

Appendix A. Systematic Literature Review - Protocol

All scientific production with investigation must be based on evidence [33], in particular, Systematic Literature Reviews (SLR) [34], which provides a critical synthesis on the state of the art of primary studies for a given topic. Such synthesis usually summarizes results, identifies methodologies, compares solutions, and finds out the contributions of such studies and potential research gaps.

For this SLR, we followed the well-established guidelines by Kitchenham [6]. We also adopted the StArt tool⁴ to support systematizing our SLR.

Figure A.2 depicts the SLR process we adopted. We start by **defining our research questions** (Section Appendix A.1), then performing the primary studies **search process** (Section Appendix A.2). With the initial set of studies, we perform the **selection process** (Section Appendix A.3) supported by the **extraction process** (Section Appendix A.4). With the final set of studies, we perform the **analysis process** by presenting (Section 3) and discussing (Section 4) the obtained results.



Figure A.2: SLR process

Appendix A.1. Aim and Research Questions

This SLR is **aimed at characterizing** existing ontologies, **with the purpose** identifying and classifying such ontologies for providing a panorama and recommendations for further investigation, **from the point of view of** DF researchers, **in the context of** primary studies presenting ontologies to support the DF area.

Therefore, we defined one main research question (RQ) with a set of secondary research questions (SRQ), as follows:

- ***RQ: How have ontologies supported the Digital Forensics area?***
 - ***SRQ1: What methodologies have been used to build ontologies for Digital Forensics?***
 - ***SRQ2: What are the types of ontologies that have contributed to the work of researchers and practitioners in Digital Forensics?***
 - ***SRQ3: What evaluation methods have been used for ontologies of Digital Forensics?***
 - ***SRQ4: What phases of Digital Forensics have ontologies supported?***
 - ***SRQ5: What subareas of Digital Forensics have ontologies supported?***

⁴http://lapes.dc.ufscar.br/tools/start_tool

In **SRQ1** we analyze the methodologies/methods used to build ontologies for DF, including tools and information sources. **SRQ2** seek to identify ontological structures, models, and domains. In addition, we want to analyze their scope and provide a classification based on relevant properties to DF. In **SRQ3** we investigate what techniques have been used to empirically evaluate the ontologies, such as, scenarios, case studies, and experiments. For **SRQ4** we analyze what DF phases the ontologies have been built for. Therefore, a mapping on the ontologies versus DF phases might be seen as a potential roadmap. Finally, in **SRQ5**, we analyze the DF subareas each ontology supports.

Appendix A.2. Search Process

Our first step in the search process was defining key terms and their synonyms to compose a search string. Therefore, we list the following:

- **ontology/ontologies**; and
- **digital forensics** - *“digital forensic”, “computer forensic”, “computer forensics”, “digital evidence”, and “forensic investigation”*.

In the next step we composed our search string to be applied in the digital database sources by using the **“and”** and **“or”** logical operators in the defined terms and synonyms. Therefore, Table A.6 presents our defined search string.

Table A.6: Search string

(ontology OR ontologies)
AND
(“digital forensic” OR “digital forensics” OR “computer forensic” OR “computer forensics” OR “digital evidence”)

Once we defined our search string, we selected the following sources (search engines and databases) to search for relevant primary studies: ACM, IEEE, Scopus, Springer, and Science Direct. We list the following criteria to select these sources:

- worldwide known Computer Science sources;
- source provides an advanced search mechanism able to handle advanced queries;
- internationally indexed source;
- source indexes most relevant conferences and journals in Digital Forensics;
- source indexes papers with high impact factor (JCR and/or H-index);
- source has not ceased publication.

These criteria seek to exclude databases that are inactive or not related to the field of computing, such that selecting articles that present theoretical foundations related to the object of study and essential to identify the techniques that produce theoretical knowledge for the researcher [6].

Table A.7 summarizes the search results per source. Thus, we came up with an initial set of 523 studies for the selection process.

Table A.7: Summary of the Search Process

Sources	Returned Studies
ACM (Digital Library)	54
IEEE	75
Scopus	61
Springer	164
Science Direct	169
Total	523

Appendix A.3. Selection Process

We started the selection process by defining inclusion (I) and exclusion (E) criteria for the primary studies as follows:

- (I.1) studies dealing with ontologies for Digital Forensics;
- (E.1) studies that do not address ontology for Digital Forensics;
- (E.2) studies that are not in the field of Computer Science;
- (E.3) studies with four pages or less;
- (E.4) studies with the StArt tool quality score⁵ = 0;
- (E.5) study not written in English to allow worldwide dissemination and reproduction;
- (E.6) study text is not completely available;
- (E.7) workshop, conference, and symposium calls;
- (E.8) duplicated study;

⁵The quality score of a study we adopted is looking for the terms defined in Section Appendix A.2 in the study title (five points each), abstract (three points each) and keywords (two points each). This terms verification is automated in StArt

- (E.9) study unavailable even contacting author(s).

To perform the selection process, we screened all 523 primary studies by applying the inclusion and exclusion criteria. A summary of papers resulting from this process is then presented in Table A.8, from which 70 studies (column “Selected”) are ready for the extraction process.

Table A.8: Summary of the Selection Process.

Sources	Returned*	Selected	Duplicated	Rejected
ACM	54	9	3	42
IEEE	75	21	5	49
Scopus	61	22	8	31
Science Direct	169	18	19	132
Springer	164	0	0	164
Total	523	70	35	418

*from the search process (Section Appendix A.2).

Appendix A.4. Extraction Process

In this SLR stage, we fully and carefully read all the 70 studies resulted from the selection process. In parallel of reading and extracting the studies data, we also double checked them according to the inclusion and exclusion criteria defined in the selection process. Therefore, we came up with 41 studies excluded based on different criteria as shown in Table A.9.

Table A.9: Rejected Primary Studies.

Ref.	Title	Year	Pub. Type	Publication Venue	Excl. Criterion
[35]	Knowledge discovery and experience modeling in computer forensics media analysis	2004	Conference	Proceedings of the 2004 international symposium on information and communication technologies	E.1
[36]	The computing ontology: application in education	2007	Journal	Association for Computing Machinery - Special Interest Group on Computer Science Education Bulletin	E.1
[37]	DFS3: automated distributed file system storage state reconstruction	2020	Conference	Proceedings of the 15th International Conference on Availability, Reliability and Security	E.1
[38]	Computer crime investigation by means of fuzzy semantic maps	2009	Conference	IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology	E.3
[39]	Research on Ontology-Based Cluster Analysis in Computer Forensics	2014	Conference	Seventh International Symposium on Computational Intelligence and Design	E.1
[40]	Forensic future of social media analysis using web ontology	2017	Conference	4th International Conference on Advanced Computing and Communication Systems (ICACCS)	E.1
[41]	Knowledge sharing and reuse in digital forensics	2010	Conference	Fifth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering	E.1
[42]	Ontology-based forensic event detection using inference rules	2017	Conference	International Conference on Engineering, Technology and Innovation (ICE/ITMC)	E.1
[43]	Cyber identity: Salient trait ontology and computational framework to aid in solving cybercrime	2018	Conference	17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)	E.1
[44]	An ontology framework for recommendation about a crime scene investigation	2014	Conference	14th International Symposium on Communications and Information Technologies (ISCIT)	E.1
[45]	Evidence/discovery-based evolving ontology (EDBEO)	2010	Conference	IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS)	E.1

[46]	Situation-based ontologies for a computational framework for identity focusing on crime scenes	2017	Conference	IEEE conference on cognitive and computational aspects of situation management (CogSIMA)	E.2
[47]	A latent semantic indexing and WordNet based information retrieval model for digital forensics	2018	Conference	IEEE International Conference on Intelligence and Security Informatics	E.1
[48]	OntoEDS: protecting energy delivery systems by collaboratively analyzing security requirements	2017	Conference	IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)	E.2
[49]	A framework for identity: Dempster-Shafer theory the flow and combination of evidence	2018	Conference	IEEE Symposium Series on Computational Intelligence (SSCI)	E2
[50]	Semantic interpretation of structured log files	2016	Journal	IEEE 17th International Conference on Information Reuse and Integration (IRI)	E.1
[51]	An ontology-based transformation model for the digital forensics domain	2017	Journal	Elektronika ir elektrotechnika	E.1
[52]	Towards a built-in digital forensics-aware framework for web services	2015	Conference	11th International Conference on Computational Intelligence and Security (CIS)	E.1 and E.3
[53]	Ontology applied in the Computer Forensics process	2016	Journal	CEUR Workshop Proceedings	E.5
[54]	Semantic modelling of digital forensic evidence	2010	Conference	International Conference on Digital Forensics and Cyber Crime	E.1
[55]	Multimedia forensics is not computer forensics	2019	Conference	International Workshop on Computational Forensics	E.1
[56]	A survey of information security incident handling in the cloud	2015	Journal	computers & security	E.1
[57]	Analyse digital forensic evidences through a semantic-based methodology and NLP techniques	2019	Journal	Future Generation Computer Systems	E.1
[58]	Study of identifying and managing the potential evidence for effective Android forensics	2020	Journal	Forensic Science International: Digital Investigation	E.1
[59]	Experience constructing the artifact genome project (agp): managing the domain's knowledge one artifact at a time	2018	Journal	Digital Investigation	E.2
[60]	Subject-based semantic document clustering for digital forensic investigations	2013	Journal	Data & Knowledge Engineering	E.1
[61]	Digital forensics as a service: Stepping up the game	2020	Journal	Forensic Science International: Digital Investigation	E.1
[62]	Knowledge representation model for crime analysis	2017	Journal	Procedia computer science	E.1
[63]	Validation and verification of computer forensic software tools—Searching Function	2009	Journal	Digital Investigation	E.1
[64]	STITCHER: correlating digital forensic evidence on internet-of-things devices	2020	Journal	Forensic Science International: Digital Investigation	E.1
[65]	Leveraging ontologies and machine-learning techniques for malware analysis into Android permissions ecosystems	2018	Journal	Computers & Security	E.1
[66]	An enhanced multiclass support vector machine model and its application to classifying file systems affected by a digital crime	2019	Journal	Journal of King Saud University-Computer and Information Sciences	E.1
[3]	Towards a conceptual model for promoting digital forensics experiments	2020	Journal	Forensic Science International: Digital Investigation	E.1
[67]	ACrime Intelligence from Social Media Using CISMO	2020	Conference	International Congress on Information and Communication Technology	E.6
[68]	Towards designing a knowledge graph-based framework for investigating and preventing crime on online social networks	2019	Conference	International Conference on e-Democracy	E.6
[69]	ECT: A Novel Architecture for Evidence Collection in Forensic Investigation	2017	Conference	International Conference on P2P, Parallel, Grid, Cloud and Internet Computing	E.6
[70]	A Knowledge Model of Digital Evidence Review Elements Based on Ontology	2020	Journal	Digital Forensics and Forensic Investigations: Breakthroughs in Research and Practice	E.6
[71]	Semantic Integration and Correlation of Digital Evidences in Forensic Investigations	2016	Conference	International Conference on P2P, Parallel, Grid, Cloud and Internet Computing	E.6
[72]	SMONT: an ontology for crime solving through social media	2017	Journal	Int. J. Metadata Semant. Ontologies	E.6
[73]	Reconstruction et analyse smantique de chronologies cybercriminelles	2014	Conference	14èmes Journées Francophones Extraction et Gestion des Connaissances	E.6
[74]	An automatically filtering blacklist model of social network based on semantic web	2014	Journal	Journal of Harbin Institute of Technology (New Series)	E.6

According to Table A.9, 27 (65.8%) studies were excluded based on **E.1 criterion**. This is because they present no ontologies in their paper sections, except for the abstract, which mentions them. Bogen and Dampier [35], which sought to model experiences during the analysis of digital evidence, present no ontology for its context. Another example for this criterion is the work of Cassel et al. [36], which only reported

the history of the research group, but no interesting approach to the DF context. The study of Chi [39], despite based on an ontology, did not present the ontological model, but a clustering algorithm. In general, all works excluded by the E.1 criterion do not present subsidies that characterize it as coherent for final acceptance and data extraction in the research questions. We excluded four (9.75%) studies based on **E.2 criterion** as they deal with crimes committed through physical actions, thus not in the digital context as, for instance, in the works of McDaniel et al. [46] and Mason et al. [49]. Two (4.87%) works were excluded regarding **E.3 criterion** as they have four pages or less, thus no representation of an ontology could be properly represented in such works. Only one (2.44%) study was excluded based on **E.5 criterion**, as its abstract was in English, but its content was in Portuguese [53]. Eight (19.5%) works were excluded as they were not available (**E.6 criterion**). No works matched **E.7, E.8, and E.9 criteria**.

We extracted the following fields from each study: metadata (title, authors, year of publication, publication venue, and publication type); category; ontological contributions; methodologies used; phase with critical points; classification of ontologies and points that support them.

Table A.10 presents the list of final primary studies, after the complete reading and data extraction phase. Therefore, data analysis of this SLR was based on the 29 final works. Each work is uniquely identified with an ID starting with the letter “S” as for “Study” and a sequence natural number, from S1 through S29 (column “ID & Ref.”). Note we also added the respective reference (“Ref.”) of each study.

Table A.10: SLR Final Set of Studies

ID & Ref.	Title	Year	Pub. Type	Publication Venue
S1 [11]	A framework for semantic annotation of digital evidence	2013	Conference	Annual ACM Symposium on Applied Computing
S2 [22]	Towards a comprehensive ontology based-investigation for digital forensics cybercrime	2015	Journal	International Journal on Communications Antenna and Propagation
S3 [12]	Ontology for reactive techniques in digital forensics	2019	Conference	Application, Information and Network Security
S4 [23]	An ontological approach to study and manage digital chain of custody of digital evidence	2011	Journal	Journal of Information and Organizational Sciences
S5 [75]	A framework for data-driven physical security and insider threat detection	2018	Conference	International Conference on Advances in Social Networks Analysis and Mining
S6 [13]	System of systems cyber effects simulation ontology	2015	Conference	Winter Simulation Conference
S7 [14]	A semantic-based methodology for digital forensics analysis	2020	Journal	Journal of Parallel and Distributed Computing
S8 [76]	Addressing the increasing volume and variety of digital evidence using an ontology	2014	Conference	Joint Intelligence and Security Informatics Conference
S9 [77]	The Phases Based Approach for Regeneration of Timeline in Digital Forensics	2020	Conference	International Conference on INnovations in Intelligent SysTems and Applications
S10 [15]	A case-based reasoning method for locating evidence during digital forensic device triage	2014	Journal	Decision Support Systems
S11 [16]	Cyber forensics ontology for cyber criminal investigation	2009	Conference	International Conference on Forensics in Telecommunications, Information, and Multimedia
S12 [78]	D4I-Digital forensics framework for reviewing and investigating cyber-attacks	2020	Journal	Array
S13 [79]	A semantic-web-technology-based framework for supporting knowledge-driven digital forensics	2016	Conference	International Conference on Management of Digital EcoSystems
S14 [10]	A complete formalized knowledge representation model for advanced digital forensics timeline analysis	2014	Journal	Digital Investigation
S15 [17]	An ontology framework for automated visual surveillance system	2015	Conference	International Workshop on Content-based Multimedia Indexing
S16 [80]	DESO: Addressing volume and variety in large-scale criminal cases	2015	Journal	Digital Investigation
S17 [81]	A comprehensive literature review of file carving	2013	Conference	International Conference on Availability, Reliability and Security
S18 [82]	DIALOG: A framework for modeling, analysis and reuse of digital forensic knowledge	2009	Journal	Digital Investigation

S19 [18]	Network security assessment using a semantic reasoning and graph based approach	2017	Journal	Computers and Electrical Engineering
S20 [19]	Forensics-aware web services composition and ranking	2015	Conference	International Conference on Information Integration and Web-based Applications and Services
S21 [20]	Toward a cybercrime classification ontology: A knowledge-based approach	2019	Journal	Computers in Human Behavior
S22 [69]	Correlation of digital evidences in forensic investigation through semantic technologies	2017	Workshop	International Conference on Advanced Information Networking and Applications Workshops
S23 [21]	The Information Retrieval Model for Computer Crime Evidence Based on Domain Ontology	2019	Conference	International Symposium on Computational Intelligence and Design
S24 [83]	An ontology-based approach for the reconstruction and analysis of digital incidents timelines	2015	Journal	Digital Investigation
S25 [24]	Semantic representation and integration of digital evidence	2013	Journal	Procedia Computer Science
S26 [10]	Automatic timeline construction and analysis for computer forensics purposes	2014	Conference	Joint Intelligence and Security Informatics Conference
S27 [84]	A multilayered semantic framework for integrated forensic acquisition on social media	2019	Journal	Digital Investigation
S28 [85]	Method ontology for intelligent network forensics analysis	2010	Conference	International Conference on Privacy, Security and Trust
S29 [86]	Leveraging CyBOX to standardize representation and exchange of digital forensic information	2015	Journal	Digital Investigation

We evaluated in a qualitative way all 29 studies of the final set. To do so, we defined two binary questions (Yes or No) to allow us to gather up evidence of their minimum contribution. Such questions are:

- Q1 - Does the study have a clear objective?
- Q2 - Has the study proposal been evaluated?

Table A.11 summarizes such evaluation. As we can see, only two studies (S15 and S28) have no clear objective stated. In addition, 17 (58,6%) studies evaluated their ontologies.

Table A.11: Quality assessment of studies.

Study ID	Q1	Q2	Study ID	Q1	Q2	Study ID	Q1	Q2
S1	✓	✓	S11	✓		S21	✓	
S2	✓		S12	✓	✓	S22	✓	
S3	✓		S13	✓	✓	S23	✓	
S4	✓	✓	S14	✓	✓	S24	✓	✓
S5	✓		S15		✓	S25	✓	✓
S6	✓	✓	S16	✓	✓	S26	✓	
S7	✓	✓	S17	✓		S27	✓	✓
S8	✓	✓	S18	✓	✓	S28		✓
S9	✓	✓	S19	✓		S29	✓	
S10	✓	✓	S20	✓				
Count	10	7	Count	9	6	Count	8	4

✓ means “Yes” for a given question.

We then performed the Analysis Process (Figure A.2). We present and discuss its results in Sections 3 and 4, respectively.

Appendix A.5. Study Data Dissemination

All data from this study is available in Zenodo at <https://doi.org/10.5281/zenodo.7248899>, a public, permanent, and citable repository. We followed Open Science practices [87, 88] to make this research available for every citizen.