Advances in the Application of Ontologies in the Area of Digital Forensic Electronic Mail

B. Parra, M. Vegetti, H. Leone

***Abstract*—This article presents a descriptive review of the research published in the last five years to identify areas of unavailability in the study of Digital Forensics problems. Particularly, it is important to define the state of the art related to the application of ontologies, especially in the forensics of emails. The following objectives of the review are proposed: identify and study the most up-to-date research contributions on Ontologies and Digital Forensics; establish the gaps in current research reltated with the application of Ontologies to Digital Forensics; and correlate these works from attributes of proximity (or distance) with the application of ontologies to the forensic analysis of emails. In addition, a systematic method is defined to select the research works that are considered of interest for this review. It is expected that it will lead to the identification of gaps in the investigation of characteristic problems in digital forensic analysis, and the definition of an updated theoretical framework linked to the forensic analysis of emails with the application of ontologies.**

***Index Terms*— Ontology, Digital Forensic, Email, Traceability.**

I. Introducción

LOS diferentes estamentos de seguridad –tanto militares como judiciales y políticos- se preocupan por encarar la lucha contra el crimen desde la óptica tecnológica, con una mirada cada vez más preocupante sobre el uso de la tecnología para delinquir. A ello puede colaborar la *Forensia Digital* definida 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*" [1].

Desde el desarrollo actual de la Forensia Digital son varios los desafíos a resolver. Garfinkel [2] señala como importante la falta de integración de las herramientas con estrategias de procesos (como la ingeniería reversa) para reducir tiempos y costos, así como las dificultades para mostrar los resultados en términos de informes que puedan ser fácilmente interpretables para los profesionales de la justicia.

El volumen de datos encontrados en el análisis forense también debe considerarse al momento de la presentación del informe pericial. Resulta necesario seleccionar, identificar y mostrar los datos de interés para la causa. Habitualmente se requieren procesos de abstracción y resumen que –por sí mismas- las herramientas forenses no los proveen, quedando en manos de la capacidad del perito mostrar los datos relevantes ajustados al individuo o contexto de la causa, con posibilidades de integrarlos al conjunto de pruebas documentales y materiales que pudieran existir. Ya no es posible considerar la forensia digital como un proceso técnico de búsqueda de datos digitales, sino que es necesario considerarla desde la óptica de la semántica de esos datos, en el contexto en que se encuentran, darles un significado que los vincule con el resto de los componentes de la causa judicial, y que sea interpretable por todos los participantes (jueces, abogados, otros peritos, policía, etc.). Resulta necesario generar un marco referencial en el que todos estos actores interactúen entre sí y puedan interpretar los resultados forenses a la luz de la causa en Litis. Las ontologías son consideradas como modelos conceptuales de referencia, ya que capturan los aspectos fundamentales del dominio y hacen explícitos los compromisos ontológicos subyacentes en dicho dominio. Estas características soportan las tareas de comunicación, comprensión del dominio, resolución de problemas y negociación de significado entre usuarios humanos[3]. Es por ello, que las ontologías resultan una herramienta adecuada para generar el mencionado marco de referencia.

Reuver et al. [4] que definen una ontología como "*...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*".

Una ontología que represente formalmente el proceso de transmisión de un correo electrónico permite derivar de esta representación la trazabilidad del mismo, demostrando con ello que, si es posible establecer el origen del correo a partir del correo recibido, entonces se evita el "no repudio" de dicha evidencia digital, o sea, el desconocimiento de la misma por parte de la contraparte en el juicio.

B. Parra, Facultad de Ingeniería, Instituto de Estudios Interdisciplinarios de Ingeniería (IEsIIng), Universidad Católica de Salta (UCASAL), Salta, Argentina, beatriz.gallo@ucasal.edu.ar.

M. Vegetti, Instituto de Desarrollo y Diseño (INGAR), Consejo Nacional de Ciencia y Técnica (CONICET), Universidad Tecnológica Nacional (UTN), Santa Fe, Argentina, mvegetti@santafe-conicet.gov.ar.

H. Leone, Instituto de Desarrollo y Diseño (INGAR), Consejo Nacional de Ciencia y Técnica (CONICET), Universidad Tecnológica Nacional (UTN), Santa Fe, Argentina, hleone@santafe-conicet.gov.ar.

En el caso particular de los correos electrónicos, el análisis forense se realiza sobre la cabecera del mail, obteniéndose un volumen de datos técnicos de difícil interpretación para el lego, y deben seleccionarse y mostrarse en el marco del resto de las pruebas de la causa judicial, ofreciendo un informe técnico que permita la interpretación de los resultados a la luz de la causa, por parte de los profesionales de la criminalística y el derecho. Se requiere mucho más que la identificación de una dirección IP (Internet Protocol) del correo electrónico. Hoy en día se exige que estos datos se presenten sistemáticamente y semánticamente en el marco de la causa judicial, en el mismo espacio de análisis que el resto de los elementos probatorios. Y en particular, las ontologías resultan una herramienta pluridisciplinar para facilitar el análisis de la prueba documental, por parte de todos los actores (abogados, jueces, investigadores y peritos). En este contexto resulta de interés realizar un estudio sistemático del estado del arte en ambas áreas: ontologías y forensia digital, y particularmente sobre la aplicación de ontologías a la forensia de correos electrónicos.

Respecto del tipo de estudio, se pretende efectuar una investigación exploratoria para identificar el estado del arte sobre la aplicación de las ontologías a la Forensia Digital, particularmente en la forensia de correos electrónicos, realizando un estudio crítico y ajustando el alcance del mismo a los objetivos propuestos. Éstos últimos se mencionan a continuación:

(i) Identificar y estudiar los aportes investigativos más actualizados sobre Ontologías y Forensia Digital.
(ii) Establecer las áreas de vacancia sobre la aplicación de las ontologías a la Forensia Digital
(iii) Relacionar los trabajos desde atributos de cercanía (o distancia) con la aplicación de ontologías para el análisis forense de correos electrónicos.

Este trabajo se organiza en secciones, la Sección II describe el método de revisión bibliométrica, incluyendo el análisis de varias metodologías, la definición de la más adecuada y la descripción de las tres fases de búsquedas que se establecen: Búsqueda Inicial, Preselección por Conteo de Palabras Claves y Selección Final. En la Sección III se aborda el análisis y discusión de los resultados, recurriendo a la clasificación de los trabajos en tres grupos: por área temática, por objeto de estudio y trabajos referidos a la forensia de correos electrónicos. En la Sección IV se discuten los aportes encontrados en la revisión, y por último la Sección V incluye las conclusiones obtenidas.

II. Método De Revisión Bibliométrica

A fin de realizar una revisión bibliográfica ordenada y ajustada a norma, se analizaron diversas metodologías de revisiones de trabajos científicos, identificando los criterios más útiles para la temática en estudio.

Kitchenham [5] definió una metodología para la revisión de la literatura vinculada a la ingeniería de software, basándose en métodos de revisión provenientes de la medicina. Su principal aporte está en la definición del paradigma basado en la evidencia, que promueve la evaluación objetiva y la búsqueda de resultados empíricos relevantes sobre un tema de investigación. Siguiendo esta misma línea de generar métodos para la revisión bibliográfica en la ingeniería de software, Biolchini [6] presenta su método de tres pasos, que permite pasar de los conceptos hacia los estudios que pueden proporcionar evidencia sobre el tema en cuestión (fase 1), luego se analizan comparativamente los contenidos de esas publicaciones, para generar nuevo tipo de evidencia si fuera posible (fase 2), y por último se arriba a las conclusiones, que podría significar la obtención de nuevos conocimientos.

También se consultó la metodología propuesta por Grant [7], en la que se realiza un análisis comparativo entre 14 tipos diferentes de revisión bibliográfica (siempre en el área de la Medicina), identificado por una parte los *tipos* de revisiones (crítica, literaria, mapeo sistemático, meta-análisis, estudios mixtos, de visión general, revisión por alcance, etc.), mostrando para cada una la conveniencia de utilización.

Por su parte Velásquez [8] sintetiza las metodologías propuestas por tres autores (Kitchenham, Sorrell y Tranfield) identificando las tres fases más comunes involucradas en una revisión: planeamiento (en la que se propone la justificación, motivación y diseño del protocolo de búsqueda), la ejecución (que incluye los procesos de búsqueda, selección, evaluación de calidad, extracción y síntesis de resultados) y el reporte final de la revisión realizada.

Para Medina López et al. en [9] la búsqueda bibliográfica ajustada a norma implica considerar al menos 5(cinco) fases de trabajo: 1) Identificación del campo de estudio y del período a analizar; 2) Selección de las fuentes de información; 3) Realización de la búsqueda (qué, dónde y cómo); 4) Gestión y depuración de los resultados de la búsqueda; y 5) Análisis de los resultados. Durante el desarrollo de estas etapas se deben considerar además criterios de éxito, tales como: establecer con claridad el objetivo que se persigue, documentar el proceso, deifnir parámetros de cualificación comparables, entre otros. En [10] se propone definir restricciones para la inclusión de trabajos a revisar, con el fin de limitar el alcance del estudio en función de la estructura de las publicaciones (descripción de un experimento por ejemplo). Y también proponen definir criterios de exclusión referidos al tipo de trabajo a considerar (tesis, patentes, etc.).

La metodología propuesta en [10] formula el proceso de revisión en fases claramente identificadas permitiendo la generación de un marco de trabajo ordenado que ayuda en todo el proceso de revisión, orientando al investigador en cada fase, evitando desvíos que insumen tiempo y esfuerzo. Se observa también, que las definiciones de los objetivos planteados por la metodología para cada una de sus fases están claramente especificadas, siendo de gran ayuda para quien se inicia en la revisión bibliográfica con el objetivo de conformar el estado del arte sobre la temática de interés. Asimismo, la metodología mencionada ha sido aplicada en casos donde se aborda el estudio de dos áreas temáticas distintas que deben estudiarse tanto individualmente como fusionadas entre sí, situación que se da en la revisión de investigaciones sobre Forensia Digital y Ontologías. Por otra parte, la propuesta presentada en [10] resuelve de manera formal la definición del

alcance de la revisión, basado en las restricciones iniciales y los criterios de exclusión para delimitar el universo de publicaciones existentes y acotar la selección de las mismas, para llegar a resultados concretos. Esta característica, permite completar la propuesta de Medina López et. al [9].

Teniendo en cuenta lo expresado en el párrafo previo se optó por tomar como guía la metodología propuesta por Medina López et al. en [9] y lo dicho en [10] para realizar una revisión del estado del arte de la aplicación de las ontologías en problemáticas de forensia digital de correos electrónicos. Así, se propuso un método de revisión sistemática para el estudio que se introduce en el presente trabajo basado en las siguientes fases: A) Definición del Marco de Estudio y Alcance de la Revisión y B) Procesos de Búsqueda y Selección.

A. Definición del Marco de Estudio y Alcance de la Revisión

A fin de cumplir con los objetivos propuestos para la revisión, resulta necesario definir los atributos o palabras claves que delimiten el *marco de estudio*, siendo términos con suficiente fuerza como para guiar los procesos de búsqueda.

Se partió de la conjunción de dos temáticas principales: las ontologías y la forensia digital. Ambas áreas, de amplísimo desarrollo por sí solas, se fusionan en trabajos de investigación particulares que cuentan cada uno de ellos con sus propios objetivos, en los que –en la generalidad- se observa la aplicación de ontologías en la resolución de problemas de la forensia digital. Aun considerando el marco teórico específico de estudios de aplicación de ontologías a la forensia digital, se definió un siguiente nivel de detalle centrado en la entidad u objeto de la pericia. Particularmente, interesaba identificar los últimos aportes referidos a los métodos, herramientas y artefactos forenses (se denomina así a los distintos componentes en los que reside la evidencia digital, sea éste un dispositivo de hardware o software), vinculados con correos electrónicos.

Se definieron los criterios de búsqueda mediante reglas de decisión enfocadas al marco teórico indicado. Los términos iniciales para la búsqueda son los siguientes: *forensia, ontologías, correo electrónico y cabecera del correo electrónico.* De la conjunción de estos términos se puede deducir los *criterios de búsqueda*:

- CB1: "*ontology AND forensic AND electronic mail*": que permite ahondar en la aplicación de ontologías a la forensia digital de correos electrónicos;
- CB2: "*forensic AND email header*": para identificar los trabajos relacionados a forensia de correos electrónicos en los que se aborden los métodos y herramientas utilizadas a partir del análisis forense de la cabecera del correo electrónico,

La literatura sobre revisión bibliográfica aconseja realizar una prueba piloto de los criterios de búsqueda seleccionados, de modo de afinarlos y adecuarlos a conveniencia. El objetivo de la prueba piloto es tener pocos falsos positivos (artículos que han sido seleccionados por la búsqueda automática pero que realmente no responden a los objetivos del estudio) y pocos falsos negativos (artículos no detectados por la estrategia de búsqueda establecida pero que son de provecho para el estudio). Por ejemplo: la palabra *email* no es una palabra clave exitosa por sí misma ya que las búsquedas automáticas devuelven textos que contienen *email* como referencia del correo electrónico de los autores, y no como tema de estudio del artículo.

Así, se decidió agregar las palabras *forensic* u *ontology* para orientar la indagación. Por otra parte, se observó que el buscador debía ajustarse particularmente en lo siguiente: a) la publicación puede contener la palabra clave *ontology* o su plural (*ontologies),* el término *forensic* puede figurar bajo un sinónimo (*investigation)* y la palabra clave *electronic mail* puede resumirse como *email, mail o e-mail.* Atendiendo a estas consideraciones se definieron los criterios de búsqueda señalados en la Tabla I.

TABLA I
CRITERIOS DE BÚSQUEDA AUTOMÁTICA

Criterio	Regla de Búsqueda
N° 1	ontology AND forensic AND email AND year>=2014 AND year<=2018
N° 2	forensic AND header email AND year>=2014 AND year<=2018

Se recurrió a las siguientes *fuentes de información*: revistas científicas especializadas y artículos publicados en congresos sobre las temáticas de estudio. La selección de revistas científicas especializadas no es un tema menor, debe cuidarse los aspectos de reconocimiento de la publicación en el contexto científico y factor de impacto o medida de la importancia de la revista. Las actas de congresos son de utilidad cuando se trabaja en áreas de investigación emergentes, cual es el caso de la Forensia Digital en general, y de la Forensia de Correos Electrónicos en particular. Así, en el estudio realizado se consideraron las siguientes bibliotecas electrónicas: IEEE Xplore Digital Library, ScienceDirect, Scopus, Scholar Google, The Journal of Digital Forensics, Security and Law (JDFSL), y ACM Library.

En base a estas consideraciones, y con el objetivo de fijar los *límites de la revisión*, se definió como restricciones iniciales de la búsqueda las siguientes:

- R1: Se incluyen trabajos referidos a la aplicación de las ontologías para definir o mejorar *metodologías de trabajo, herramientas forenses y análisis forense de artefactos forenses*.
- R2: se considera un espacio temporal de cinco años, tomando el período 2014-2018.

Los criterios de exclusión definidos son los siguientes:

- CE1: Se excluyen del estudio los libros, capítulos de libros, cartas, notas, tesis de grado o posgrado y patentes.
- CE2: Se excluyen publicaciones impresas en papel, considerando solo textos electrónicos, y de éstos, aquellos que cuentan con acceso público o a los que se puede acceder mediante las vías institucionales disponibles.
- CE3: Se excluyen las publicaciones escritas en otros idiomas que no sean en inglés.
- CE4: En caso de que un mismo estudio se repitan en dos o más búsquedas, se lo considera una única vez.

- CE5: Se excluyen los trabajos de los que no puede acceder al texto completo del mismo.
- CE6: Se excluyen artículos que no estén en formato de texto portátil (PDF) y que superen los 15Mb de tamaño.

B. *Procesos de Búsqueda y Selección*

El proceso de búsqueda y selección se dividió en tres fases claramente identificables: Búsqueda Inicial, Preselección por Conteo de Palabras Claves y Selección Final.

La fase de *Búsqueda Inicial* consistió en el uso de los buscadores automáticos de las bibliotecas digitales visitadas, activando las palabras claves definidas en la Tabla I como parámetros de las consultas. Se debe destacar que en el caso de Scholar Google el criterio de búsqueda genera una cantidad masiva de artículos obtenidos, por ello se decidió tomar los 100 primeros trabajos, entendiendo que el resto no es de interés para el estudio aprovechando el rankeo por *relevancia* que Scholar Google define para su algoritmo de búsqueda.

Una vez realizadas las búsquedas automáticas de nivel inicial, o sea, aquellas generadas por los buscadores de las propias bibliotecas parametrizados según los criterios señalados en la Tabla I, se obtuvieron 1091 trabajos.

La fase dos, denominada *Preselección por Conteo de Palabras Claves,* toma como insumo los 1091 textos encontrados en la fase de Búsqueda Inicial, y se realizó la preselección por conteo de palabras claves, aplicando el algoritmo que se muestra en la Figura 1.

A continuación se explica el algoritmo de conteo de palabras claves, diseñado expresamente para este caso.

El procedimiento -de carácter semiautomático- se basa en un proceso ETL (Extraction, Transformation and Load), en el cual se uniformizan los metadatos de las publicaciones, respetando el formato que ofrece cada biblioteca, y con una carga manual cuando la misma no permite exportar la búsqueda.

Primeramente se debe conformar una hoja de cálculo con los metadas de las publicaciones (Título, Autores, Palabras Claves, Año de Publicación, etc.), que -de acuerdo a las posibilidades de los buscadores de las bibliotecas consultadas- se pueden exportar en formato CSV desde la página de resultados obtenidos y de allí a una hoja de cálculo, o bien los metadatos se cargan manualmente en una hoja de cálculo.

La siguiente tarea consiste en acceder a los archivos PDF de las publicaciones seleccionadas y contar la cantidad de palabras claves que figuran en el texto completo del trabajo.

Para esta actividad se desarrolló una aplicación web denominada *KeyWordFinder* (disponible en https://digilab.ucasal.edu.ar/keywordfinder) que permite la carga de un texto en formato portátil (PDF) y realiza un proceso de barrido del texto contando el número total de veces que aparece una palabra clave previamente ingresada. Se considera que un trabajo aborda la temática de estudio cuando la palabra clave figura un mínimo de tres veces en el texto.

Las palabras claves que se cuentan son las mismas que las propuestas en la Tabla I y, además, deben figurar *todas* las palabras del criterio en una cantidad de tres o más.

En particular, la palabra *mail* y sus sinónimos (email, e-mail, electronic mail), pueden figurar como dato de identificación de los autores, por ello, se incluye aquellos trabajos que superan en cinco el conteo de esa palabra.

Por último, se aplica una función lógica para indicar la condición de *pre-seleccionado/descartado* de cada trabajo, registrando esta situación en la hoja de cálculo de metadatos.

La revisión de los textos en formato portátil se realiza para los dos criterios de búsqueda de la Tabla I, considerando todas las bibliotecas señaladas, logrando 88 publicaciones preseleccionadas.

El último paso del procedimiento, denominado *Selección Final,* consiste en leer los trabajos preseleccionados en la fase anterior para confirmar si efectivamente resultan de interés para el estudio.

De este modo, se analiza cada trabajo teniendo presente su identificación con los objetivos de la revisión, se revisa el cumplimiento de los criterios de exclusión definidos y se confirma la condición final de *seleccionado/descartado,* incluyendo la identificación de los falsos positivos ([11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] y [22]).

Por último, en esta fase de revisión exahustiva se aprovecha para identificar el área temática de cada trabajo, su objeto de estudio y separar los que tratan sobre correos electrónicos.

Así, de los 1091 trabajos hallados en la búsqueda inicial, se pre-seleccionaron 88 aplicando el algoritmo de conteo de palabras claves, y de éstos últimos, se seleccionaron 76 a partir de la lectura individual de los textos.

El resultado final de la selección, se indica en la Tabla II que detalla –por fuente bibliográfica- la cantidad de trabajos encontrados en cada una de las tres fases de selección. La tabla se completa con la referencia bibliográfica de cada publicación seleccionada.

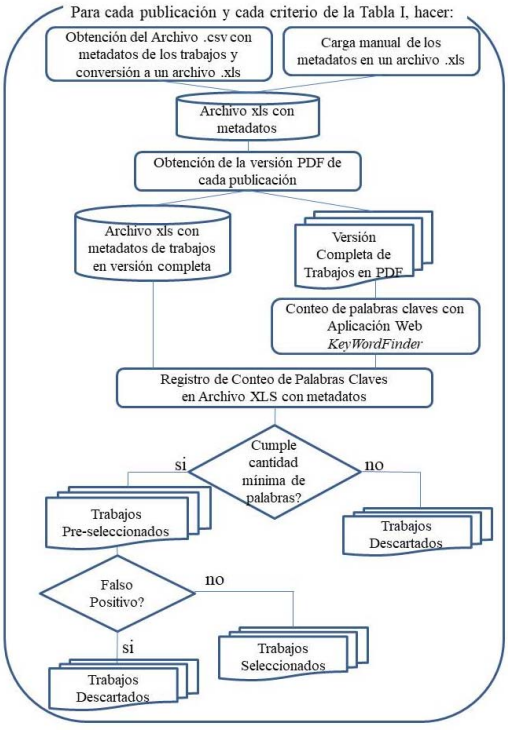

Fig. 1. Algoritmo de Preselección por Conteo de Palabras Claves.

Tabla II
RESULTADOS DEL PROCEDIMIENTO DE BÚSQUEDA

Publicación	Criterios	Búsqueda Inicial	Preselección por Conteo	Selección Final	Referencias de los trabajos seleccionados
IEEE Xplore Digital Library	N° 1	56	4	4	[23], [24], [25], [26]
	N° 2	185	25	24	[27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50]
ScienceDirect	N° 1	104	2	2	[51], [52]
	N° 2	104	21	18	[53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70]
Scopus	N° 1	120	2	2	[71], [72]
	N° 2	5	1	1	[73]
Scholar Google	N° 1	84	8	5	[74], [75], [76], [77], [78]
	N° 2	18	8	8	[79], [80], [81], [82], [83], [84], [85], [86]
The Journal of Digital Forensics, Security and Law (JDFSL)	N° 1	62	3	3	[87], [88], [89]
	N° 2	78	8	5	[90], [91], [92], [93], [94]
ACM Library	N° 1	77	4	2	[95], [96]
	N° 2	198	2	2	[97], [98]
		1091	88	76	Total: 76 trabajos

III. Análisis Y Discusión De Los Resultados

A nivel descriptivo, se analiza el conjunto de los trabajos seleccionados, para luego abordar cada uno en particular.

Si se tiene en cuenta el *año de publicación*, se observa que los trabajos se distribuyen de manera pareja en el quinquenio considerado (ver Figura 2), y no se encuentra un incremento de investigaciones de un año a otro como sería de esperarse a fin conformar un marco científico que sustente la Forensia Digital.Respecto de la *temática principal* de cada publicación, se observó que los trabajos se enfocan hacia Tráfico de Redes, Ontologías, Data Mining, Big Data, Seguridad Informática y otras áreas diversas como lenguaje natural y máquinas virtuales entre otros. La Figura 3 muestra la distribución porcentual de las publicaciones por área temática.

También se pueden clasificar los trabajos en función de su aporte al estudio de los *métodos de análisis forense, las herramientas y los artefactos forenses*. En ese sentido se detectó que el 45% de los estudios tratan sobre métodos para el análisis forense, 38% abordan el tema de herramientas utilizadas para dicho análisis y el 17% restante trata sobre forensia en dispositivos o artefactos forenses (Figura 4).

En particular, sobre *correos electrónicos*, se encontró que 27 trabajos estudian y describen el análisis forense de correos electrónicos, y de éstos, cinco analizan ataques cibernéticos utilizando correos electrónicos. En cinco se aborda el estudio de la cabecera del correo electrónico y de éstos sólo en uno se estudia expresamente el proceso de transmisión.

En función del resumen enunciado, se pueden describir los aportes más importantes de cada trabajo, en base a tres grupos:

- *Análisis por área temática*: considerando los estudios que abordan el Tráfico de Redes; Ontologías; técnicas y herramientas de Data Mining; o de Big Data; Seguridad Informática y Otras Áreas de la informática.
- *Análisis por objeto de estudio*: identificando los estudios sobre Métodos y Herramientas de Análisis Forense y sobre Forensia de Dispositivos y/o Artefactos.
- *Análisis de trabajos referidos a Forensia de Correos Electrónicos,* detallando los estudios sobre ataques cibernéticos vía correo electrónico, aquellos que abordan la cabecera del correo electrónico y los que consideran el proceso de transmisión.

Se detallan los trabajos incluidos en cada apartado.

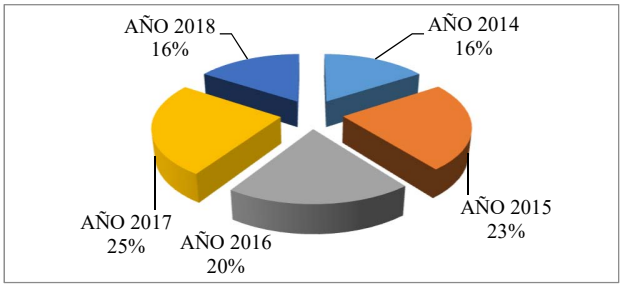

Fig. 2. Distribución de Trabajos por Año de Publicación.

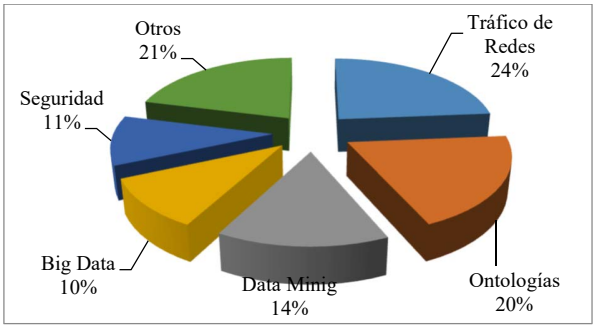

Fig. 3. Distribución de Trabajos por Área Temática.

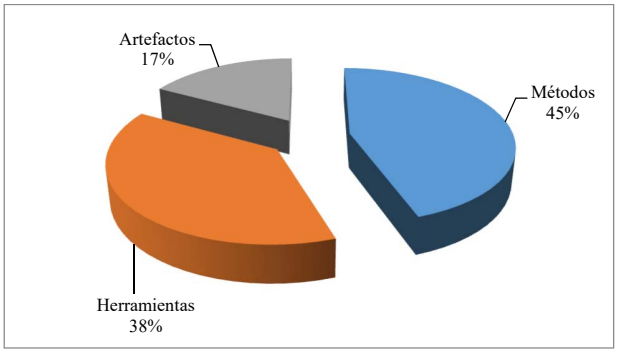

Fig. 4. Distribución de Publicaciones por Objeto de Estudio.

A. Análisis por Área Temática

En lo referente a *Tráfico de Redes* se destacan estudios que se pueden agrupar según dos enfoques: ataques cibernéticos y forensia de redes de datos.

Se encontraron estudios referidos a *ataques cibernéticos* en las siguientes investigaciones: en [39] los autores proponen TRIG un sistema de almacenamiento de tráfico y generación de información relacionada, que puede almacenar el tráfico de red de 20 Gbps en tiempo real; en [51] se presenta un enfoque de colaboración para la gestión de la información de incidentes cibernéticos en procesos industriales; en [91] se abordaron los sistemas de prevención de fuga de datos (DLPS) que analizan el tráfico de red y alertan en caso de una fuga de datos. El trabajo [84] analiza los ataques cibernéticos con ayuda de herramientas forenses digitales como *WinHex* mientras que el trabajo [38] describe una metodología de análisis de paquetes de red para detectar espionaje industrial a partir del tráfico de red de una organización.

En referencia a la *Forensia de Redes de Datos,* la investigación [81] presenta NetFox, una herramienta para el análisis forense de redes con posibilidades avanzadas de reconstrucción de aplicaciones y técnicas avanzadas para detectar actividades ilegales o no autorizadas; por su parte, el trabajo [60] trata sobre el análisis forense completo del tráfico de red en sistemas SCADA; en [57] se estudia la herramienta de intercambio de archivos P2P BitTorrent versión 2.x y propone métodos para recuperar potenciales evidencias forense en dispositivos que ejecutan Windows, Mac OS, Ubuntu, iOS y dispositivos Android.

Considerando ahora la aplicación de *Ontologías*, se encontraron trabajos referidos a dos enfoques: trabajos que presentan ontologías expresamente desarrolladas y aplicación de las ontologías como marco de trabajo.

Los trabajos que presentan *desarrollo de ontologías* son los siguientes estudios: la herramienta CyBox propuesta en el estudio [52] basada en la ontología DFAX permite representar e intercambiar información forense digital, incorporando aspectos procesales de la forensia, como ser la cadena de custodia, manejo de casos y procesamiento forense, mientras que en [72] se propone DF Disciplines una ontología para categorizar las disciplinas forenses digitales, así como algunas metodologías que puedan ofrecer orientación en diferentes áreas de la forense digital. Mediante la ontología OBMO (Online Banking Malware Ontology) que se aborda en [25], que a su vez se basa en ontologías prediseñadas OSCAF (Open Semantic Collaboration Architecture Foundation) se define un método para modelar las organizaciones criminales intervinientes e identificar a los desarrolladores de malware; esta misma ontología OBMO se trabaja en [75] para la investigación de malware bancario en línea con nuevos enfoques que correlacionen la evidencia digital. En [88] se describe una ontología que cataloga herramientas forenses digitales comunes mientras que el estudio [77] presenta una revisión exhaustiva de las soluciones y modelos existentes para recopilar información y utilizarla para resolver crímenes mediante una ontología prototipo llamada SC-Ont mientras que el estudio [95] describe una ontología para servicios web (Fi4SOA) propuesta para fusionar propiedades forenses con requisitos comerciales en la fase de diseño del servicio. El estudio [23] presenta PSO como marco ontológico para mejorar la seguridad física y la detección de amenazas internas; en este mismo espacio [74] presenta INFORENSIC Ver2 para la gestión del conocimiento en seguridad y análisis forense digital. También se analizó el trabajo [89] que estudia los teléfonos celulares, proponiendo la ontología F-DOC para modelar cada componente del contenido de un teléfono inteligente con el propósito de realizar el análisis forense. En el estudio [71] se describe una ontología para representar el direccionamiento semántico de una lista de distribución de correo electrónico.

Otros trabajos abordan la *aplicación de las ontologías* en diferentes contextos: en [24] se describe un marco ontológico que recupera y muestra modelos de evidencias obtenidos a través de diferentes herramientas forenses, presentando un sistema capaz de añadir afirmación semántica a los datos generados por las herramientas de análisis forense durante los procesos de extracción. Los autores del trabajo [63] proponen una taxonomía de ataques de seguridad en la nube y las posibles estrategias de mitigación; mientras que en [76] se describe un marco para integrar la Ingeniería de Requisitos (RE) con Ontologías Forenses Digitales (SDFO) científicamente optimizadas. En el trabajo [64] trata acerca del correo electrónico como vehículo para amenazas tipo Phishing.

En la mayoría de estos trabajos no se indica la metodología utilizada para construir la ontología, salvo [23] y [24] que presentan una metodología propia; [74] y [89] que utilizan Methontology con algunas variantes; en el trabajo [95] se utiliza la metodología Sherwood Applied Business Security (SABSA) y en el estudio [75] se recurre a NeOn.

Muy pocos artículos ([23], [24], [75], [88], [77]) describen la utilización de herramientas y lenguajes utilizados en la propuesta. Entre los lenguajes y herramientas utilizados se encuentran: RDF, RDFS, SPARQL, OWL, OWL 2, SWLR y SPARQL Endpoint.

Solo describen gráficamente la taxonomía definida en [25], [63], [75], [88], [77], [89] y [72]; mientras que [23] y [25] incluyen casos de uso y en [74], [25], [52] se menciona la reutilización de otras ontologías.

Por otra parte, son varias las investigaciones abordadas desde *Data Mining.* Aplicando *aprendizaje automático* se encontraron varios trabajos: en [28] se propone un enfoque híbrido para detectar, filtrar y archivar pruebas provenientes de correos electrónicos; en tanto la herramienta DYNAMINER es introducida en [30] para la identificación de malware como un problema de aprendizaje basado en análisis de tráfico de red. También se encontraron trabajos relacionados a *redes neuronales* para la identificación de Malware en el estudio [37] . Las herramientas relacionadas a la *extracción de datos* son utilizadas en [66] para la definición de mejores atributos de una cámara de video, mediante la aplicación de técnicas de procesamiento de imágenes y extracción de datos y así aumentar la precisión de detección, la solidez y la eficiencia computacional del dispositivo de video.

El estudio detallado en [46] describe el manejo de datos privilegiados mediante técnicas de extracción de datos, basadas en un script dentro de la herramienta forense digital Nuix. Por su parte, en [65] se estudia la aplicación de técnicas de extracción de datos para la detección de fraudes en los estados financieros de auditoría y propone una taxonomía para apoyar y guiar la búsqueda de datos, considerando la inclusión en el análisis de correos electrónicos como evidencia digital, utilizando técnicas de minería de textos.

Se encontraron aplicaciones de la minería de datos tanto a la Forensia como a seguridad informática. [93] utiliza técnicas de *clustering* para el agrupamiento de correos electrónicos no deseados en función de diferentes atributos que permiten acortar los tiempos de búsqueda en el repositorio de SPAM. En tanto que en [40] se estudian la capacidad de propagación de gusanos que amenazan cada vez más a los hosts y servicios de Internet, en donde se recurre a la minería de datos para explotar vulnerabilidades desconocidas y además estudiar sus propias representaciones cambiantes. Finalmente, en [67] se identifican y analizan los vectores de ataque de exfiltración de datos (fuga de datos confidenciales o privados a una entidad no autorizada) y las contramedidas vigentes.

Por otra parte, se detectaron algunos trabajos que estudian analíticas sobre grandes volúmenes de datos o *Big Data*. En [87] los autores describen como realizar el análisis forense digital de un repositorio de big data con datos heterogéneos provenientes de diversas fuentes. En tanto, el estudio [58] muestra cómo aplicar Fuzzy Hashing en investigaciones forenses y así identificar datos complejos y no estructurados que tienen cierta similitud de nivel de bytes.

También se describen herramientas analíticas para big data. En [98] se presenta LIFTR que permite priorizar información selectiva recuperada de los teléfonos Android. El estudio [26] muestra PROFORMA, un sistema prototipo que evalúa continuamente la confiabilidad y el riesgo de las comunicaciones sociales, en el cual el usuario otorga permiso explícito para acceder a las redes sociales. Por su parte, [30] estudia los remanentes de datos de valor forense del servicio de almacenamiento en la nube privada de Syncany, un motor de almacenamiento popular para plataformas de big data, orientando el estudio a la reducción de tiempos de búsqueda de la evidencia. En [42] se propone una herramienta de big data para analizar el tráfico de Internet y extraer información de alto nivel, como enlaces visitados, credenciales de usuario y cookies de sesión de los protocolos de red utilizados.

En la revisión realizada se hallaron trabajos relacionados a la *seguridad informática*. Entre estos trabajos se destacan la investigación descripta en el trabajo [67] donde se identifican las lagunas legales y las tecnologías que facilitan la comisión de actos delictivos cibernéticos; así como la propuesta [54] que describe las técnicas de ingeniería inversa que pueden utilizarse para acceder a datos encriptados. Por su parte, en el estudio [78] se muestra una clasificación de la evidencia digital en función del momento de obtención de las pruebas (análisis en vivo o post mortem), análisis de dispositivos móviles y herramientas de análisis forense. Por otra parte, también se abordaron *técnicas de ataque y mitigación de amenazas.* En tal sentido el trabajo [73] presenta un análisis para detectar y mitigar los ataques o amenazas esteganográficas en teléfonos celulares mientras que el estudio señalado en [41] aborda técnicas de análisis correlacional para identificar familias de ransomware.

Entre *otras áreas temáticas* identificadas se puede mencionar estos trabajos: forensia de máquinas virtuales ([27] y [69]); análisis forense de archivos en memoria o discos ([79], [32], [33], [35], [55], [36]), y aplicación de técnicas de procesamiento de lenguaje natural. Relacionado con esta última temática, el artículo [96] propone una herramienta para la traducción automática de políticas de seguridad escritas en lenguaje natural a lenguaje de máquina, mediante el lenguaje ABAC (Attribute Based Access Control).

El análisis de la revisión según la clasificación por áreas temáticas se completa con aquellos trabajos referidos a la aplicación de dichas áreas en la forensia de correos electrónicos, los cuales se detallan en la sección C.

B. Análisis por Objeto de Estudio

Si se considera el elemento central del estudio, se pueden identificar trabajos que desarrollan *métodos o metodologías específicas* para determinadas actividades. Así, los estudios se pueden agrupar en métodos para: gestionar incidentes cibernéticos ([51], [91], [80], [50], [23], [63], [82], [40]), analizar el tráfico de red ([53], [31], [38], [83]), detectar datos falsificados ([29], [34], [70], [44]), identificar virus ([36], [37], [93]), así como identificar y procesar datos complejos y no estructurados ([58], [62], [24], [87], [78], [69], [49], [94]).

También se encontraron algunos métodos innovadores o de atención para el análisis forense, como por ejemplo: en [84] se detallan los procesos de investigación sobre fuentes de ataques cibernéticos y en [46] se proponen métodos para privilegiar datos forenses y minimizar la exposición de los contenidos al investigador forense.

Se analizaron cinco trabajos que describen *herramientas* orientadas al análisis de trafico de redes (CARONTE [97], NETFOX [81], TRIG [39], RDAP [59], RSLogix [60]) y una herramienta orientada al seguimiento de incidentes de seguridad (DYNAMINER [30]).

Asimismo, en esta revisión se identificaron un conjunto de herramientas basadas en ontologías: INFORENSIC Ver 2 [74], OBMO [25] y [75], Fi4SOA [95], CyBox [52], OTM [76], SC-Ont [77], PROFORMA [26], F-DOC [89] y DF [72]. También se encontraron herramientas basadas en procesamiento de lenguaje natural ([96]), en técnicas de minería de datos ([65], [66], [68]) y en técnicas de análisis de grandes volúmenes de datos ([67], [41], [42], [43]).

En cuanto a los trabajos que abordan el análisis forense de *artefactos*, se pueden citar los siguientes: estudios para análisis de dispositivos móviles ([48],[54],[98],[73]), para análisis de computadoras ([42],[76],[32],[35]) y para análisis de redes ([57],[38],[61],[92]) considerando diversos sistemas operativos y software de base; estudios sobre análisis de memorias ([79] y [32]) y máquinas virtuales ([27]); sobre análisis de imágenes ([92]) y mensajería instantánea en redes sociales ([83]).

En el caso particular de los estudios que tratan sobre métodos, herramientas y artefactos que involucran correos electrónicos, se detallan en la sección siguiente.

C. Análisis de Estudios Sobre Correos Electrónicos

En este tercer criterio de clasificación se detallan los trabajos que abordan el análisis de correos electrónicos, detallando particularmente como se encararon desde las áreas temáticas señaladas en la clasificación de la sección A, y como se identificaron según el objeto de estudio según la clasificación de la sección B.

Así, se pueden considerar los trabajos relacionados a *tráfico de redes*, en donde particularmente interesan aquellos estudios centrados en el *análisis de la dirección IP*. En [53] se presenta un algoritmo para reducir el volumen de direcciones IP a analizar considerando la identificación e interacción de los usuarios mediante el análisis de metadatos del tráfico de red. El trabajo [97] propone un método para recuperar la dirección IP oculta en servicios de TOR (The Onion Router). En tanto, el estudio [59] desarrollaron la herramienta RDAP que resuelve consultas sobre registro de direcciones IP, nombres de dominio, sistemas autónomos, características de seguridad e internacionalización para una dirección IP determinada.

Es de destacar que solo se encontró un trabajo que recurre a las *ontologías* aplicadas a correos electrónicos. El estudio [71] describe una ontología que representa el direccionamiento semántico del correo electrónico, que permite a los usuarios dirigir correos electrónicos a grupos especificados semánticamente, proporcionando autenticación segura a grupos de cuentas de correo.

Desde la *minería de datos* se realizaron varias investigaciones sobre correos electrónicos. En [93] se recurre a la agrupación de correos electrónicos no deseados en función de sus diferentes atributos para conformar clústeres que permiten acortar los tiempos de búsqueda en el repositorio de correos spams. El estudio [28] utiliza aprendizaje automático para detectar, filtrar y archivar evidencia proveniente de correos electrónicos, que permitan identificar a quienes actúan en el ciberespacio cometiendo delitos. Los autores del trabajo [65] proponen una aplicación de técnicas de *extracción de datos y minería de textos* para la detección de fraudes en los estados financieros de auditoría y propone una taxonomía para apoyar y guiar la búsqueda de datos, considerando la inclusión en el análisis de correos electrónicos como evidencia digital. Finalmente, [82] apela a la *minería de textos* para generar un marco de detección de acciones de ciberacoso en mensajes; servicio de mensajes cortos, servicio de mensajes multimedia, chat, mensajes de instancia y correos electrónicos.

En [56] se proponen técnicas analíticas de *Big Data* para la búsqueda de evidencia en grandes conjuntos de datos de correo electrónico.

Desde la *Seguridad Informática* se encontraron varios trabajos que se pueden identificar en base al tipo de *ataque cibernético* planteados desde correos electrónicos.[70] analiza cinco escenarios forenses en los que el investigador forense tiende a pasar por alto la información incriminatoria crucial que se ha disfrazado de spam. También referido a la seguridad informática, se analizó el trabajo [37] en el cual se realiza una encuesta a los usuarios finales y se concluye que es habitual la violación de las políticas de seguridad, particularmente con el uso indebido del correo electrónico. El estudio [34] aborda el Spearphishing (variación de Phishing que ataca a organizaciones especialmente seleccionadas) con un novedoso enfoque automatizado basado en modelos probabilísticos de metadatos de correo electrónico y características estilométricas del contenido de correo electrónico. Acerca de investigaciones sobre Phishing se encontró el estudio [64] que trata el correo electrónico como vehículo para este tipo de amenaza de seguridad, en dicho estudio se describe una encuesta que investiga los ataques de phishing y las técnicas antiphishing desarrolladas no solo en entornos tradicionales, como correos electrónicos y sitios web, sino también en entornos nuevos, como las redes sociales y teléfonos móviles.

Entre los trabajos que abordan la aplicación de *otras áreas de la informática* aplicada a la forensia de correos electrónicos, se encontró el estudio [68] que propone técnicas de procesamiento de lenguaje natural (clasificación asociativa personalizada) para abordar el problema de atribución de autoría de correos electrónicos a partir de las características que definen el estilo de escritura de una persona.

Considerando los *métodos forenses* es de interés el estudio [72] en el que los autores definieron un método forense basado en las directrices disponibles preparadas por el Instituto Nacional de Estándares y Tecnología (NIST) para forensia de correos electrónicos. Por su parte, en [49] se propone un algoritmo para encontrar vínculos y repeticiones de datos en un contexto de investigación forense, se evaluó la efectividad del algoritmo buscando similitud en cadenas de correos electrónicos y permitió segregar direcciones de correo similares de las no similares.

A continuación se detallan las *herramientas forenses* para el procesamiento de datos de correos electrónicos. Se identificaron las siguientes: Sistema de Visualización de correlaciones para Foxmail [33] que permite extraer la información del archivo de evidencia de correo mostrando gráficamente la asociación entre los contactos y permite la búsqueda en el cuerpo del correo así como el archivo adjunto mediante la recuperación de texto completo; EMAILFINDER [35] para acceder a información de correos electrónicos residente en la memoria de teléfonos móviles; INVEST [56] que posibilita la búsqueda de evidencia en grandes conjuntos de datos de correo electrónico; en el estudio [47] se presenta un script en SQLite Index Recovery que se probó con datos de la aplicación Apple Mail.

Considerando el correo electrónico como *artefacto forense*, la revisión permitió identificar investigaciones de interés. El estudio [31] refiere al comportamiento de las diferentes aplicaciones cliente de correo electrónico mientras recibe los correos electrónicos falsificados del remitente, el estudio plantea un algoritmo para la identificación de direcciones falsas mediante el análisis de los campos SPF, DKIM, DKIM-Signature y DMARC recibidos. Los autores de [61] investigan acerca de las ventajas y amenazas de la infraestructura de correo electrónico basada en la nube y analizan las amenazas

de ataque a los servidores de correo corporativos. Finalmente, el estudio [86] considera los correos electrónicos salientes y la fuga de información involuntaria considerando los metadatos que son una parte natural de los encabezados de los correos electrónicos, marcando un nivel notable de exposición de la información de identidad personal y organizativa que puede quedar a disposición de un atacante.

Particularmente, los trabajos [28],[30],[78],[84] y [90] abordan el análisis forense de correos electrónicos desde el encabezado del mismo. El último trabajo mencionado, compara las cinco herramientas forenses de código abierto más populares para forensia de correos electrónicos a partir del encabezado, señalando el detalle de funcionalidades que ofrecen como resultado (determinación de la IP, del ID del correo, identificación del emisor/receptor, fecha y hora, búsqueda y visualización de datos, grado de usabilidad, entre otros). El estudio [29] discute acerca de diferentes métodos para detectar la falsificación de correos electrónicos (spoofing) analizando el encabezado del correo. En [85] se define un método forense para teléfonos Android basado en las directrices preparadas por el Instituto Nacional de Estándares y Tecnología (NIST) para forensia de correos electrónicos a partir de los metadatos de la cabecera. En [79] se analizan varias herramientas forenses que se basan en los registros del encabezado de los correos electrónicos, con énfasis en la delincuencia en línea y las restricciones legales, examinando la amplitud de la información que se puede obtener con esas herramientas. Por su parte, en el trabajo [31] se propone un algoritmo de tres niveles para identificar correos maliciosos a partir de los registros de servidores y dispositivos locales que se encuentran en el encabezado de los correos.

Respecto del *proceso de transmisión* de correos electrónicos, los trabajos precitados en el párrafo anterior dan por sentado que al considerar las direcciones IP del encabezado se analiza el proceso de transmisión pero no lo enfocan de manera directa y concreta, salvo en el trabajo [31].

IV. Aportes De La Revisión Realizada

En esta sección se resumen los resultados de la revisión, desde su contribución a los objetivos planteados para la revisión bibliográfica.

Así, respecto del primer objetivo, referido a *Identificar y estudiar los aportes investigativos más actualizados sobre Ontologías y Forensia Digital,* se puede decir que el mismo se cumplió ya que se encontraron contribuciones destacadas para la formalización científica de la Forensia Digital, particularmente, aquellos vinculados a las ontologías, que en un total de 15 trabajos ([52],[72],[25],[75],[88],[77],[95],[23],[74],[89],[24],[63],[76],[64] y [71]) representan el 20% de las investigaciones consideradas, y están enfocados a cuatro temáticas particulares: direccionamiento de cuentas, desarrollo de herramientas forenses, estrategias para mitigación de ataques y representación semántica de artefactos forenses.

El segundo objetivo de esta revisión es la propuesta de *establecer las áreas de vacancia sobre la aplicación de las ontologías al análisis forense de correos electrónicos.* Al respecto, se observó que los estudios realizados se enfocaron en tres áreas temáticas: el estudio del *Tráfico de Redes*, encontrándose 18 trabajos, que principalmente abordaron a temáticas relacionadas a los *ataques cibernéticos* ([39],[51],[91],[84] y [38]), al *direccionamiento de IP* ([53],[97] y [59]) y a *Forensia de Redes de Datos* ([81],[60],[57],[83]); en la *Minería de Datos* se analizaron 11 estudios, de los cuales se destacan aquellos que trabajan la minería de textos y técnicas de extracción de datos ([66],[46],[65],[82],[40],[67]); y en la *Seguridad Informática* con ocho trabajos que se enfocan principalmente en la atención de ataques cibernéticos ([73],[41]). Mientras que las áreas con escasa aplicación en la forensia de correos electrónicos serían las siguientes: *Big Data* con cuatro ([87],[43],[56],[78]) de los ocho trabajos analizados*, Ontologías* con tres trabajos ([64],[71],[89]) de los 15 considerados *y Procesamiento de Lenguaje Natural* ([68]).

En particular, la conjunción entre ontologías y forensia digital aplicada a correos electrónicos solo se encontró en un trabajo de investigación: el estudio [71] en el que se propone el direccionamiento de correo electrónico semántico (SEA) para dirigir correos a grupos de usuarios especificados semánticamente; pero se debe destacar que este trabajo no aborda el desarrollo de una ontología que modele específicamente la cabecera del correo electrónico. Los otros dos estudios encontrados, [64] y [89] incluyen ontologías que representan al correo electrónico como un único concepto, sin desagregar los componentes que lo integran.

Asimismo se observa que si bien en varios trabajos [28],[30],[78],[84] y [90]) se aborda el sistema de transmisión del correo electrónico a partir de su encabezado, en ninguno se aplica el concepto de *trazabilidad* como elemento vinculante de los distintos equipos utilizados en la transmisión. Las guías procesales para periciar correos electrónicos ([99]) establecen que se debe acceder al correo *recibido* y verificar los equipos de emisión y recepción mediante la dirección IP que figura en la cabecera del mismo, realizando un recorrido *inverso* del proceso de transmisión para llegar desde el equipo receptor al equipo emisor. Este recorrido inverso puede sostenerse técnicamente si se aplica el concepto de *trazabilidad del proceso de transmisión,* y puede sostenerse científicamente si dicha trazabilidad se representa mediante una ontología.

El último objetivo de la revisión, referido a *relacionar estos trabajos desde atributos de cercanía (o distancia) con la aplicación de ontologías para el análisis forense de correos electrónicos,* puede decirse que se cumplió, ya que la revisión se desarrolló con el detalle suficiente como para encontrar dos resultados concretos: a) las ontologías se utilizan para el desarrollo de herramientas de forensia digital (se observó esto en 10 trabajos, que representan el 13% de las publicaciones revisadas); y b) solo un trabajo trata sobre la aplicación de ontologías a la forensia de correos electrónicos en particular con el grado de detalle que se busca ([71]).

Se prestó atención a las escasas investigaciones (solo cinco que representa el 7% de las publicaciones revisadas), dirigidas a estudiar el análisis forense de correos electrónicos basados en los metadatos de la cabecera, y de este grupo, solo en un trabajo se abordó detalladamente el proceso de transmisión.

V. Conclusiones

Los objetivos propuestos para la revisión se cumplieron, ya que se pudieron identificar investigaciones actualizadas sobre aplicación de ontologías a la Forensia Digital, se encontraron áreas de vacancia de interés para trabajar desde el espacio de la investigación académica, y se pudo relacionar las publicaciones revisadas en términos de cercanía o distancia en la aplicación ontologías.

Desde la Forensia Digital se debería considerar las fortalezas y ventajas de los métodos y herramientas propios de Big Data y Procesamiento de Lenguaje Natural. Y respecto de las ontologías, aunque fue considerada por un conjunto importante de investigadores sobre Forensia Digital, se abordó escasamente en el análisis pericial de correos electrónicos. De lo dicho, se puede concluir que son varios los ámbitos en los que sería importante generar estudios sobre Forensia Digital.

Por una parte, las técnicas y herramientas analíticas de Big Data son adecuadas para procesar grandes volúmenes de datos no estructurados, como los contenidos en las cuentas de correo electrónico, para realizar estudios sobre el volumen, variedad y valor de esos datos, considerando además las velocidades de procesamiento que permiten estas herramientas analíticas.

El correo electrónico cuenta con un componente textual de mucho interés para la Forensia Digital, particularmente porque la documentación impresa fue reemplazada poco a poco por los mensajes de correo electrónico, en los que también se observan características sobre el estilo de escritura, sintaxis y semántica de las palabras. Desde este enfoque, los métodos y herramientas propias del Procesamiento de Lenguaje Natural pueden aportar un marco científico adecuado.

En la búsqueda bibliográfica realizada se encontraron trabajos asociados a una o más de estas dos temáticas: *ontologías y forensia digital*, en los que se observan modelos, criterios o componentes que son útiles para la formulación de herramientas y marcos ontológicos para resolver problemáticas referidas a la forensia de correos electrónicos.

Pero también se debe destacar que no se encontraron trabajos sobre el caso particular de aplicación de la trazabilidad para validar correos electrónicos, basado en una representación ontológica, lo cual evidencia la necesidad del desarrollo de criterios científicos que certifiquen y formalicen las actividades de la forensia digital. La ausencia de este tipo de trabajos abre posibilidades para la investigación en esta línea. En última instancia, esta característica –la comprobación de la existencia del correo electrónico- es la que permite sostener la condición de no repudio de esta evidencia digital.

El tema no se agota aquí. Considerando que la tecnología avanza con velocidad en el desarrollo de nuevas áreas (Internet de las Cosas por ejemplo) y en la profundización de áreas existentes (Inteligencia Artificial por citar alguna), seguramente surgirán nuevas investigaciones basadas en la web semántica y que traten la forensia digital en esas áreas.

Referencias

[1] D. Ce, D. E. L. Parlamento, and E. Y. Del, "Directiva 2002/58/Ce Del Parlamento Europeo y del Consejo Sobre Protección De Datos," vol. 39, pp. 1–22, 2008.

[2] S. L. Garfinkel, "Digital forensics research: The next 10 years," *Digit. Investig.*, vol. 7, no. SUPPL., pp. S64–S73, 2010.

[3] G. Guizzardi, "Theoretical foundations and engineering tools for building ontologies as reference conceptual models," *Semant. Web*, vol. 1, no. 1–2, pp. 3–10, 2010.

[4] M. de Reuver and T. Haaker, "Designing viable business models for context-aware mobile services," *Telemat. Informatics*, vol. 26, no. 3, pp. 240–248, 2009.

[5] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, 2007.

[6] J. Biolchini, P. G. Mian, A. C. C. Natali, and G. H. Travassos, "Systematic Review in Software Engineering," 2005.

[7] M. J. Grant and A. Booth, "A typology of reviews: an analysis of 14 review types and associated methodologies.," *Health Info. Libr. J.*, vol. 26, no. 2, pp. 91–108, 2009.

[8] J. D. Velásquez, "Una Guía Corta para Escribir Revisiones Sistemáticas de Literatura Parte 3," *Dyna*, vol. 82, no. 189, pp. 9–12, 2015.

[9] C. Medina Lopez, J. A. Marin Garcia, and R. Alfalla Luque, "Una propuesta metodológica para la realización de búsquedas sistemáticas de bibliografía," *WPOM-Working Pap. Oper. Manag.*, vol. 1, no. 2, pp. 13–30, 2010.

[10] I. Portugal, P. Alencar, and D. Cowan, "The use of machine learning algorithms in recommender systems: A systematic review," *Expert Syst. Appl.*, vol. 97, pp. 205–227, 2018.

[11] M. Edwards, A. Rashid, and P. Rayson, "A Systematic Survey of Online Data Mining Technology Intended for Law Enforcement," *ACM Comput. Surv.*, vol. 48, no. 1, pp. 1–54, 2015.

[12] J. Slay and F. Schulz, "Development of an Ontology Based Forensic Search Mechanism: Proof of Concept," *J. Digit. Forensics, Secur. Law*, vol. 1, no. 1, pp. 25–44, 2006.

[13] F. Amato, L. Barolli, G. C. B, A. Mazzeo, and F. Moscato, *ECT: A Novel Architecture for Evidence Collection in Forensic Investigation*, vol. 13. 2018.

[14] M. K. Pandya, S. Homayoun, and A. Dehghantanha, "Forensics investigation of openflow-based SDN platforms," in *Advances in Information Security*, vol. 70, 2018, pp. 281–296.

[15] J. M. Spring, T. Moore, and D. Pym, "Practicing a Science of Security," *Proc. 2017 New Secur. Paradig. Work. ZZZ - NSPW 2017*, pp. 1–18, 2017.

[16] K. F. Hong, C. C. Chen, Y. T. Chiu, and K. Sen Chou, "Ctracer: Uncover C&C in Advanced Persistent Threats Based on Scalable Framework for Enterprise Log Data," *Proc. - 2015 IEEE Int. Congr. Big Data, BigData Congr. 2015*, pp. 551–558, 2015.

[17] N. Hoque, M. H. Bhuyan, R. C. Baishya, D. K. Bhattacharyya, and J. K. Kalita, "Network attacks: Taxonomy, tools and systems," *J. Netw. Comput. Appl.*, vol. 40, no. 1, pp. 307–324, 2014.

[18] S. Michael and T. J. Gollins, "Our Digital Legacy: an Archival Perspective," *J. Contemp. Arch. Stud.*, vol. 4, p. 3, 2017.

[19] X. Liu, Q. Liu, T. Peng, and J. Wu, "Dynamic access policy in cloud-based personal health record (PHR) systems," *Inf. Sci. (Ny).*, vol. 379, pp. 62–81, 2017.

[20] P. Ohm, "Good Enough Privacy," *U. Chi. Leg. F.*, vol. 1, no. 1, p. 1, 2008.

[21] J. Kim, A. Park, and S. Lee, "Recovery method of deleted records and tables from ESE database," *Digit. Investig.*, vol. 18, pp. S118–S124, 2016.

[22] G. J. Williams and A. Winner, "We ' ve Got Mail : Email Preservation at a Small , Private University by Nahali (Holly) Croft," vol. 13, pp. 65–90, 2016.

[23] V. Mavroeidis, "A Framework for Data-Driven Physical Security and Insider Threat Detection," *2018 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min.*, pp. 1108–1115, 2018.

[24] F. Amato, G. Cozzolino, A. Mazzeo, and F. Moscato, "An application of semantic techniques for forensic analysis," *2018 32nd Int. Conf. Adv. Inf. Netw. Appl. Work.*, pp. 380–385, 2018.

[25] R. Carvalho, M. Goldsmith, and S. Creese, "Applying Semantic Technologies to Fight Online Banking Fraud," *Proc. - 2015 Eur. Intell. Secur. Informatics Conf. EISIC 2015*, pp. 61–68, 2016.

[26] A. Gupta, S. Dasgupta, and A. Bagchi, "PROFORMA: Proactive Forensics with Message Analytics," *IEEE Secur. Priv.*, vol. 15, no. 6, pp. 33–41, 2017.

[27] H. Riaz and M. A. Tahir, "Analysis of VMware virtual machine in forensics and anti-forensics paradigm," *6th Int. Symp. Digit. Forensic*

Secur. ISDFS 2018 - Proceeding, vol. 2018-Janua, no. Isdfs, pp. 1–6, 2018.

[28] Z. Ghasem, I. Frommholz, and C. Maple, "A hybrid approach to combat email-based cyberstalking," *2015 4th Int. Conf. Futur. Gener. Commun. Technol. FGCT 2015*, no. Fgct, 2015.

[29] A. Jayan and S. Dija, "Detection of spoofed mails," in *2015 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2015*, 2015, pp. 1–4.

[30] B. Eshete and V. N. Venkatakrishnan, "DynaMiner: Leveraging Offline Infection Analytics for On-the-Wire Malware Detection," *Proc. - 47th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Networks, DSN 2017*, pp. 463–474, 2017.

[31] D. L. Msongaleli, "Electronic Mail Forensic Algorithm for Crime Investigation and Dispute Settlement," *Digit. Forensic Secur. (ISDFS), 2018 6th Int. Symp.*, pp. 1–5, 2018.

[32] R. P. Iyer, P. K. Atrey, G. Varshney, and M. Misra, "Email spoofing detection using volatile memory forensics," *2017 IEEE Conf. Commun. Netw. Secur. CNS 2017*, vol. 2017-Janua, pp. 619–625, 2017.

[33] Z. Chen *et al.*, "Email Visualization Correlation Analysis Forensics Research," *2017 IEEE 4th Int. Conf. Cyber Secur. Cloud Comput.*, pp. 339–343, 2017.

[34] S. Duman, K. Kalkan-Cakmakci, M. Egele, W. Robertson, and E. Kirda, "EmailProfiler: Spearphishing Filtering with Header and Stylometric Features of Emails," *Proc. - Int. Comput. Softw. Appl. Conf.*, vol. 1, pp. 408–416, 2016.

[35] L. Chen and Y. Mao, "Forensic analysis of email on android volatile memory," *Proc. - 15th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 10th IEEE Int. Conf. Big Data Sci. Eng. 14th IEEE Int. Symp. Parallel Distrib. Proce*, pp. 945–951, 2017.

[36] S. Samuel, J. Graham, and C. Hinds, "Hunting Malware: An Example Using Gh0st," *Proc. - 2017 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2017*, pp. 97–102, 2018.

[37] E. M. Rudd, R. Harang, and J. Saxe, "MEADE: Towards a Malicious Email Attachment Detection Engine," *2018 IEEE Int. Symp. Technol. Homel. Secur. HST 2018*, pp. 1–7, 2018.

[38] D. Patil and B. Meshram, "Network Packet Analysis for Detecting Malicious Insider," *2018 3rd Int. Conf. Converg. Technol. I2CT 2018*, pp. 1–8, 2018.

[39] Y. Choi, J. Lee, S. Choi, J. Kim, and I. Kim, "Traffic Storing and Related Information Generation System for Cyber Attack Analysis," pp. 1052–1057, 2016.

[40] R. Kaur and M. Singh, "A survey on zero-day polymorphic worm detection techniques," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 3, pp. 1520–1549, 2014.

[41] K. P. Subedi, D. R. Budhathoki, and D. Dasgupta, "Forensic analysis of ransomware families using static and dynamic analysis," *Proc. - 2018 IEEE Symp. Secur. Priv. Work. SPW 2018*, no. June, pp. 180–185, 2018.

[42] A. Amro, S. Almuhammadi, and S. Zhioua, "NetInfoMiner: High-level information extraction from network traffic," *2017 IEEE Int. Conf. Big Data Smart Comput. BigComp 2017*, pp. 143–150, 2017.

[43] Y.-Y. Teing, D. Ali, K. Choo, M. T. Abdullah, and Z. Muda, "Greening Cloud-Enabled Big Data Storage Forensics: Syncany as a Case Study," *IEEE Trans. Sustain. Comput.*, pp. 1–1, 2017.

[44] S. Gupta, E. S. Pilli, P. Mishra, S. Pundir, and R. C. Joshi, "Forensic analysis of E-mail address spoofing," *Proc. 5th Int. Conf. Conflu. 2014 Next Gener. Inf. Technol. Summit*, pp. 898–904, 2014.

[45] H. Pieterse, M. S. Olivier, and R. P. Van Heerden, "Playing hide-and-seek: Detecting the manipulation of Android Timestamps," *2015 Inf. Secur. South Africa - Proc. ISSA 2015 Conf.*, 2015.

[46] D. Fleurbaaij, M. Scanlon, and N. A. Le-Khac, "Privileged data within digital evidence," *Proc. - 16th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 11th IEEE Int. Conf. Big Data Sci. Eng. 14th IEEE Int. Conf. Embed. Softw. Syst.*, pp. 737–744, 2017.

[47] F. Ramisch and M. Rieger, "Recovery of SQLite Data Using Expired Indexes," *Proc. - 9th Int. Conf. IT Secur. Incid. Manag. IT Forensics, IMF 2015*, pp. 19–25, 2015.

[48] H. Pieterse, M. S. Olivier, and R. P. Van Heerden, "Reference architecture for android applications to support the detection of manipulated evidence," *SAIEE Africa Res. J.*, vol. 107, no. 2, pp. 92–103, 2016.

[49] S. Jo, J. Kim, and D. Choi, "The study of document filter for smart device," *17th Asia-Pacific Netw. Oper. Manag. Symp. Manag. a Very Connect. World, APNOMS 2015*, pp. 515–518, 2015.

[50] G. S. Tanwar and A. S. Poonia, "Live forensics analysis: Violations of business security policy," *Proc. 2014 Int. Conf. Contemp. Comput. Informatics, IC3I 2014*, pp. 971–976, 2014.

[51] G. Settanni *et al.*, "A collaborative cyber incident management system for European interconnected critical infrastructures," *J. Inf. Secur. Appl.*, vol. 34, pp. 166–182, 2017.

[52] E. Casey, G. Back, and S. Barnum, "Leveraging CybOXtm to standardize representation and exchange of digital forensic information," *Digit. Investig.*, vol. 12, no. S1, pp. S102–S110, 2015.

[53] N. Clarke, F. Li, and S. Furnell, "A novel privacy preserving user identification approach for network traffic," *Comput. Secur.*, vol. 70, pp. 335–350, 2017.

[54] X. Zhang, I. Baggili, and F. Breitinger, "Breaking into the vault: Privacy, security and forensic analysis of Android vault applications," *Comput. Secur.*, vol. 70, pp. 516–531, 2017.

[55] Y. Y. Teing, A. Dehghantanha, K. K. R. Choo, and L. T. Yang, "Forensic investigation of P2P cloud storage services and backbone for IoT networks: BitTorrent Sync as a case study," *Comput. Electr. Eng.*, vol. 58, no. 2016, pp. 350–363, 2017.

[56] J. Koven, E. Bertini, L. Dubois, and N. Memon, "InVEST: Intelligent visual email search and triage," *Digit. Investig.*, vol. 18, pp. S138–S148, 2016.

[57] M. Scanlon, J. Farina, and M. T. Kechadi, "Network investigation methodology for BitTorrent Sync: A Peer-to-Peer based file synchronisation service," *Comput. Secur.*, vol. 54, pp. 27–43, 2015.

[58] P. C. Bjelland, K. Franke, and A. Årnes, "Practical use of Approximate Hash Based Matching in digital investigations," *Digit. Investig.*, vol. 11, no. SUPPL. 1, pp. S18–S26, 2014.

[59] B. Nikkel, "Registration Data Access Protocol (RDAP) for digital forensic investigators," *Digit. Investig.*, vol. 22, pp. 133–141, 2017.

[60] S. Senthivel, I. Ahmed, and V. Roussev, "SCADA network forensics of the PCCC protocol," *Digit. Investig.*, vol. 22, pp. S57–S65, 2017.

[61] A. Bhardwaj and S. Goundar, "Security challenges for cloud-based email infrastructure," *Netw. Secur.*, vol. 2017, no. 11, pp. 8–15, 2017.

[62] E. Casey, "Using computed similarity of distinctive digital traces to evaluate non-obvious links and repetitions in cyber-investigations," vol. 24, pp. 2–9, 2018.

[63] S. Iqbal *et al.*, "On cloud security attacks: A taxonomy and intrusion detection and prevention as a service," *J. Netw. Comput. Appl.*, vol. 74, pp. 98–120, 2016.

[64] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey," *Comput. Secur.*, vol. 68, pp. 160–196, 2017.

[65] G. L. Gray and R. S. Debreceny, "A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits," *Int. J. Account. Inf. Syst.*, vol. 15, no. 4, pp. 357–380, 2014.

[66] M. Jahanirad, A. W. A. Wahab, and N. B. Anuar, "An evolution of image source camera attribution approaches," *Forensic Sci. Int.*, vol. 262, pp. 242–275, 2016.

[67] F. Ullah, M. Edwards, R. Ramdhany, R. Chitchyan, M. A. Babar, and A. Rashid, "Data exfiltration: A review of external attack vectors and countermeasures," *J. Netw. Comput. Appl.*, vol. 101, pp. 18–54, 2018.

[68] M. R. Schmid, F. Iqbal, and B. C. M. Fung, "E-mail authorship attribution using customized associative classification," *Digit. Investig.*, vol. 14, no. S1, pp. S116–S126, 2015.

[69] J. Park, "TREDE and VMPOP: Cultivating multi-purpose datasets for digital forensics – A Windows registry corpus as an example," *Digit. Investig.*, vol. 26, pp. 3–18, 2018.

[70] S. Yu, "Covert communication by means of email spam: A challenge for digital investigation," *Digit. Investig.*, vol. 13, pp. 72–79, 2015.

[71] R. Mehta, "SEMANTIC E-MAIL ADDRESSING USING," 2017.

[72] N. M. Karie and H. S. Venter, "Toward a general ontology for digital forensic disciplines," *J. Forensic Sci.*, vol. 59, no. 5, pp. 1231–1241, 2014.

[73] W. Mazurczyk and L. Caviglione, "Steganography in Modern Smartphones and Mitigation Techniques," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 1. pp. 334–357, 2015.

[74] D. Ellison, H. Venter, and ikuesan R. Adeyemi, "An Improved Ontology for Knowledge Management in Security and Digital Forensic," 2017.

[75] R. Carvalho and R. Carvalho, "CDT Technical Paper 06/14 Online Banking Malware Ontology Rodrigo Carvalho."

[76] L. M. Maake, V. R. Kebande, and N. M. Karie, "Onto-Engineering : A Conceptual framework for Integrating Requirement Engineering Process with scientifically tuned Digital Forensics Ontologies," no. June, 2017.

[77] E. Kalemi and S. Yildirim-Yayilgan, "Ontologies for Social Media Digital Evidence," *Int. J. Comput. Electr. Autom. Control Inf. Eng.*, vol. 10, no. 2, pp. 335–340, 2016.
[78] V. Jusas, D. Birvinskas, and E. Gahramanov, "Methods and tools of digital triage in forensic context: Survey and future directions," *Symmetry (Basel).*, vol. 9, no. 4, 2017.
[79] B. Romaios, K. Nikolaos, K. George, and A. Andreas, "Email forensic tools: A roadmap to email header analysis through a cybercrime use case," *J. Polish Saf. Reliab. Assoc. Summer Saf. Reliab. Semin.*, vol. 7, no. 1, pp. 21–28, 2016.
[80] C. S. D. Brown, "Investigating and prosecuting cyber crime: Forensic dependencies and barriers to justice," *Int. J. Cyber Criminol.*, vol. 9, no. 1, pp. 55–119, 2015.
[81] J. Pluskal, P. Matoušek, O. Ryšavý, M. Kmeť, V. Veselý, and F. Karpíšek, "Netfox Detective : A Tool for Advanced Network Forensics Analysis," no. i.
[82] Z. Ghasem, I. Frommholz, and C. Maple, "A machine learning framework to detect and document text-based cyberstalking," in *CEUR Workshop Proceedings*, 2015, vol. 1458, no. October, pp. 348–355.
[83] T. Y. Yang, A. Dehghantanha, K. K. R. Choo, and Z. Muda, "Windows Instant Messaging App Forensics: Facebook and Skype as Case Studies," *PLoS One*, vol. 11, no. 3, p. e0150300, 2016.
[84] K. K. Shashidhar and D. H. Manjaiah, "Forensic Investigation Processes for Cyber Crime and Cyber Space," *Adv. Intell. Syst. Comput.*, vol. 216, pp. 169–178, 2014.
[85] R. Umar, I. Riadi, and B. F. Muthohirin, "Acquisition of Email Service Based Android Using NIST," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 3, no. 3, p. 263, 2018.
[86] J. Nurse, A. Erola, M. Goldsmith, and S. Creese, "Investigating the leakage of sensitive personal and organisational information in email headers," *J. Internet Serv. Inf. Secur.*, vol. 5, no. 1, pp. 70–84, 2015.
[87] H. Mohammed, N. Clarke, and F. Li, "An Automated Approach for Digital Forensic Analysis of Heterogeneous Big Data," *J. Digit. Forensics, Secur. Law*, vol. 11, no. 2, 2017.
[88] H. Wimmer, L. Chen, T. Narock, and L. Chen, "Ontologies and the Semantic Web for Digital Investigation Tool Selection," *J. Digit. Forensics, Secur. Law*, vol. 13, no. 3, pp. 21–46, 2018.
[89] M. Alzaabi, T. A. Martin, K. Taha, and A. Jones, "The use of ontologies in forencis analysis of smartphone content," *J. Digit. Forensics, Secur. Law*, vol. 10, no. 4, pp. 105–114, 2015.
[90] V. K. Devendran, H. Shahriar, and V. Clincy, "A Comparative Study of Email Forensic Tools," *J. Inf. Secur.*, vol. 06, no. 02, pp. 111–117, 2015.
[91] F. Breitinger and I. Baggili, "File Detection on Network Traffic Using Approximate Matching," *J. Digit. Forensics, Secur. Law*, vol. 9, no. 2, 2017.
[92] Y. Wu, D. Ye, Z. Wei, Q. Wang, W. Tan, and R. H. Deng, "Situation-aware Authenticated Video Broadcasting over Train-trackside WiFi Networks," *IEEE Internet Things J.*, 2018.
[93] R. Khan, M. Mizan, R. Hasan, and A. Sprague, "Hot Zone Identification: Analyzing Effects of Data Sampling On Spam Clustering," *J. Digit. Forensics, Secur. Law*, no. c, 2017.
[94] D. Jeong, H. Kang, and S. Lee, "Towards Syntactic Approximate Matching - A Pre-Processing Experiment," *J. Digit. Forensics, Secur. Law*, vol. 11, no. 2, 2017.
[95] A. Akremi, H. Sallay, M. Rouached, R. Bouaziz, and M. Abid, "Forensics-aware web services composition and ranking," *Proc. 17th Int. Conf. Inf. Integr. Web-based Appl. &Services - iiWAS '15*, pp. 1–10, 2015.
[96] R. C. Turner, "Proposed Model for Natural Language ABAC Authoring," pp. 61–72, 2017.
[97] S. Matic, P. Kotzias, and J. Caballero, "CARONTE: Detecting Location Leaks for Deanonymizing Tor Hidden Services," *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur. - CCS '15*, pp. 1455–1466, 2015.
[98] S. Varma, R. J. Walls, B. Lynn, and B. N. Levine, "Efficient Smart Phone Forensics Based on Relevance Feedback," pp. 81–91, 2014.
[99] M. E. Darahuge and L. E. Arellano González, *Manual de Informática Forense III*. 2016.

Beatriz Parra, is a Computer Engineer (UCSE) and obtained a Masters in Business Administration (UCASAL). Professor and researcher at the Catholic University of Salta (UCASAL, Argentina). She worked as an Official Computer Expert of the Court of Justice of Salta (Argentina) during 2000-2011 and is currently working as an Expert of Party. He is currently developing his doctoral thesis on 'An ontology of e-mail and its traceability as a support for Digital Forensic' to obtain the title of Doctor in Information Systems Engineering from the Santa Fe Regional School, National Technological University. (Argentina).

Marcela Vegetti is a professor at the Department of Information Systems Engineering of the Facultad Regional Santa Fe, Universidad Tecnológica Nacional (Santa Fe, Argentina). She also holds a position as assistant researcher at the National Council of Scientific and Technical Research of Argentina (CONICET), working at the "Instituto de Desarrollo y Diseño". She obtained her PhD. Degree in Engineering in Information Systems from Universidad Tecnológica Nacional in 2007. Her current research activities focus on the application of ontologies and semantic technologies for conceptual modeling and supporting interoperability in different domains.

Horacio Leone is a full Professor at the Department of Information Systems Engineering of the Facultad Regional Santa Fe, Universidad Tecnológica Nacional (Santa Fe, Argentina). He also holds a Researcher position at the National Council for Scientific and Technical Research of Argentina (CONICET), working at Instituto de Desarrollo y Diseño. He obtained his PhD degree in Chemical Engineering from Universidad Nacional del Litoral (Santa Fe, Argentina) in 1986 and was a Postdoctoral Fellow at the Massachusetts Institute of Technology (1986–1989). He current research activities focus on software architectures, models for supporting the design process, semantic web applications to supply chain information systems, and enterprise modelling. He has supervised several PhD students.