APLICACIÓN DE LAS TECNOLOGÍAS SEMÁNTICAS A LA FORENSIA DIGITAL: ONTOLOGÍA DEL CORREO ELECTRÓNICO Y SU TRAZABILIDAD PARA EL ANÁLISIS FORENSE

Beatriz Parra de Gallo. bgallo@ucasal.edu.ar
IEsIIng – Instituto de Estudios Interdisciplinarios de Ingeniería (Universidad Católica de Salta), Salta, Argentina

Marcela Vegetti. mvegetti@santafe-conicet.gov.ar
INGAR – Instituto de Desarrollo y Diseño (Conicet/UTN), Santa Fe, Argentina

MODALIDAD: Ponencia Presencial
LÍNEA TEMÁTICA: Ingeniería y Tecnología

RESUMEN

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

En sus inicios la Forensia Digital hizo su aparición en el ámbito de la justicia con el surgimiento de las primeras pruebas digitales, requiriendo mayormente la presencia de peritos para constatar la veracidad de esas pruebas que técnicamente no representaban grandes desafíos (constatar la existencia de un correo del cual se adjuntaba su impreso, o verificar la funcionalidad de un software o el contenido de un archivo por ejemplo), pero a partir de la popularización de Internet, la Forensia Digital ha entrado en una crisis producto del impacto de dos elementos que marcan la época actual de la tecnología informática: la masividad de los datos y la multiplicidad de plataformas tecnológicas. Las investigaciones encontradas señalan que las herramientas forenses actuales presenta varios desafíos al respecto, involucrando no solo los modelos de "visibilidad y búsqueda" de los datos obtenidos sino también la falta de integración de estrategias (como la ingeniería reversa) para reducir tiempos y costos.

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

Aunque la Forensia Digital avanzó en concordancia con la tecnología, es necesario aún trabajar un aspecto que no es propiamente del ámbito tecnológico y que genera un conjunto de interrogantes que impactan en gran medida en los resultados que se obtienen, i.e., la interpretación de esos resultados. Diversos autores, entre ellos Harichandran señalan la importancia de mejorar las instancias de comunicación entre los técnicos y los profesionales del derecho, incrementando la accesibilidad y usabilidad de las herramientas de análisis forense para facilitar su interpretación por parte de los no técnicos. El volumen de datos que se obtiene al realizar el análisis forense debe ser interpretado a la luz de la pesquisa. Cualquiera sea el componente sobre el cual se realiza el análisis forense (celulares, correo electrónico, discos, etc) se generará una cantidad de información técnica que es necesario insertar en el conjunto de pruebas documentales de la causa judicial, colocándolo en un estadio de lectura que facilite la interpretación de esos datos técnicos por parte de los profesionales de la criminalística y el derecho. Resulta necesario contar con un marco de referencia basado en la conceptualización formal del universo de discusión que permita extraer la información tecnológica y establecer una correspondencia unívoca con una descripción de evidencia digital. En particular, las ontologías resultan una herramienta universal o pluridisciplinar para facilitar el análisis de la prueba documental, por parte de todos los actores (abogados, jueces, investigadores y peritos).

Por otra parte, la fuerza probatoria del dictamen pericial se sustenta en "*los principios científicos o técnicas en que se funda, en concordancia con la aplicación de las reglas del arte"*, es decir, es importante sostener la actividad pericial mediante la aplicación de herramientas y técnicas probadas científicamente. En el caso de las evidencias digitales, el perito recurre a una revisión manual de la misma, o a la utilización de herramientas informáticas ad-hoc, pero de las cuales no se conoce el marco científico que sustenta la funcionalidad de las mismas. De allí la necesidad de generar herramientas construidas desde un contexto metodológico y científico que sustente los resultados que se obtienen.

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

Como resultado de esta investigación, desarrollada en el marco del Proyecto de investigación denominado "Aplicación de las Tecnologías Semánticas a la Forensia Digital: Etapa 1", con asiento en el Instituto de Estudios Interdisciplinarios de Ingeniería de la Facultad de Ingeniería de la Universidad Católica de Salta, se ha desarrollado ***OntoFoCe*** una ontología que representa el proceso de transmisión del correo electrónico permitiendo derivar su trazabilidad y respondiendo a 21 preguntas de competencia relativas a los puntos de pericia más habituales en forensia de correos electrónicos. Y a partir de esta ontología se construyó ***ObE Forensic*** una aplicación web destinada al análisis forense de correos

electrónicos. Particular atención se prestó a las cuestiones de seguridad informática de la aplicación, específicamente con la definición de una política de control de acceso, de un recurso de uso obligatoria para aceptación de los términos y condiciones de uso de la aplicación, y con la condición de almacenamiento volátil del espacio destinado al procesamiento de los datos que se ingresan en la aplicación (es decir, los datos se borran cuando el usuario cierra sesión).
Mayores detalles sobre la estructura, funcionalidad y resultados de esta aplicación se puede leer en http://www.risti.xyz/issues/risti32.pdf (páginas 17-32, DOI: 10.17013/risti.32.17–32), una publicación expresa sobre ObE Forensics, presentada por los mismo autores.
Por último, cabe mencionar que una versión prototipo de la aplicación está publicada en https://digilab.ucasal.edu.ar/login (Se puede ingresar con las siguientes credenciales, usuario: beagallo@gmail.com; clave: bgallo). Por cuestiones de seguridad, la aplicación requiere de la identificación del usuario mediante credenciales que se le entregan, previa solicitud de las mismas, sin costo involucrado.

PALABRAS CLAVE: ontología; forensia digital; correo electrónico

ABSTRACT

The different security forces -both military, judicial and political- are concerned about facing the fight against crime from the technological point of view, that is, with an increasingly worrying look at the use of technology to commit a crime. This resulted in the generation of an own space, within the computer's science, called Digital Forensics. In 2001 the Digital Forensics Research Conference defined the "Digital Forensics" as "*The use of scientifically obtained and proven methods for the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from sources digital for the purpose of facilitating or favoring the reconstruction of criminal acts or for the prevention of unauthorized actions that are considered harmful for planned operations*".
In its beginnings Digital Forensics appeared in the field of justice with the emergence of the first digital tests, requiring mostly the presence of experts to verify the veracity of those tests that technically did not represent major challenges (note the existence of a mail which was attached your printed, or verify the functionality of a software or the content of a file for example), but after the popularization of the Internet, Digital Forensics has entered into a crisis product of the impact of two elements that mark the current era of computer technologies: the massiveness of the data and the multiplicity of technological platforms. The investigations found indicate that current forensics tools present several challenges in this regard, involving not only the models of "visibility and search" of the data obtained but also the lack of integration of strategies (such as reverse engineering) to reduce time and costs.
In the forensic context, it is very important to link the data from the meaning of each thing. It is not only about "finding the digital evidence", but about interpreting it in the context of the situation, linking it with the rest of the components of the investigation (physical evidence, interrogation, legal and procedural framework of the case, etc.). So it is essential to advance in digital forensics from the perspective of semantics -as a binding element of

all the components of the system- as well as from a referential framework that can interpret it -an ontology-.
Although the definition most referenced in the literature is that of "*an ontology is an explicit specification of a conceptualization*", it is worth detailing the concept a little more, taking what Reuver said about "*An ontology is the conceptual and terminological description of a shared knowledge about a specific domain. Leaving aside the formalization and interoperability of applications, this is nothing more than the main competence of the term: to make improvements in communication using the same system in the terminological and conceptual* ".
Although the Digital Forensics advanced in accordance with the technology, it is still necessary to work on an aspect that is not strictly technological and that generates a set of questions that greatly impact on the results obtained, ie, the interpretation of those results. Several authors, including Harichandran, point out the importance of improving the communication mechanisms between technicians and legal professionals, increasing the accessibility and usability of forensic analysis tools to facilitate their interpretation by non-technicians. The volume of data obtained when performing the forensic analysis must be interpreted in light of the investigation. Whatever the component on which the forensic analysis is carried out (cell phones, e-mail, disks, etc.), a quantity of technical information will be generated that must be inserted in the set of documentary evidence of the court case, placing them in a stadium that facilitates the interpretation of these technical data on the part of professionals in criminology and law. It is necessary to have a frame of reference based on the formal conceptualization of the universe of discussion that allows to extract the technological information and establish a univocal correspondence with a description of digital evidence. In particular, ontologies are a universal or multidisciplinary tool to facilitate the analysis of documentary evidence, by all actors (lawyers, judges, researchers and experts).
On the other hand, the probative force of the expert opinion is based on the scientific principles or techniques on which it is based, in accordance with the rules of art, that is, it is important to sustain the expert activity through the application of scientifically proven tools and techniques. In the case of digital evidence, the expert uses a manual review of it, or the use of ad-hoc computer tools, but of which the scientific framework that supports the functionality of the same is not known. Hence the need to generate tools built from a methodological and scientific context that supports the results obtained.
The digital documents that can be proposed as "evidence" in a court case are very diverse, in terms of structure, format and origin. In order to limit the study of the application of semantic technologies to Forensia Digital, e-mail was considered as an object of study, and in particular, these technologies were applied to formulate an ontology that allows incorporating the forensic analysis of e-mails as digital evidence not repudiable.
As a result of this research, developed within the framework of the research project called "Application of Semantic Technologies to Digital Forensics: Stage 1", with a seat at the Institute of Interdisciplinary Engineering Studies of the Faculty of Engineering of the Universidad Católica de Salta, an ontology called ***OntoFoCE*** has been developed that represents the process of transmission of e-mail allowing to derive its traceability and answering 21 questions of competence related to the most common skill points in email forensic. And from this ontology ***ObE Forensic*** was built a web application for the forensic analysis of e-mails.

Particular attention was paid to the issues of security of the application, specifically with the definition of an access control policy, with a resource of access compulsory for acceptance of the terms and conditions of use of the application, and with the condition of volatile storage of the space destined to the processing of the data entered in the application (i.e. the data is cleared when the user logs off).
Further details on the structure, functionality and results of this application can be read in http://www.risti.xyz/issues/risti32.pdf (pages 17-32, doi: 10.17013/RISTI.32.17-32), a specific publication on ObE Forensics, submitted by the same authors.
Finally, it should be noted that a prototype version of the application is published in https://digilab.ucasal.edu.ar/login (You can login with the following credentials, User: beagallo@gmail.com, Pass: bgallo). For security reasons, the application requires the identification of the user using credentials that are delivered, upon request, at no cost involved.

KEYWORDS: ontology; digital forensics; e-mail