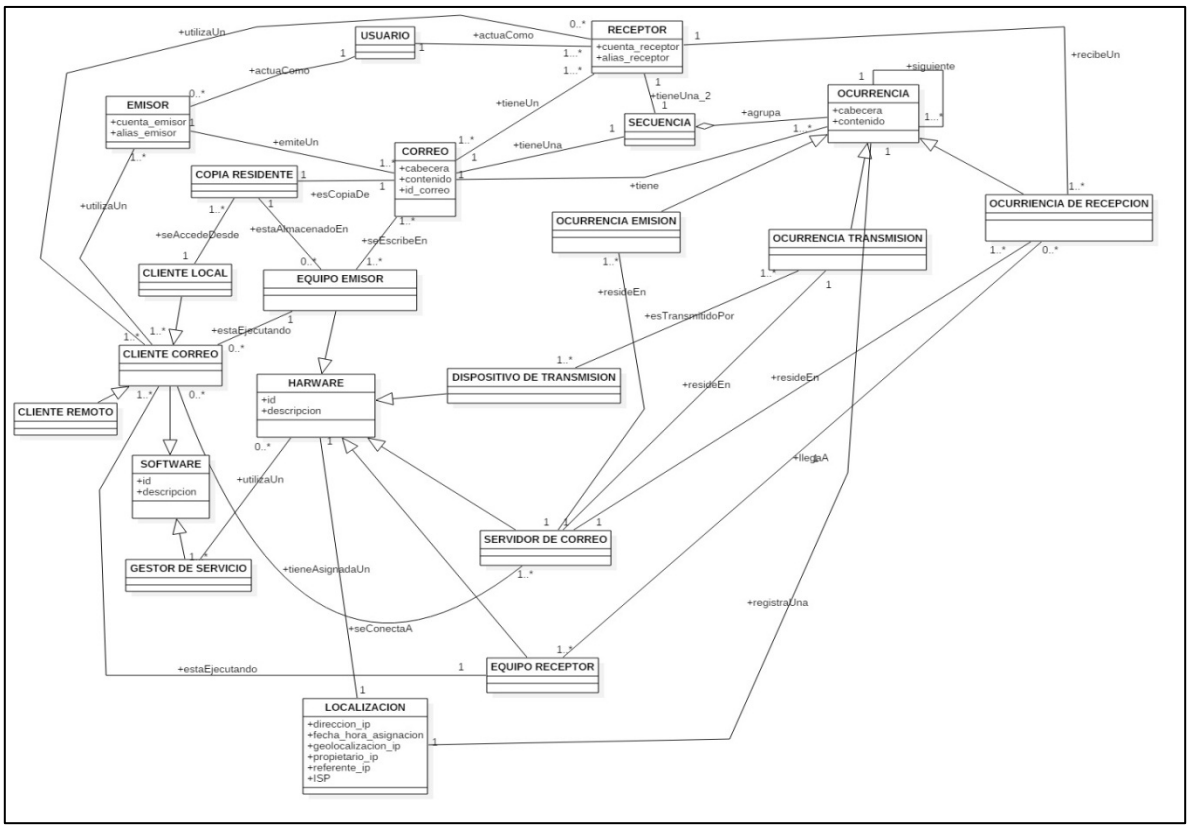

Figura 1: Ontología Para el Análisis Forense de Correo Electrónico

A partir de los puntos de pericia se formularon las siguientes preguntas de competencia:

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?
4. Dado un correo electrónico ¿Cuáles son los datos que permiten identificar la autoría y recepción del mismo?

Y sobre estas preguntas se trabajó en un primer ciclo de la iteración metodológica en el que se definieron los conceptos, atributos de instancias y relaciones intervinientes, los que se describen ampliamente en [5].

A la fecha, el proyecto se encuentra en desarrollo de acuerdo al plan de trabajo inicialmente formulado.

Se cita entre los principales resultados:

- Estudio de la naturaleza jurídica del correo electrónico como prueba digital, desarrollado en [5].
- Se investigó acerca de la utilización de técnicas de minería de datos para el procesamiento previo de los datos para poblar la ontología, descripto en [6].
- Verificación de la trazabilidad de un correo electrónico, desarrollado en [7]
- Se presentaron los avances logrados en un evento académico específico de esta temática [8].

En estos momentos, se está investigando sobre la incorporación del concepto de trazabilidad del correo electrónico en el modelo ontológico,

considerando que durante el proceso de transmisión el correo consta de sucesivas ocurrencias que se almacenan en servidores transitorios hasta llegar a destino.

Formación de Recursos Humanos

La RF N° 153/12 de la Facultad de Ingeniería de la UCASAL avala la conformación de grupos de investigación en el ámbito de esa unidad académica, destinados a fomentar la asociación de los investigadores en conjuntos que sustenten un programa estable para favorecer el trabajo de investigación multi-disciplinario, optimizando así el uso de los recursos, facilitando la inserción en el medio socio-productivo, la constitución de redes y aumentando la competitividad de los investigadores en la captación de recursos desde agencias nacionales e internacionales. De esta acción surge el Grupo de Informática Forense, al cual pertenecen los investigadores del presente proyecto.

El equipo de trabajo del proyecto está conformado por el Dr. Horacio Leone, director, MBA Ing. H. Beatriz P. de Gallo, co-directora del proyecto, el Ing. Esteban Rivetti y la Esp. Abog. Marìa Isabel Rodriguez Virgili.

El proyecto prevé además la incorporación de alumnos investigadores: los alumnos Esteban Rivetti (luego incorporado al equipo de investigación), Natalia Farfán y Ebaneo Kao desarrollaron sus trabajos de finalización de carrera en temáticas vinculadas a esta línea de investigación.

Es intención que este proyecto de investigación actúe como marco de contención para el desarrollo de un trabajo de tesis doctoral, a cargo de la Ing. H. Beatriz P. de Gallo, actualmente doctorando en la carrera de Doctorado en Ingeniería Mención en Ingeniería en Sistemas de Información, de la Universidad Tecnológica Nacional – Facultad Regional Santa Fe.

Referencias

[1] DFRWS TECHNICAL REPORT.2001. A Road Map for Digital Forensic Research, http://www.dfrws.org/2001/dfrws-rm-final.pdf Página vigente al 15/11/2013

[2] 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

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

[4] 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)

[5] Beatriz P. de Gallo, Marcela Vegetti, Horacio Leone, “Ontología para el Análisis Forense de Correo Electrónico”, CoNaIISI 2014 - ISSN: 2346-9927 - Página 1008-1018

[6] 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

[7] Rivetti E., Parra H.B., “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.

[8] Gallo Beatriz P. de, Vegetti Marcela, Leone Horacio, “Avances en la Construcción de una Ontología para el Análisis Forense de Correo Electrónico”, presentado al VI CIIDDI (Congreso Iberoamericano de Investigadores y Docentes de Derecho e Informática), en revisión.