Cybersecurity Ontologies: A Systematic Literature Review

Revisión sistemática de la literatura sobre ontologías en ciberseguridad

William Fernando Borja Rivadeneira²
william.f.borja.r@pucesa.edu.ec

Omar S. Gómez¹,²
ogomez@espoch.edu.ec

¹GrIISoft Research Group, Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador.
²Pontificia Universidad Católica del Ecuador Sede Ambato, Ambato, Ecuador
Abstract: Cybersecurity is a young discipline that has gained relevance in our modern society. This research reports the findings of a systematic review of the literature on ontologies in the field of cybersecurity. From an initial set of 214 papers on the subject, 50 relevant papers were selected for this SLR. With these documents we answered research questions related to the domains in which ontologies are reported, the methodologies, tools and languages used, and the verification and validation mechanisms reported. As results, we observed that the largest number of ontologies are classified in the domains of infrastructure and networking, software and human factor. Regarding the papers that report the use of a methodology for developing the ontologies (12%), Methontology is the commonly used one. Protégé, in conjunction with the OWL language, are the preferred tools for ontology development. Regarding verification and validation (V&V) mechanisms, we observe that more than half (62%) report the application of V&V mechanisms to their ontologies.

Keywords: Systematic Literature Review, Ontology, Cyber-security, Cybersecurity Ontologies, ICT.

Resumen: La ciberseguridad es una disciplina joven que ha tomado relevancia en nuestra sociedad moderna. En la presente investigación se reportan los hallazgos de una revisión sistemática de la literatura sobre ontologías en el ámbito de la ciberseguridad. De un conjunto inicial de 214 documentos sobre el tema, para la presente SLR se utilizaron 50 documentos relevantes. Con estos documentos se respondieron preguntas de investigación relacionadas con los dominios en los que se reportan las ontologías, las metodologías, herramientas y lenguajes usados, así como los mecanismos de verificación y validación reportados. Con respecto a los resultados observamos que el mayor número de ontologías se clasifican en los dominios de infraestructura y networking, software y factor humano, dentro del porcentaje de documentos en los que se reporta el uso de alguna metodología para el desarrollo de ontologías (12%), methontology es la metodología comúnmente usada. Protege en conjunto con el lenguaje OWL es la herramienta de preferencia para el desarrollo de ontologías. En cuanto a los mecanismos de verificación y validación, observamos que poco más de la mitad (62%) reporta algún mecanismo de V&V.

Palabras Claves: Revisión Sistemática de la Literatura, Ontología, Ciber-seguridad, Ontologías de Ciberseguridad, ICT.
INTRODUCTION

There is no doubt that Information and Communications Technology (ICT) has significantly revolutionized the knowledge society in which we are immersed. Advances in software, hardware and telecommunications have managed to converge in different sectors of our society, offering modern solutions. However, due to the massification of TICs, issues related to vulnerabilities and threats in software, hardware and telecommunications have arisen.

Cybersecurity is a young discipline aimed at protecting vulnerabilities or minimizing threats to technological infrastructure such as software, hardware and telecommunications (Thakur and Pathan, 2014). Although knowledge about cybersecurity issues is mostly held by people involved in the ICT arena, due to massive use of ICT, knowledge about cybersecurity should be extended to the general public (Singer and Friedman, 2014).

Knowledge about cybersecurity has been gradually built up thanks to the diversity of contributions made by different experts in this field. A portion of these contributions have focused on creating ontologies that help to define, represent and organize a vocabulary of concepts (Neches et al., 1991) related to this discipline. These ontologies provide a shared knowledge on the different aspects of cybersecurity. In order to have a better understanding of ontologies reported in the field of cybersecurity, this paper presents the results of a systematic review of the literature on the various ontologies that have been reported in the context of cybersecurity.

The rest of the document is organized as follows: Section II presents some generalities about ontologies; Section III describes the method used in which the research questions to be answered by this systematic review are framed. Section IV presents the results with respect to the specified research questions. Finally, section V discusses the results and presents the conclusions.

ONTLOGIES

In the origins of Western thought, ontology was considered a discipline related to philosophy. It was oriented to the study of the existing (entities) and their relationships. In a general way, an ontology can be defined as a vocabulary, in which entities, classes, properties, predicates, functions and the relations between these elements are stated. An ontology is important because it enables sharing knowledge regarding a particular domain.

In the literature we can find different approaches or methods to create ontologies (Lenat and Guha, 1990; Uschold and King, 1995; Grüninger and Fox, 1995; Bernaras et al., 1996; Fernandez et al., 1997; Swartout et al., 1997; Staab et al., 2001; KBSI, 1994). For example, in (Lenat and Guha 1990) authors propose an approach for the creation of ontologies that support intelligent systems based on knowledge. Similarly, in (Uschold and King, 1995) authors propose an ontology construction method aimed at capturing knowledge, coding it and integrating it with existing ontologies. In the case of (Grüninger and Fox, 1995), authors propose a methodology that allows the development of systems based on first-order knowledge; in addition, taking advantage of the robustness of classical logic as a guide to transform informal systems into computational ones.
This methodology focuses on identifying the main scenario for the construction of ontologies. On the other hand, in (Bernaras et al., 1996), the Kactus methodology is presented. It focuses on the construction of ontologies considering a knowledge base that uses a process of abstraction, where the context of the entities is specified. One of the widely known methodology is Methontology, proposed by Juristo (Fernandez et al., 1997). It helps to create a new ontology as well as reusing existing ones. This methodology fits into a development process based on the creation of prototypes.

The SENSUS method (Swartout et al., 1997) is another approach, it uses existing ontologies creating an ontology skeleton, the resulting prototype eliminates terms irrelevant to the domain knowledge. Another methodology is Onto-Knowledge (Staab et al., 2001), a project that supports the development of ontologies for knowledge management. Finally, in (KBSI, 1994), authors mention the KBSI IDEF5 method, which is a method that allows and helps in the creation, modification and maintenance of ontologies.

Tools are also relevant in developing ontologies. Examples of these tools are: Ontolingua Server (Farquhar et al., 1997), Ontosaurus (Swartout et al., 1997), Protégé (Noy et al., 2000), WebODE (Arpírez et al., 2003), OntoEdit (Sure et al., 2002), among others. One tool that is commonly used in the creation of ontologies is Protégé (Noy et al., 2000), which is an independent open-source tool that has an extensible architecture. Its main core is the ontology editor, that its functionality can be extended through the use of plug-ins.

Another tool that supports the creation of ontologies is Ontosaurus (Swartout et al., 1997). It consists of two modules: an ontology server that uses a knowledge representation system and a web browser that allows editing and exploring ontologies using HTML. It is worth to note that several of these tools have their own ontology development language. Examples of these languages are: XOL (XML-Based Ontology Exchange Language) (Karp et al., 1999); SHOE (Simple HTML Ontology Extensions) (Luke and Heflin, 2000); which is an extension of HTML, DAML+OIL (Horrocks and van Harmelen, 2001) and OWL (Web Ontology Language) (Dean and Schreiber, 2003).

These languages vary according to their use; besides considering that these languages can be integrated through extensions or APIs established by the provider of these languages. OWL (Dean and Schreiber, 2003) is one of the languages commonly used by ontology developers. OWL is aimed at publishing and sharing ontologies developed on the Web. OWL is a derivation of DAML+OIL (Horrocks and van Harmelen, 2001) which shares some of its functionalities.

When an ontology is developed, an important aspect to take into account is the one related to their verification and validation (V&V) (Raad and Cruz, 2015). It can be approached from two perspectives: with regards to its quality and with regards to its correctness (Raad and Cruz, 2015). In the literature we can find some approaches that address the verification and validation of an ontology, such as: the gold standard approach (Ulanov et al., 2010); corpus-based (Brewster et al., 2004); task-based (Welty et al., 2003); and criteria-based (Fernandez et al., 2009). As for the gold standard approach (Ulanov et al., 2010), it focuses on comparing the developed ontology with a reference ontology created with certain criteria.
On the other hand, the corpus-based approach (Brewster et al., 2004), consists in evaluating the coverage of an ontology with one or several ontologies through a corpus in which the determined domain is significantly covered. In the case of the task-based approach (Welty et al., 2003), the evaluation of an ontology is aimed at a specific task, based on the results obtained to improve the knowledge of this task. Another approach consists in validating the ontology according to a desirable criterion (Raad and Cruz, 2015), such as its structure (Fernandez et al., 2009), where for example the number of nodes that an ontology has is used, or more complex criteria may be used.

Another approach is based on experts (Alani and Brewster, 2006), for example, evaluations are based on the coincidences of classes, density and intermediation that are detailed in the ontology. Some support tools for V&V tasks are OntoMetric (Lozano et al., 2004), natural language application metrics (Hartmann et al., 2004), OntoClean (Gangemi et al., 2002), EvaLexon (Spyns et al., 2004) and OOPS! (Poveda et al., 2015). For example, OntoClean (Gangemi et al., 2002), allows the evaluation of an ontology based on its taxonomic structure. Another case is OOPS! (Poveda et al., 2015), a tool that scans an ontology by using a URL to find possible inconsistencies that may affect the modeling of it.

METHOD

The guideline described in (Kitchenham, 2004) was followed in order to conduct the Systematic Literature Review (SLR) here reported. According to (Kitchenham, 2004), the realization of a SLR is divided into three phases: planning, execution and reporting, following we describe each of these phases in our context.

As part of the planning phase, the protocol for this SLR was developed. It sets out the research questions, as well as the objectives of this research. In this phase, the source for searching, the search string, and the inclusion and exclusion criteria are also defined. In the second phase (execution) the protocol is run, in this phase we proceed with the searches of documents with respect to the search string defined in the protocol, we also carry out the discrimination of the documents according to the inclusion and exclusion criteria defined in the protocol, we also carry out the analysis and synthesis of the relevant documents (selected papers). Finally, the third phase corresponds to the presentation of the results of this SLR.

The main objective of this SLR is to gain a better understanding of reported ontologies related to cybersecurity domain. With respect to this objective, the following research questions have been posed.

- RQ1. What areas of cybersecurity are the ontologies reported in?
- RQ2. What methods or approaches have been used for the development of the selected ontologies?
- RQ3. What tools have been used for the construction of the reported ontologies?
- RQ4. Have the reported ontologies been validated?
Identification and selection of the searching source

For the present work, the Scopus database has been chosen. It contains the largest number of abstracts and citations in the scientific literature as well as offers a search engine with advanced searching options.

Definition of the search string

The search string has been defined with different terms based on the subject matter of this SLR, these terms have been combined with logical operators, resulting in the following search string:

TITLE-KEY-ABS(ontology AND (cyber-security OR cybersecurity OR “cyber security”))

Inclusion Criteria (CI) and Exclusion Criteria (CE)

We filtered the selection of primary studies (documents on the subject) in terms of the following inclusion and exclusion criteria. The primary studies were selected based on the title, abstract and keywords in order to determine whether they are identified as relevant ones, documents were selected taking into account compliance with the following inclusion criteria: papers reporting cybersecurity ontologies written in English language. These papers should be published in prestigious indexed venues such as journals, proceedings and book chapters subject to a peer review process. In the same way, those papers that met some of the following exclusion criteria were not considered for this SLR: duplicated papers, papers whose main contribution is not related with cybersecurity ontologies, poster papers, non-English written papers, and short communications such as letters to the editor.

Execution

Once the protocol was defined, we proceeded with the execution phase. The previously defined search string was executed in September 2020, using the Scopus database search engine. It is worth to note that Scopus is one of the biggest abstract and citation database in the arena of scientific literature. Initially 214 document results were obtained after running the search string. After applying the inclusion and exclusion criteria on titles, abstracts and keywords, 72 documents were selected (first filter).

From this selection, the complete content of 69 documents was accessed. After analyzing the content of these documents, 19 were discarded, so finally 50 documents were considered for this review (relevant papers). According to the previously described, we used the Scopus database to run the search string, and the contents of the relevant documents was accessed through their respective publishers such as IEEE, Springer, ACM, Science Direct, among others.

As shown in Figure 1, the first ontology related to cybersecurity was reported in 2004 (Simmonds et al., 2004), from that year onwards, other works sporadically arise, it is from 2014 that the number of ontologies related to cybersecurity increases, for example in 2019 eleven ontologies were reported (Burita, 2019; Vega Barbas et al., 2019; Gasmi et al., 2019; Doynikova et al., 2019; Scarpato et al., 2019; Niyazova et al., 2019; Islam et al., 2019; Baesso Moreira et al., 2019; Katsantonis et al., 2019; Shaaban et al., 2019; Zamfira et al., 2019).
It is worth to note that the relevant papers used in this SLR were subjected to a quality assessment. Nine evaluation criteria adapted from (Kitchenham, 2004) were considered, these criteria are related to the year of publication, the type of publication (journal or conference) as well as criteria related to the structure of the selected documents. A Likert scale was used for this purpose, where the maximum evaluation score was set 50 points, Table 1 shows the nine criteria used for the assessment.

**TABLE 1. Quality assessment criteria used for this SLR.**

The authors of this paper carried out the evaluation of the relevant documents and the results of the evaluation were averaged. As can be seen in Table 2, none of the papers was rejected based on the evaluation criteria used.
<table>
<thead>
<tr>
<th>Quality percentage score by category</th>
<th>Number of papers</th>
<th>Percentage of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor (&lt;26%)</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Regular (26%-45%)</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Good (46%-65%)</td>
<td>19</td>
<td>38%</td>
</tr>
<tr>
<td>Very Good (66%-85%)</td>
<td>26</td>
<td>52%</td>
</tr>
<tr>
<td>Excellent (&gt;86%)</td>
<td>5</td>
<td>10%</td>
</tr>
</tbody>
</table>

**TABLE 2. Quality assessment outcome.**

The total score from each paper was computed using a percentage scale. We observe that all of the relevant papers yielded a quality score that ranges from good to excellent. The average score of this evaluation was 76%, which was considered a good enough quality indicator for this SLR.

**RESULTS**

This section presents the results of the information synthesis process. These results are structured with respect to the research questions stated for this SLR.

**RQ1. What areas of cybersecurity are the ontologies reported in?**

Regarding this research question, we have identified four categories in which the ontologies are grouped: General, Networking, Software and Human Factor. The general category refers to those ontologies that involve a mix of concepts related to the other categories, such as networking, software, or the human factor. In the category of networking the ontologies address concepts related to equipment, protocols and network modeling. In the software category, the ontologies are mainly focused on describing cybersecurity concepts from a software development perspective. In the category of human factor are those ontologies that describe concepts related to the personnel involved in aspects of cybersecurity in ICTs. Table 3 presents the ontologies grouped by the categories previously described. As shown in Table 3, the largest number of ontologies analyzed are grouped in the general and networking categories.
<table>
<thead>
<tr>
<th>Scope / Domain</th>
<th>Number of papers</th>
<th>Percentage of papers</th>
<th>References</th>
</tr>
</thead>
</table>

**TABLE 3. Classification of Ontologies according to their scope.**

**RQ2. What methods or approaches have been used for the development of the selected ontologies?**

Regarding this research question, we observe that most of the authors of the reported ontologies (88%, 44 papers) do not mention to use existing methods for the development of their ontologies. In other words, the authors describe their own approach that they followed for the development of their ontologies. To a lesser extent we observe that only in six relevant papers (12%), the authors mention following some existing methodology for the development of their ontologies. For example, four authors mention the use of the Methontology methodology (Fernandez et al., 1997), while two authors mention the use of the methodology Ontology Development 101 (Noy and McGuinness, 2001).
Table 4 shows the resulting classification with regards this research question.

<table>
<thead>
<tr>
<th>Existing Method</th>
<th>Number of papers</th>
<th>Percentage of papers</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Following Existing Methods</td>
<td>6</td>
<td>12%</td>
<td>Zamfira et al. (2019), Maathuis et al. (2018), Obrst et al. (2014), Razzaq et al. (2014), Katsantonis et al. (2019), Van Vuuren et al. (2014)</td>
</tr>
</tbody>
</table>

TABLE 4. Classification of Ontologies according to their methodology.

As shown in Table 4, only six papers report the use of a methodology for developing their ontologies. Methontology approach was reported in four papers (Zamfira et al., 2019; Maathuis et al., 2018; Obrst et al., 2014; Razzaq et al., 2014), whereas the ontology development 101 method was reported in two works (Katsantonis et al., 2019; Van Vuuren et al., 2014).

**RQ3. What tools have been used for the construction of the reported ontologies?**

In the case of the tools that support the construction of ontologies, 40% of the documents (20 papers) authors mention the use of some tool to support the development of their ontologies. We observe that Protégé is the most used tool for ontology development; its use is mentioned in 16 out of 20 relevant documents that mention the use of a tool. To a lesser extent, the use of other tools is also reported, such as Atom-Tool; Cyber Security Ontology Expert Tool; CYBEX; and IntelMQ. Table 5 shows the ontologies grouped by tools.
The tools used for the ontology development usually incorporate languages that help in the definition of ontologies. We observe that 24% (12 papers) of the relevant documents (Chukkapalli et al., 2020; Vega Barbas et al., 2019; Doynikova et al., 2019; Zheng et al., 2018; Petrenko and Makoveichuk, 2017; Elnagdy et al., 2016; Falk, 2016; Syed et al., 2016; Iannacone et al., 2015; Oltramari et al., 2015; Salem and Wacek, 2015; Laskey et al., 2015) report the use of the OWL language (Dean and Schreiber, 2003).

To a lesser extent, the use of languages such as SPARQL, SWRL, XML and OWL 2 is observed (Baesso Moreira et al., 2019; Onwubiko et al., 2018; Albalushi et al., 2018; Tseng et al., 2017; Bergner and Lechner, 2017; Fontenele and Sun, 2016; Maines et al., 2015; Geller et al., 2014).

**RQ4. Have the reported ontologies been validated?**

Regarding the evaluation and validation of the ontologies reported in this SLR, we observe that 62% of them (31 primary studies) mention the use of some verification or validation mechanisms. For example, in 18 relevant papers, authors mention the use of information extraction rules similar to those proposed in (Boley et al., 2001). In the case of the ontologies here reported, some authors perform the verification and validation of their ontologies based on the comparison with existing ontologies (3 relevant papers). Another type of validation observed is the validation by experts, approach mentioned in other three relevant papers. The use of tools for assessing ontologies is also mentioned, tools like OntoClean (Gangemi et al., 2002), the OQure metrics tool (Duque and Fernandez, 2011), and a Protégé extension called HermiT Reasoner (Data and Knowledge Group) are mentioned. We also observe hybrid approaches in which validation is conducted through the use of criteria (Fernandez et al., 2009) and tasks (Welty et al., 2003). Finally, we also observe the use of a metrics-based validation approach. Table 6 shows the resulting classification regarding this research question.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Number of papers</th>
<th>Percentage of papers</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOM-TOOL</td>
<td>1</td>
<td>2%</td>
<td></td>
</tr>
<tr>
<td>Cyber Security Ontology Expert Tool</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYBEX</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IntelM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 5. Ontology classification by used tools**
<table>
<thead>
<tr>
<th>Evaluation / Validation</th>
<th>Number of papers</th>
<th>Percentage of papers</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison with Existing Ontologies</td>
<td>3</td>
<td>6%</td>
<td>Zheng et al. (2018), Syed et al. (2016), Iannacone et al. (2015)</td>
</tr>
<tr>
<td>Experts' validation</td>
<td>3</td>
<td>6%</td>
<td>Fontenele and Sun (2016), Chun and Geller (2015), Wali et al. (2013)</td>
</tr>
<tr>
<td>OntoClean</td>
<td>3</td>
<td>6%</td>
<td>Zamfira et al. (2019), Zamfira et al. (2018), Razzaq et al. (2014)</td>
</tr>
<tr>
<td>OQuare metrics</td>
<td>1</td>
<td>2%</td>
<td>Mozzaquatro et al. (2018)</td>
</tr>
<tr>
<td>HermiT Reasoner</td>
<td>1</td>
<td>2%</td>
<td>Maathuis et al. (2018)</td>
</tr>
<tr>
<td>Hybrid Approach</td>
<td>1</td>
<td>2%</td>
<td>Syed (2020)</td>
</tr>
<tr>
<td>Metrics-based Validation</td>
<td>1</td>
<td>2%</td>
<td>Gasmi et al. (2019)</td>
</tr>
</tbody>
</table>

**TABLE 6.** Evaluation and validation approaches reported.
DISCUSSION AND CONCLUSIONS

The highest percentage of ontologies reported in the field of cybersecurity are in the general and networking category. The general category addresses a mix of concepts belonging to the networking, software and human factor categories. These findings suggest that work is being done on the definition of ontologies in specific cybersecurity domains. However, we also observe work on the development of ontologies addressing a more general domain of the cybersecurity.

Regarding the methods or approaches used for the development of ontologies, we observe that in most of the relevant documents analyzed, authors do not mention following existing methodologies for developing their ontologies. Only in 12% (six papers) authors mention using some methodology as a reference. From this percentage, the most used methodology is Methontology (Fernandez et al., 1997). These findings seem to suggest a lack of motivation in the use of existing ontology development methodologies. We observe that the most widely used tool for the construction of ontologies is Protégé (Noy et al., 2000).

From the ontologies that report the use of some tool, 32% (16 papers) report the use of Protégé. OWL (Dean and Schreiber, 2003) is the most used language among the ontologies that report the use of a language.

Finally, regarding the use of verification and validation (V&V) mechanisms, we observe that 62% of the primary studies report the use of some verification or validation mechanism, among which the following stand out: the use of information extraction rules, making of comparisons with respect to existing ontologies, validation by experts, validation through the use of metrics, as well as the use of tools for this purpose. We observe that gradually there is interest in applying V&V mechanisms that help to correct deficiencies in the development of ontologies in the field of cybersecurity.

Study Limitations

It is worth to note that secondary studies as the one here reported are subject to interpretation in its different phases, thus implying the presence of bias. In order to minimize a possible bias, we followed the main phases with its activities of the SLR methodology. Although the risk of missing relevant papers was present, we consider that the selected documents for this review (relevant papers) represent a good enough sample of reported cybersecurity ontologies.

In this SLR, we did not consider gray literature, so we assume that good quality grey literature of this subject will be reported in journals or conferences, because of this, possible publication bias may arise due to negative findings are not usually published. We did not consider documents published in a non-English language, although this is not a limitation in our regional context, it can be a reflection of the limitations imposed on us by the available research in this area (updated and peer-reviewed literature is commonly published in English).

Cybersecurity is a young discipline in which disciplines such as telecommunications, electronics and computing have converged. The arrival of different ontologies in the field of cybersecurity has fostered to have more knowledge of the concepts related to this discipline. In this research we have reported the results of a systematic literature review on cybersecurity ontologies. The main contribution of this secondary study is the synthesis of the findings from the different ontologies reported with respect to areas or domains, methodologies, tools and languages used, as well as the V&V mechanisms reported. The results of our work can serve as a reference for future research on this topic.
REFERENCES


