Terminology Management in Cybersecurity through Knowledge Organization Systems: an Italian Use Case

Claudia Lanza*, Elena Cardillo†, Maria Taverniti†, Roberto Guarasci*

*University of Calabria, Rende, Italy
Email: c.lanza@dimes.unical.it; roberto.guarasci@unical.it
†Institute of Informatics and Telematics, National Research Council, Rende, Italy
Email: {elena.cardillo;maria.taverniti}@iit.cnr.it

***Abstract*—Specialized terminology is usually managed by Knowledge Organization Systems (KOSs), which manipulate and organize concepts and terms through standardized structured techniques. In this paper, an approach to organize, manage, and subsequently update specialized terminologies, specifically related to the domain of Cybersecurity, is proposed. A preliminary analysis and comparison between KOSs showing a higher level of semantic representation, i.e., thesauri and ontologies, is presented in the first section with the objective of clarifying the conceptual framework of these resources. A concrete use case in the domain of Cybersecurity is then described to show the context of application of these two semantic resources, i.e., a project funded by the Institute of Informatics and Telematics of the National Research Council aimed at providing terminology management and representation in the frame of the Italian Cybersecurity Observatory. A transaction between the thesaural and ontological representation of the domain knowledge represents the core of the approach showing the higher qualitative value that ontologies are able to provide to represent the domain of interest, due to the more precise formalization of semantic relationships existing among concepts.**

Keywords- Cybersecurity; KOS; Thesauri; Ontologies; Semantic relations.

I. INTRODUCTION

Managing technical terms proper to specialized languages represents one of the main tasks of Knowledge Organization Systems (KOSs). In the context of KOSs, semantic resources, as, for example, thesauri and ontologies, are useful tools to organize domain specific knowledge and to support processes like document indexing, information searching and retrieval and, in some cases, automatic reasoning (e.g., for decision making), above all in those specialized domains where semantic ambiguity between terms represents a step to be avoided. During the last few years some effort has been spent, as shown in Section III, on the definition of ontological models, used in the domain of Cybersecurity, aimed at supporting systems to better identify vulnerabilities and, thus, supporting decision making. Nevertheless, the specificity of the domain and the constant updates of the related information and data, the need for more appropriate semantic resources, based on standards, and highly structured to better represent the domain knowledge, is still evident. This is even more true in the Italian context, where there is a lack of highly semantically structured ways to manage the terminology of this field of study. Taking inspiration by this scenario, the present paper, which is an invited extension of [1], is focused on presenting a preliminary analysis of the main differences existing in the way of organizing and representing the information related to highly specialized domains, targeting the analysis on Cybersecurity. Amongst the KOSs [2] the comparison will focus on two means of semantic knowledge configuration: thesauri and ontologies. The reason why these two types of resources have been selected among others mainly relies on one of the main objectives of the Italian *OCS Project* coordinated by the Cyber Security Observatory of the Institute of Informatics and Telematics, National Research Council (IIT-CNR) [3], presented in detail in Section IV, which provides the understanding of the technical domain of Cybersecurity for a community of users demanding a guided orientation in this field of knowledge. The second purpose of the present work is twofold: (i) to show the results of the above mentioned project, whose main objectives are the development of an Italian and standardized controlled vocabulary, in other words a thesaurus [2] for the Cybersecurity domain, which can be considered a reliable knowledge organization system that structures the information related to specialized domains; (ii) to enhance of its semantic relationships and representation by exploiting a more formal language, i.e., the Web Ontology Language (OWL) [4], the recommended Semantic Web language for authoring ontologies.

The utility of this resource provided in the Italian scenario (and for this reason in Italian language), is specifically addressed to Italian medium-sized companies, citizens, stakeholders and scholars at different levels who need a key access point to better understand and reduce ambiguity dealing with Cybersecurity terminology. The vagueness of certain terms is due to the fact that the majority of them, coming from a domain, which, by essence, is characterized by a predominant usage of English multi word units, are given in their original English version to keep their meaning even when applied to other language use cases and contexts. The present use case implies the involvement of Italian Cybersecurity institutions and training organizations, so the transfer learning process is essential to guarantee the uniformity of key concepts in the Cybersecurity domain either found in sector-oriented magazines and laws or regulations (also in grey literature).

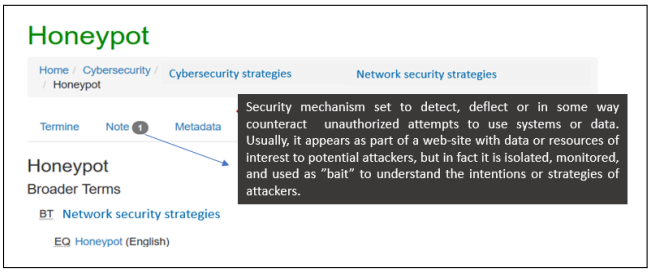

Fig. 1. Thesaurus representation of *Honeypot*.

To give an example, the term *Honeypot* has no corresponding term in Italian language; consequently, to maintain its practical meaning, terminologists in the transfer learning operations should leave the English form as to provide a strong homogeneous informative flow within organizations that are supposed to share common official knowledge (see Figure 1 above to see the use of *honeypot* in a thesaurus structure). To enable users to refer to a uniform resource that spreads specialized information onto several technical databases in a unique modality, the structure of the thesaurus allows the insertion of a Scope Note (*SN*), that is a targeted definition of the terms. This definition is taken from authoritative sources, such as sector-oriented glossaries, standards, official guidelines, etc. This additional feature provides a better unified structure between systems shared under different languages. Moreover, one of the main outcomes of this research activity is strictly linked to the possibility of integrating the Italian thesaurus and the ontology in an automatic threats recognition system, which is intended to monitor terms and concepts and to detect the appearance of new ones without much human effort.

Some of the considered resources to build the source corpus useful to obtain a list of representative terms are hereafter summarized. Representative terms synthesize the concepts belonging to a specific domain and provide the starting model to realize, in a second step, an ontology for Cybersecurity, which is, consequently, based on the structure created for the Italian thesaurus. The ontology has been developed with the goal of representing the classes linked to each other through more precise properties that could, at times, specify the interconnections between them better than a flat visualization that belongs to a thesaural organization of terms. The paper is structured as follows: Section II presents the theoretical background for both thesauri and ontologies in order to highlight which are their main characteristics and the advantages in using them for organizing and representing highly technical domains. Section III gives an overview of the state of the art, presenting related works focused on Cybersecurity information management, both in English and Italian, and on the construction of KOSs. Section IV describes the construction of the Italian thesaurus for Cybersecurity and its enhancement through an ontological representation. Section V provides a discussion about the main advantages derived from exploiting thesauri and ontologies in the described Italian use case. Finally, Section VI sums up the key issues underlined in the paper giving some overall remarks and future perspectives.

II. Background

In this section, a theoretical background is presented to describe and clarify characteristics, purposes, differences and advantages of the two main structured KOSs, i.e., theusauri and ontologies. This will introduce the reader to the approach proposed in Section IV to build such terminological resources for the Cybersecurity domain.

A. Thesauri

Thesauri's main scope is that of structuring information and organizing it in a layered network of semantic connections, and its management and usability is piloted by KOSs functionalities [5][6]. As Soergel affirms in his work, "A thesaurus is a structured collection of concepts and terms for the purpose of improving the retrieval of information. A thesaurus should help the searcher to find good search terms, whether they be descriptors from a controlled vocabulary or the manifold terms needed for a comprehensive free-text search — all the various terms that are used in texts to express the search concept" [7]. The way thesauri are structured follows standardized rules that should be respected, as the ones included in the ISO standards 25964-1:2011 and 25964-2:2013 [8][9], and the objective of uniforming a lexicon meant to be a reference for a community of domain-oriented users is pursued. A thesaurus should provide a reliable and a well structured semantic means to guide the understanding of technical terms representing concepts belonging to a specific field of knowledge. Its indexing function proves to be helpful in the way the users are able to analyze documents according to an informative organization of descriptors. In other words, the abstraction of knowledge occurs indirectly by exploiting terminological units that take on the status of descriptors or indexing units. The latter is the element that language uses to describe, synthesize and extract information from documents [10]. Thesauri's terms undergo both quantitative and quality control. Quantity control refers to thesaurus' terms selection among those that represent in a better way the concepts of the domain of study. These latter become descriptors of the thesaurus (i.e., preferred terms) and usually are followed by the non-preferred terms that act as synonym entries, e.g., *Malicious software* is the preferred term instead of *Malware* in the Italian Cybersecurity thesaurus. In detail, as suggested by the mentioned standard, countable terms have to be expressed in plural form (trees and not tree), and semantically the control is always granted by the respect of the biunivocal relationship existing between terms and concepts (only one concept corresponds to a term and viceversa). That means that the ambiguity of the natural language is controlled and reduced to zero through the use of a limited set of terms (indexing terms) that represent the concepts in a given domain. In this scenario the user who selects a search term and the indexer who chooses indexing terms are both guided to use the same term for the same concept [8]. Thesauri present three main standardized

forms of connections that are generated for structuring the information, and five abbreviation codes used to represent such relationships within the controlled vocabulary:

1) Equivalence relation, with the tags Use (USE) and Used For (UF), expresses the synonymy property:
   - Usage:
     Cyber minacce UF *Cyber Threat Actors*; *Cyber Threat Actors* USE *Cyber Minacce*
   - Acronyms:
     Virtual Private Network UF *VPN*; *VPN* USE *Virtual Private Network*
   - Synonymy control:
     Cyber attacks UF *Cibernetic attacks*; *Cibernetic attacks*; USE *Cyber attacks*
2) Hierarchical relation, with the tags Broader Term (BT) and Narrower Term (NT), exists when having two concepts and one of them is part, or is included in the other:
   - Whole/parts:
     Vulnerabilities NT *Software vulneabilities*; *Software vulneabilities* BT *Vulnerabilities*
   - Class/member:
     Logic bombs NT *Elk Cloner*;
     Elk Cloner BT *Logic bombs*
3) Associative relation, with the tag Related Term (RT), covers associations between pairs of concepts that are not hierarchically related [8]:
   Cyber war RT *Cyber weapon*.

The aforementioned standards also guide the way terms should be defined to indicate a unique and unambiguous meaning. The use of a Scope Note is useful when an indexer needs to fix the boundaries of a concept within a domain. Scope Note is marked with the tag SN. An example of SN can be:

> *Phishing* - SN: Tricking individuals into disclosing sensitive personal information through deceptive computer-based means.

The choice to privilege a thesaurus structure instead of other semantic resources, such as glossaries or taxonomies, relies on its peculiarity of managing the representative terms of a specific domain as an entangled network of semantic relations that guide the comprehension of a conceptual model proper of a field of knowledge to be studied [11].

B. Ontologies

The term ontology, which has been borrowed by the Artificial Intelligence (AI) community from phylosophy, gained new definitions and found a broad spectrum of applications in various branches of computer science [12]. In AI, an ontology is considered to be an engineering artefact, which is constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary. Gruber defines it as "An explicit specification of a conceptualization" [13], so, in simple words, a formal specification of a domain of knowledge. In order to formally represent a certain domain, ontologies use a set of constructs describing the world in terms of classes, properties, and individuals. To enrich the formalization, other constructs are used for expressing complex descriptions in terms of relations between classes, cardinality, equality, etc. Consequently, it is possible to say that an ontology consists of a set of definitions of classes, properties, and individuals, together with a set of axioms (i.e., formal restrictions) expressing the relations between classes and properties, and a set of facts about particular individuals. Just like thesauri, ontologies define a common vocabulary (for a specific domain) and a shared understanding. We can have different ontologies according to the used level of formalism: (i) light-weight ontologies (i.e., ontologies that represent only the hierarchical level of concepts and relations in a domain, so, more commonly, taxonomies); and (ii) heavy-weight ontologies (i.e., lightweight ontology enriched with axioms used to fix the semantic interpretation of concepts and relations). Ontologies are used to share knowledge between people, agents, and software thanks to their characteristics of enabling the reuse of domain knowledge and making domain assumptions explicit. Another important feature is that through ontologies it is possible to represent both domain knowledge and operational knowledge and reuse them separately, enabling in any case automated reasoning. The importance of an ontology as a means of structuring knowledge is well recognized in different areas, such as, knowledge representation, knowledge management, natural language processing (NLP), multi-agent systems, database integration, web services, and others. The literature is full of significant academic research devoted to the development of a theoretical and practical basis of ontology technology. Among others, the most notable developments have been the world wide web consortium standardization of expressive representational languages for publishing ontologies on the web [14] [15]. From a practical point of view, the methods followed for building ontologies observe basic principles that can be found in guidelines like the one published by Noy and Mcguinness [16] or Bourigault [17].

The OWL language helps in building formal, sound and consistent domain-specific terminologies, and provides a standard web accessible medium for interoperability, access and reuse. OWL uses RDF (Resource Description Framework) for its syntax, the prescribed framework for representing resources in a common format, describing information in the form of *subject-predicate-object* triples, thus enabling to represent them in the form of a graph. Three different OWL sublanguages can be used according to the formalism we want to give to our ontology and to the performances in reasoning and inference we want to obtain: OWL Lite, OWL DL, and OWL Full. The first sublanguage is the least powerful one, in fact it allows to represent taxonomies and uses less

constructs (it includes cardinality restrictions). For this reason it has the lowest computational complexity among the OWL sublanguages. The second one, OWL DL (i.e., Description Logic) provides a more formal representation since it imposes restrictions on the usage of OWL/RDF constructors. This sublanguage is used when the maximum decidable expressivity is required and is able to maintain computational completeness (that means that all conclusions are computable). Finally, OWL Full is the most expressive one, since it uses all the OWL language primitives and all of the RDF Schemas (RDFS) and, with respect to the other two sublanguages, it is undecideable, semantically difficult to understand and to work with, and, as a result, standard automatic reasoning techniques cannot be applied. Differently, because of its formalism, OWL DL allows reasoning and inference. Reasoning is the act of making implicit knowledge explicit. To infer knowledge from ontologies, reasoning engines are used, which allow determining also subsumption, classification, equivalence, and identifying ontology inconsistencies [15].

Ontology similarities with thesauri can be easily identified after this theoretical description. In particular, both describe and organize a domain, include concepts and relations between them; they use hierarchies, and describe instances belonging to concepts. Both of them can be applied for information management, for cataloguing and in search engines. However, several differences must be considered. First of all, thesauri had as their original purpose that of being used in librarian contexts as indexing tools and controlled vocabularies. So, it is understandable that they are thought to represent knowledge in a less formal and comprehensive way with respect to ontologies. On the contrary, because of their philosophical origin, ontologies are characterized by a high level of conceptual abstraction, which is accepted, and formal ways of describing domain knowledge. Regarding their structure, as seen above, ontologies are characterized by the explicit representation of the types of relationships and by the use of powerful formalisms, which are not possible to define within thesauri (e.g., axioms, relationships, cardinality). Therefore, to represent hierarchical relations between classes and subclasses, two declared relations are used, i.e., “is-a” and “kind-of”, while, to represent meronymy relations between classes, the “part-of” relation is employed. By contrast, in the thesaurus those relationships are treated as hierarchical relationships. Finally, the associative relations in an ontology are made explicit according to the exact connection (predicate) that exists between two classes. For example, taking up concepts already used in Section II-A , *cyber war* `RT` *cyber weapon*, is specified in an ontology as *cyber war* `uses` *cyber weapon*, where “used” is the ObjectProperty.

The interoperability of semantic resources like thesauri and ontologies, is given by the principle of linked open data [18][19][20], which guarantees a shareable knowledge organization system that can facilitate the coordination among several users for different terminological tasks. To generate a language that can guarantee a higher form of interaction between informative systems, without losing the exact meaning of the shared information, the ontology seems to route towards a constant reuse of the managed information by providing conceptual representations of a domain [21][22].

III. Related works

When terminologists’ activity involves the construction of knowledge organization and representation systems, the phase of taking into account which could be conceived as gold standards represents a key step in order to align the information retrieved by source corpora to texts that represent the reference standards [23]. The research activity presented in this paper starts as a monolingual - Italian - study for Cybersecurity terminology. Therefore, the starting point to develop an Italian controlled vocabulary on Cybersecurity has represented the census of the gold standards. Among the existing examples of Cybersecurity glossaries and vocabularies, of great importance are: for English, the ones contained in the NIST 7298 [24] and ISO 27000:2016 [25] standards for Information and Communication Technologies (ICT) security, and, for Italian, the Italian book ”*Libro Bianco*” (White Book for Cybersecurity) realized by the National Laboratory of Cybersecurity of the Consorzio Interuniversitario Nazionale per l’Informatica (CINI) [26], which thoroughly sheds light on the key issues related to Cybersecurity guidelines and on the latest related episodes that have changed the way to defend informative systems and to conceive some specific concepts proper to Cybersecurity. Another relevant existing resource for Italian is the ”*Glossario Intelligence*” [27], a technical glossary published by the Italian Presidency of the Council of Ministers, which contains several terms belonging to the Cybersecurity domain and which has been used as a basis for the creation of the Italian thesaurus and the ontology for Cybersecurity under investigation.

With respect to ontologies, it is worth mentioning the works targeted at the creation of ontology models for Cybersecurity, i.e., [28][29][30], and the studies focused on the approaches for developing an architecture for Cybersecurity standards [31] and enterprise’s Cybersecurity metrics [32]. In particular, in [33] an ontology has been designed to integrate data from different heterogeneous sources, in the absence of a common terminology, offering a sufficiently complete knowledge on the possible threats, thus allowing organizations to perform reasoning and support decision-making processes related to security. Another study proposed a reference ontology for Cybersecurity operational information, developed, as in our case, in collaboration with Cybersecurity organizations, and which had the aim to review industry specifications. Here, types of Cybersecurity information are defined along with the roles and operation domains (see [34] for details). Finally, a more recent work describes the development of an ontology of metrics for Cybersecurity assessment [35]. This ontology is based on determining the concepts and relations between primary features of initial security data and forming a set of hierarchically interconnected security metrics. Application of the approach is shown on a case study. The main feature of this work is the representation of security metrics as separate

instances of the ontology, which allows using the relations between the concepts of ontology for calculating integral metrics reflecting the security state.

Processing the information belonging to specific domains of interest involves the analysis of those documents which semantically tend to represent concepts through a technical language [36]. The creation of terminological databases follows some given criteria linked to gathering the related documents that have to constitute the reference corpus from which terms can be retrieved [36]. To achieve this first informative structure, the corpus firstly aims at including documents that can represent the domain in an official way [37], i.e., the gold standards [38], collecting a terminological standardized repository made up of terms that are meant to be closely specific to the technical field of knowledge under review [39].

To obtain a matching system between the terminology shared by a community of experts from a particular domain and the terms contained in a list derived from the processing of a reference corpus, the documents gathered in the corpus undergo a process of terminology extraction, which shall compare the equivalence between the representative terms of a domain with the ones of the gold standards [40].

This last step is usually implemented by exploiting semi-automatic term extraction tools. Nazarenko *et al.* [41] and Loginova [42] gave in their works detailed lists of several tools for extracting terminologies from texts. With regards to the Cybersecurity domain and the research activity treated in this paper, various existing sources, both in English and in Italian, have been analyzed in order to retrieve an accurate terminological basis from which to build a more sophisticated semantic resource to guide the knowledge representation process. The intent of this project task, as aforementioned, is to provide an Italian resource, firstly conceived as a thesaurus, to configure the terminology of Cybersecurity in a network of semantic relations that can better orientate to a lexical understanding of specialized concepts represented by terms belonging to this field. The goal of this research activity is also based on the reuse of the terms contained in the thesaurus to realize in a consequential way an ontology system that could support the inclusion of customized properties between classes and more comprehensively clarify the associative relationships used in the thesaurus [43][44][45]. This represents the reason why ontologies can usually be considered as resources that can provide a more exhaustive and explicit frame for knowledge representation.

IV. The OCS Project

In this section, the project use case is presented. The first part is focused on the description of the Cybersecurity context and the Italian Cybersecurity Observatory (OCS) scopes and services. The second part presents the thesaurus itself for managing the information about Cybersecurity and its enhancement through its migration into an ontology system.

The main objective of the activity, as mentioned before, is the creation of a thesaurus in Italian language to be used as a semantic tool to organize the terminology related to Cybersecurity, and to be inserted amongst the services of the online platform of the Italian Cybersecurity Observatory [46]. The OCS online platform is a joint work with the experts of the Cybersecurity domain that aims at gathering different services to guide the comprehension of the phenomena occurring in this field of study. For instance, apart from the semantic tools section, to which the Italian thesaurus and the ontology for Cybersecurity belong, this web service includes the analysis and detection of tweets, threats, vulnerabilities, exploits, spam mails, attacks, malware, self-assessment.

The convergence of the semantic tasks with the experts of the domain can be achieved in considering their documentation collections, consisting, among others, of the lists derived from the Common Vulnerabilities Exposure (CVE)[1], or of internal detections of the main cyber attacks, as sources to be used to update the terminology of the domain to be represented.

Indeed, the list of vulnerabilities, the spam detections or the analysis of the latest cyber threats, could represent, in a future perspective, the meeting point between the goal, by the OCS platform, of sharing technical information to defend informative systems and, by terminologists, of providing extra knowledge that can empower the terminological organization of the domain. In this way, both the thesaurus and the ontology can undergo a rethinking phase both on new highly technical term inclusion level and, consequently, the relational one.

A. The Cybersecurity context

The Cybersecurity domain is mainly characterized by a technical terminology. Given that Cybersecurity is a synergy of different sub-fields, the schematization of this specialized domain reflects this high level of heterogeneity. Cybersecurity is permeated by: (i) its multidisciplinary nature that involves Information and Communication Technologies (ICT) and its sub-areas, such as, audiovisual techniques, computer software, electronics; (ii) its specificity with respect to technical and standardized terms; and (iii) its cross-fielding thematic coverage, i.e., computer science field, legislative systems, regulations. Given these premises, the treatment of its internal language, which derives from the textual content extracted from the source corpus documents, is meant to be managed by formal semantic systems in order to obtain shareable standardized lists of the domain's representative terms, organized according to their semantic relations, which, in turn, will orientate the understanding of the conceptual model of the domain [47].

As can be observed by looking at Figure 2, the OCS website, developed for the purposes of spreading the information about Cybersecurity for the Italian community of experts and common users, registered many views on its overall level range. This high number of users coming from several countries denotes the significant interest the organization of the platform has. Nonetheless, the superior percentage of Italian users shows how the target language played an important role in orientating the ways in which the technical structuring of

[1] https://cve.mitre.org/

information about the domain has been set up. The thesaurus and the ontology presented in this paper have been included inside the OCS web page as two tools that provide a semantic outline about the information meant to be structured on the Cybersecurity domain. Even though the numbers reported are not remarkably outstanding, it can be stated that both of them have received attention especially during two Italian events during which they have been presented to an audience.

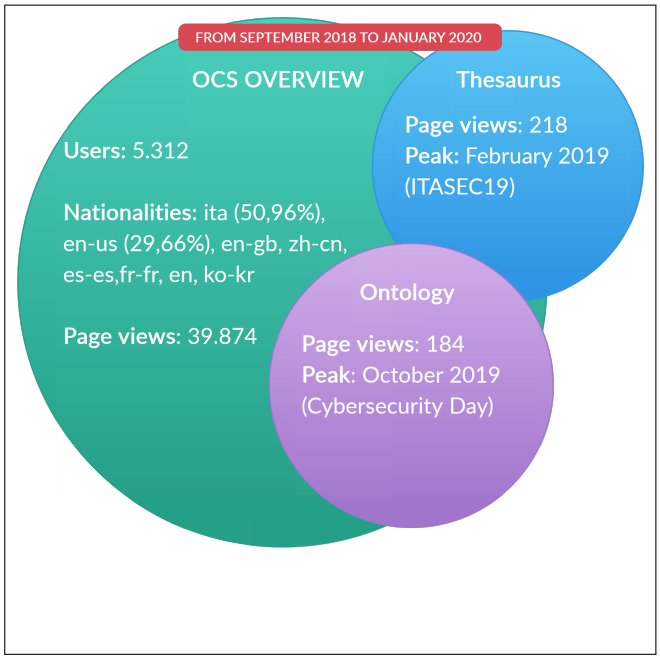

Fig. 2. Statistics OCS website.

B. The Italian thesaurus for Cybersecurity

The main focus of this paper is the creation of the Italian thesaurus on Cybersecurity for the OCS project [46], carried out in collaboration with the Institute of Informatics and Telematics of the National Research Council.

The methodology followed for the realization of the thesaurus covered classical sequences. As primary step, the terminology to be included in the thesaurus has been extracted from reliable sources which made up the corpus characterized by documents distinctively selected for their content oriented to Cybersecurity issues [37]. This collection of texts made the information retrieval highly oriented to the domain to be represented [48], and covered different types of documents, such as, standards and laws [49], Cybersecurity-related magazines or guidelines and certifications. The conceptual content of these documents was meant to be processed to obtain lists of terms (a glossary) sorted according to statistical measures able to provide a first semantic organization [50]. Indeed, the second phase concerned the semi-automatic processing of the information included in the source corpus by exploiting a term extractor software [**?**] (more specifically the Italian native tool, *Text to Knowledge* (T2K)) [51] that provided, as outputs, lists of terms ranked according to their occurrence's value in the texts. Terms selection has been based on frequency, in particular terms with the highest scores in TF-IDF values have been considered as candidate terms to be part of the Italian thesaurus for Cybersecurity. The list of the most representative terms accompanied by their frequency scores has undergone an evaluation process carried out by a group of domain experts.

Indeed, only once having received the validation by domain experts,– the third phase of the methodology –, the terms have been selected as candidate terms to be integrated in the thesaurus and their semantic relations with other terms of the domain, derived from the corpus, have been created. The current Italian Cybersecurity thesaurus contains 246 candidate terms, already validated by domain experts collaborating on the project, and mapped to the taxonomies contained in the main gold standards for Cybersecurity, i.e., NIST 7298 [24] and ISO 27000:2016 [25]. The alignment with the terms contained in the standards for ICT security granted a coordination between the knowledge shared by an international Cybersecurity community of experts and the one represented in the structured thesaurus, which is composed of preferred terms selected amongst those extracted by the T2K tool as the most frequent inside the source documents. In order to carry out a matching configuration with the standards as predictable and stable as possible, the terms included in the standards, and selected with the support of domain experts as key guidance representing the domain, have been translated using the Interactive Terminology for Europe (IATE) term banks [52]. This is considered an important step given the instructive purpose of the application, i.e., the use of the thesaurus in the web portal of the Cybersecurity Observatory. The main entries in the Italian thesaurus for Cybersecurity are the four macro categories finely selected from the extracted glossary, also according to the frequency of terms, and from the mapping with the standards alongside the approval by the domain experts. These macro categories are:

- Cybersecurity;
- Cyberdefence;
- Cyberbullism;
- Cybercriminality.

Almost each of the candidate terms included in the thesaurus network, generated by the semantic relations among the terms, are accompanied by their definitions, i.e., *Scope Note (SN)*, which helps in understanding the terms in their specific contexts giving their definition taken from the source documents [53].

For a better understanding of the actual size of the Italian Thesaurus for Cybersecurity, Table I gives a metrics of the numbers of terms, as well as of the semantic relations (SemRel).

TABLE I. Features of the Italian thesaurus for Cybersecurity.

	Terms	SemRel	Non-preferred Terms	SN
Total	246	280	33	74

TABLE II. Cybersecurity ontology metrics.

Metric	Total
Axiom	640
Logical axiom count	316
Declaration axioms count	233
Class count	157
Object property count	37
Data property count	7
Individual count	31
Annotation Property count	5
CLASS AXIOMS	
SubClassOf	58
EquivalentClasses	0
DisjointClasses	24
OBJECT PROPERTY AXIOMS	
SubObjectPropertyOf	7
InverseObjectProperties	1
FunctionalObjectProperty	1
TransitiveObjectProperty	0
SymmetricObjectProperty	1
AsymmetricObjectProperty	0
ObjectPropertyDomain	40
ObjectPropertyRange	39
DATA PROPERTY AXIOMS	
SubDataPropertyOf	1
DataPropertyDomain	8
DataPropertyRange	5
INDIVIDUAL AND ANNOTATION AXIOMS	
ClassAssertion	31
AnnotationAssertion	89

C. Ontology enhancement

Another activity of the OCS project has also been focused on the migration of the thesaurus elements into a more formal semantic resource, i.e., an ontology, to better organize and represent the information about Cybersecurity, addressed to users who want to get closer to this field of knowledge [54]. Details on the ontology structure are provided in Table II. Among the main objectives in rengineering a thesaurus into a system working with OWL language there is that referring to the capture of significant real time new terms occurrences in the future, especially following the updates given by the major official sources in the Cybersecurity domain. Indeed, what ontologies allow more than a thesaurus is to exploit the query system operations that enable users to activate reasoning engine operations which are meant to infer semantic connections from several resources given in input as conceptual models. The formalization of a thesaurus into an ontology is a task that has been attracting much interest. In fact, in the literature, different approaches have been proposed for reusing thesaurus semantic content to build ontology meta-models and to populate knowledge bases in different domains, see for example [43][55][56].

The need for migrating the content included in the thesaurus into an ontology lies in the decision to better clarify the associative relationships between the terms of the thesaurus [57]. In particular, the flat modality in which the associative relationship between terms is represented in the thesaurus, i.e., via the RT relation, turned out to be not fully satisfactory in the seek of getting a complete terminological outline for Cybersecurity [58].

As shown in Figure 3 and Figure 4, there is a clear distinction between the two systems used to organize and represent the terminology belonging to Cybersecurity. The example taken into account to represent the differences is referred to the semantic relationship linked to the idea of opposition, i.e., *Spoof* and *Antispoof*: in the thesaurus, even though a definition is present (within the black square), which corresponds to the Scope Note (*SN*), proper to thesauri, giving many details on the context from which terms come from, the "opposition" is not so well represented because it is only shown through the associative relation (*RT*) [8] between these aforementioned terms without giving other explications on the way the two terms are related as the OWL language does.

On the other hand, in the ontology, these two concepts are connected through the *ObjectProperty* "HasAsContrary" that helps in considering the *Domain* and the *Range* as linked by a precise relationship.

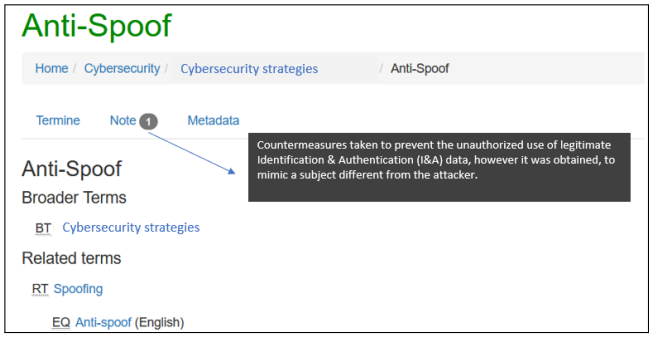

Fig. 3. Thesaurus representation of the semantic relationship that describes opposition.

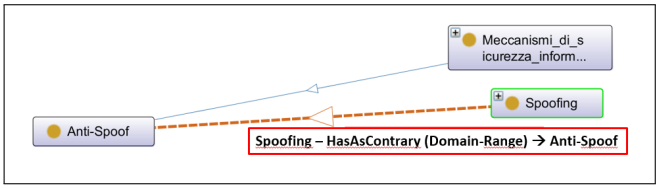

Fig. 4. Protégé representation of the semantic relationship that describes opposition.

Another representative case is depicted by Figure 5 and Figure 6, which show how a thesaurus sometimes provides a weak visualization of some attributes associated to a concept.

In the following case, the relation that had to be demonstrated was related to several attributes that security properties proper to informative systems own. For this specific purpose, the ontology resource gives more advantages in the visualization of the informative structure allowing a higher accurate organization and representation of the attributes related to the concepts. In detail, the main difference that makes ontologies

a good semantic means to represent the conceptual model connected to certain semantic classes is related to the fact that, in this case, the security properties, i.e., integrity, authenticity, confidentiality, availability, reliability, non-repudiation, and privacy, are represented as *Data Properties* and are conceived as attributes. In the thesaurus, as shown in Figure 6, they are related to the hyperonym *BT* "Data" and are represented as its specific terms, i.e., the *NT* [9].

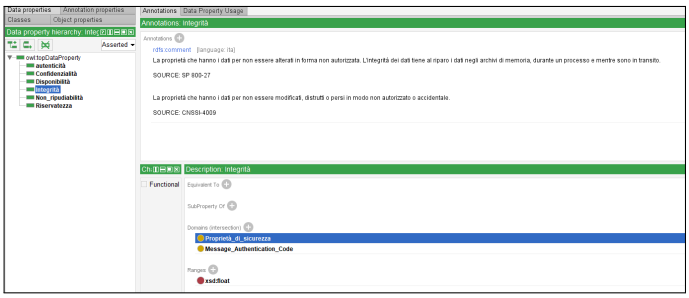

Fig. 5. Ontology representation of Security properties as *Data Properties*.

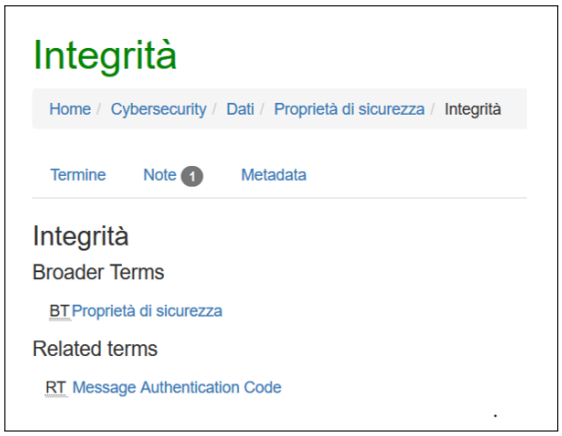

Fig. 6. Thesaurus representation of Security properties as hierarchical relations.

As mentioned before, the ontology has been forged under the basis of the thesaurus structure to organize the Italian terminology about Cybersecurity. For this reason, and considering that the main purpose of the ontology is terminology control and the appropriate semantic representation of the domain concepts, the OWL sublanguage selected for this scenario is OWL Lite. The connections between the terms have been transposed to the ontology object properties, and referred to the information contained in the source corpus documents. To increase the level of accuracy and domain-oriented information representation, the ontology has been enhanced using pattern path-variables configuration [59]. In particular, after having collected a group of passive verb pairs from the Italian Cybersecurity source corpus, a filtering procedure over the most technical ones related to the domain has been launched. From a list of verbs that have been considered domain-dependent, several queries in the source corpus have been run and analysed in order to create the associative connections among the concepts [60]. As Table III shows, the relations that have been retrieved from the reference corpus by using certain pattern paths are very detailed and they are progressively being added to the existing ontology, which has been developed by migrating the content of the Italian Cybersecurity thesaurus. The aim is to guarantee a more precise semantic system that can structure the interconnections among Cybersecurity concepts with the help of interoperable languages. In the case of ontologies, the *Object Properties* are the ones to cover the purpose of providing a more targeted form of associative relationships to represent the information of the domain.

To give a clearer idea of how these associations have been reported into the ontology so far, with the perspective of continuously augmenting the range of relations, Figure 7 gives an highlight of how, for example, the concept *Backdoor* has been connected with *Malware* and *Cracker* by using complementary patterns configurations on Protégé console. Up to now, 160 new associative relations referred to the domain verbal constructions have been selected among the semantic information contained in the texts making up the source corpus, and they are currently being analysed from a linguistic point of view in the co-occurrency level and text scope to increase the semantic relationships meant to be represented. This last passage assumes the activity by terminologists to retrace the semantic entangled network in the text correspondences, and doing so, isolating other constructs as drills of new connections.

TABLE III. Associations retrieval in ontology by patterns configurations.

Associative relationships List			
Aggirare (*bypass*)	Attaccare (*attack*)	Sfruttare (*exploit*)	Attivare (*activate*)
worm→software antivirus	exploit sql injection→web applications	crackers → vulnerabilities	backdoor ← malware
cracker → cybersecurity	script kiddies → DDoS	trojan horses → vulnerabilities	trojan → cyber attacks
virus→ cybersecurity	network file systems→ DNS spoofing	spam → botnet	payload← virus/worm

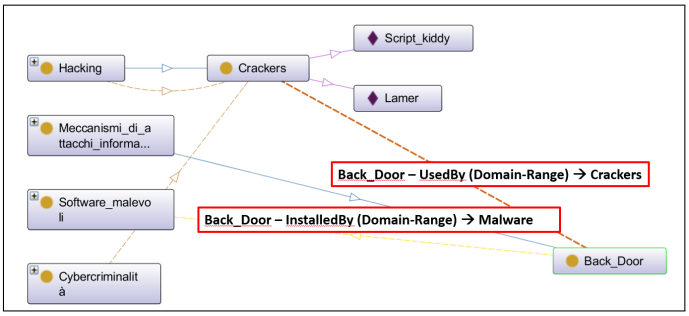

Fig. 7. Additional Object Properties through patterns path-variables.

V. Discussion

Across the phases that have characterized this research activity towards the realization of two semantic sources to

monitor and manage the terminology of Cybersecurity in Italian language, the configuring procedure for the thesaurus and ontology development proved to be different in their application to the use case. By describing the steps carried out to build the Italian Cybersecurity thesaurus, the importance of providing a semantic structure that could be as much reliable as possible with respect of the information about the domain has been underlined. This reliability system mainly focuses on the ways a thesaurus can guarantee a reflection of the domain by using a semantic relationship network to connect terms with each other and provide a guided orientation to the organization of the domain knowledge. The connections among the representative terms dependent from the domain of study have been structured following the standard guidelines that give the basis for the arrangement of hierarchical, equivalence and associative interrelations. Nonetheless, the thesaurus outline at times proved to be less accurate in the way it depicts the association among certain terms, in particular for what concerns the usage of the *Related Term* association, which shows some vagueness in how it matches domain-oriented terms mainly because of its lack of deeper semantic descriptions. Therefore, the research activity has been finalized to create another semantic resource able to make the semantic relations between the domain concepts, i.e., the ontology, explicit. We observed that this latter knowledge representation system allowed a more customized organization of the concepts that facilitates the process of combining the semantic links. The ontology has also been implemented, in a latest phase, with the inputs resulted by the execution of some pattern configuration approaches. The use of recurrent variables to be searched in the source corpus proved to be an efficient means through which concepts of the Cybersecurity domain could have been correctly correlated. In fact, we used different domain-oriented verb pairs to show how the structure of the ontology, which has been built following the thesaurus' outline, has been enhanced.

Although thesauri and ontologies belong to the same family of knowledge organization systems and some of their functionalities are the same (e.g., their use for improving information retrieval and knowledge organization), they are built for different purposes. In fact, it has been demonstrated in this contribution that ontologies allow higher formal representation of knowledge for a given domain, by providing explicit relationships between concepts, disjunctions, by applying data properties for each concept or instance and by providing restrictions that avoid ambiguity in the representation of the meaning and the context of use of a concept and their terms in the domain of reference. However, the two semantic resources might be used together or, as widely demonstrated both in this paper and in the literature, one can be reused to build or populate the other, thus they prove to complement each other, improving the end user's search experience.

The natural structural rigidity of thesauri, given by the use of *a priori* defined semantic relationships (hierarchical, associative and equivalence), seems to be a point against these type of controlled vocabularies; by contrast, such weakness seems to be overcome by the flexibility, scalability and reusability of ontologies that, as stressed by the semantic staircase of Blumauer and Pellegrini [61], compared to other KOSs, bring to a highest level of semantic richness thanks to an internal formal description of concepts. This latter combines a system of relations and properties of the concepts themselves.

Despite this, one of the strengths of the thesaurus compared to the ontology, when used in a specialized domain, is its greater capacity to eliminate ambiguity between the terms through the use of synonymy control [2] and the choice of preferred terms, compared to non-preferred terms for representing the concepts. This guarantees a standardization of technical terms in specialized domains, which can help in the process of unifying, and, by consequence, sharing, a specific field of knowledge's terminology.

VI. Conclusion

The objective of this paper concerned the presentation of the main advantages that could be achieved by using two different types of KOSs, i.e., thesauri and ontologies, to organize and represent a technical domain of study. The field of knowledge on which this paper focuses on refers to that of Cybersecurity, and the main task described is specifically linked to its specialized terminology management.

At the beginning of this paper a general overview of Knowledge organization and representation systems has been provided, successively the analysis has been addressed to the thesaurus organization system overview. In detail, the paper underlined the way this semantic monitoring tool has proved to be a reliable system to structure the information derived from heterogenous sources belonging to the Cybersecurity domain, which is widely characterized by technical terms. Concurrently, attention has also been given to the comparison between the modality of representing in the thesaurus some of the relationships existing among terms, which represent the relevant concepts of the domain, with the ones proper to ontologies through OWL language. The perspective has been oriented to provide a demonstrative outline of ontology peculiarities and advantages when using an existing thesaurus, like the one created in the Italian OCS project framework, as a basis for building the meta-model and populating the knowledge base. The perspective of the research activity both for the thesaurus and the knowledge base in OWL is oriented towards a terminological population extension, and this will involve relationships and restrictions where needed, and new evaluations to be executed. Starting from this objective, pattern configurations have been added as means to retrieve additional relations among domain-oriented concepts. Indeed, we have observed that the use of recurrent linguistic structures helped in trace back which could be considered as genre specific relations. Another motivation that lies behind the choice of taking into account pattern constructs is that of improving the preciseness of the associative relationships proper to thesauri that sometimes proved to be rather vague, i.e., *RT* relation. By selecting some verb pairs targeted to the domain of study, is possible, for instance, to create a new range of more accurate

Object Properties in the ontology, and, consequently, enhance the system that has been converted starting from a thesaurus. This last step clearly implies a rearrangement of the source thesaurus, which will continue to be updated in the source texts to provide a representative set of terms that helps in understanding the technical range of information.

Future works will include a translation in other languages (firstly English) to allow, within the OCS project team, the automatic recognition of cyber threats even from non-Italian sources and improve the thesaurus/ontology usability and sharing them also at an international level. Moreover, the remainder of this work targets at taking into account the insertion of several other types of documents to be part of the source corpus. In particular, following the perspective of getting updated on the changes related to the Cybersecurity domain, documents shall be taken from the social media world, adjusting all the analysis related to the processing of information to the treatment of texts written in a specialized form.

REFERENCES

[1] C. Lanza, E. Cardillo, M. Taverniti, and R. Guarasci, "Knowledge representation frameworks for terminology management in cybersecurity: The ocs project use case," in SEMAPRO 2019, The Thirteenth International Conference on Advances in Semantic Processing, U. o. A. S. G. P. L. U. o. H. A. F. Michael Spranger, Hochschule Mittweida, Ed., Porto, Portugal, September 2019.

[2] M. Zeng, "Knowledge organization systems (kos)," Knowledge Organization, vol. 35, pp. 160–182, 01 2008.

[3] Cybesecurity osservatorio. https://www.cybersecurityosservatorio.it\. Accessed: 2019-08-08.

[4] W3C Web Ontology Language (OWL). https://www.w3.org/OWL/. Accessed: 2019-08-08.

[5] R. Davis, H. Shrobe, and P. Szolovits, "What is a knowledge representation?" AI Magazine, vol. 14, p. 17, 03 2002.

[6] A. Miles and S. Bechhofer, SKOS Simple Knowledge Organization System Reference, ser. W3C Recommendation. United States: World Wide Web Consortium, 8 2009.

[7] D. Soergel, "The art and architecture thesaurus (aat): A critical appraisal," Visual Resources, vol. 10, pp. 369–400, 01 1995.

[8] ISO 25964-1:2011 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval, International Organization for Standardization, August 2011.

[9] ISO 25964-2:2013 Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies.

[10] M. Taverniti, "Tra terminologia e documentazione: estrazione automatica di voci indice da corpora documentali della pubblica amministrazione," Ainformazioni, vol. 1-2/2018, pp. 227–238, 2008.

[11] V. Broughton, Essential Thesaurus Construction. Facet, 2006.

[12] N. Guarino, ""formal ontology and information systems"," in *In Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS-1998)*, 1998.

[13] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, pp. 199 – 220, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1042814383710083

[14] A. Gómez-Pérez and O. Corcho, "Ontology specification languages for the semantic web," *IEEE Intelligent Systems*, vol. 17, no. 1, p. 54–60, Jan. 2002. [Online]. Available: https://doi.org/10.1109/5254.988453

[15] S. Staab and R. Studer, *Handbook on Ontologies (International Handbooks on Information Systems)*. SpringerVerlag, 2004.

[16] N. F. Noy and D. L. Mcguinness, "Ontology development 101: A guide to creating your first ontology," Tech. Rep., 2001.

[17] D. Bourigault and N. Aussenac-Gilles, "Construction d'ontologies á partir de textes," pp. 11–14, 01 2003.

[18] A. A. Shiri and C. Revie, "Thesauri on the web: current developments and trends," Online Information Review, vol. 24, no. 4, pp. 273–280, 2000.

[19] D. Soergel, "The art and architecture thesaurus (aat): A critical appraisal," Visual Resources, vol. 10, pp. 369–400, 01 1995.

[20] M. van Assem, V. Malaisé, A. Miles, and G. Schreiber, "A method to convert thesauri to skos," 06 2006, pp. 95–109.

[21] N. Guarino, D. Oberle, and S. Staab, "What is an ontology?" Springer, Berlin, Heidelberg, 05 2009, pp. 1–17.

[22] D. W. Embley, S. W. Liddle, D. W. Lonsdale, and Y. A. Tijerino, "Multilingual ontologies for cross-language information extraction and semantic search," in *ER*, 2011.

[23] A. R. Terryn, V. Hoste, and E. Lefever, "A Gold Standard for Multilingual Automatic Term Extraction from Comparable Corpora: Term Structure and Translation Equivalents," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, N. C. C. chair), K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, and T. Tokunaga, Eds. Miyazaki, Japan: European Language Resources Association (ELRA), May 7-12, 2018 2018.

[24] R. Kisserl, Glossary of Key Information Security Terms, National Institute of Standards and Technology, May 2013, NISTIR 7298 Revision 2.

[25] ISO/IEC 27000, Information technology — Security techniques — Information security management systems — Overview and vocabulary, International Standard, February 2016.

[26] R. Baldoni, R. De Nicola, and P. Prinetto, Il Futuro della Cybersecurity in Italia: Ambiti Progettuali Strategici Progetti e Azioni per difendere al meglio il Paese dagli attacchi informatici. Laboratorio Nazionale di Cybersecurity (CINI) - Consorzio Interuniversitario Nazionale per l'Informatica, 2018.

[27] Presidenza del Consiglio dei Ministri - Sistema di informazione per la sicurezza della Repubblica, Il linguaggio degli organismi informativi, Glossario intelligence. https://www.sicurezzanazionale.gov.it/sisr.nsf/quaderni-di-intelligence/glossario-intelligence.html\. Accessed: 2019-08-08.

[28] B. Barnett and A. Crapo, "A semantic model for cyber security," 2011.

[29] L. Obrst, P. Chase, and R. Markeloff, "Developing an ontology of the cyber security domain," in STIDS, 2012.

[30] A. Oltramari, L. Cranor, R. Walls, and P. McDaniel, "Building an ontology of cyber security," CEUR Workshop Proceedings, vol. 1304, pp. 54–61, 01 2014.

[31] M. C. Parmelee, "Toward an ontology architecture for cyber-security standards," in *STIDS*, 2010.

[32] A. Singhal and D. Wijesekera, "Ontologies for modeling enterprise level security metrics," ACM International Conference Proceeding Series, 01 2010.

[33] A. Aviad, K. Wecel, and W. Abramowicz, "The semantic approach to cyber security. towards ontology based body of knowledge," vol. 2015, 01 2015, pp. 328–336.

[34] T. Takahashi and Y. Kadobayashi, "Reference ontology for cybersecurity operational information," The Computer Journal, vol. 58, no. 10, p. 2297–2312, 2015.

[35] E. Doynikova, A. Fedorchenko, and I. Kotenko, "Ontology of metrics for cyber security assessment," in *Proceedings of the 14th International Conference on Availability, Reliability and Security*, ser. ARES '19. New York, NY, USA: Association for Computing Machinery, 2019.

[36] A. Condamines, "Sémantique et corpus spécialisés : Constitution de Bases de Connaissances Terminologiques," Habilitation à diriger des recherches, Université Toulouse Le Mirail, Jun. 2003. [Online]. Available: https://halshs.archives-ouvertes.fr/tel-01321042

[37] G. Leech, The state of the art in corpus linguistics, K. Aijmer and B. Altenberg, Eds. London: Longman, 1991.

[38] G. Bernier-Colborne, "Defining a gold standard for the evaluation of term extractors," in Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), 2012, pp. 15–18.

[39] J. Pearson, Terms in Context. John Benjamins, Amsterdam, 1998.

[40] A. Rigouts Terryn, V. Hoste, and E. Lefever, "A gold standard for multilingual automatic term extraction from comparable corpora : term structure and translation equivalents," in Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), 2018, pp. 1803–1808. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2018/index.html

[41] A. Nazarenko, H. Zargayouna, O. Hamon, and J. Van Puymbrouck, "Evaluation des outils terminologiques : enjeux, difficultés et propositions," *Traitement Automatique des Langues*, vol. 50, no. 1 varia, pp. 257–281, 2009. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00516698

[42] E. Loginova Clouet, A. Gojun, H. Blancafort, M. Guegan, T. Gornostay, and U. Heid, "Reference Lists for the Evaluation of Term Extraction Tools," in *Terminology and Knowledge Engineering Conference (TKE)*, Madrid, Spain, Jun. 2012, pp. http://www.oeg–upm.net/tke2012/proceedings. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00816566

[43] E. Cardillo, A. Folino, R. Trunfio, and R. Guarasci, "Towards the reuse of standardized thesauri into ontologies," in Proceedings of the 5th International Conference on Ontology and Semantic Web Patterns - Volume 1302, ser. WOP'14, 2014, pp. 26–37.

[44] F. Giunchiglia, I. Zaihrayeu, and F. Farazi, "Converting classifications into owl ontologies."

[45] S.-J. Kang and J.-H. Lee, "Semi-automatic practical ontology construction by using a thesaurus, computational dictionaries, and large corpora," in *Proceedings of the Workshop on Human Language Technology and Knowledge Management - Volume 2001*, ser. HLTKM '01. USA: Association for Computational Linguistics, 2001. [Online]. Available: https://doi.org/10.3115/1118220.1118226

[46] Cybesecurity Osservatorio - Thesaurus. https://www.cybersecurityosservatorio.it/it/Services/thesaurus.jspt\. Accessed: 2019-08-08.

[47] J. E. Rowley, J. E. Rowley, and R. J. Hartley, Organizing knowledge: an introduction to managing access to information / Jennifer Rowley and Richard Hartley , 4th ed. Ashgate Aldershot, England ; Burlington, VT, 2008.

[48] C. Barrière, "Semi-automatic corpus construction from informative texts," in Text-Based Studies in honour of Ingrid Meyer, ser. Lexicography, Terminology and Translation, L. Bowkes, Ed. University of Ottawa Press, January 2006, ch. 5.

[49] G. Zagrebelsky, Il sistema costituzionale delle fonti del diritto, EGES, Ed. Turin: UTET, 1984.

[50] A. Condamines, "L'interprètation en sémantique de corpus : le cas de la construction de terminologies," Revue française de linguistique appliquée, vol. Vol. XII, no. 2007/1, pp. 39–52, 2007.

[51] F. Dell'Orletta, G. Venturi, A. Cimino, and S. Montemagni, "T2K: a system for automatically extracting and organizing knowledge from texts," in Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), N. C. C. Chair), K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Reykjavik, Iceland: European Language Resources Association (ELRA), may 2014.

[52] IATE European Union Terminology. https://iate.europa.eu/home\. Accessed: 2019-08-08.

[53] C. Lanza, "Italian domain-specific thesaurus as a means of semantic control for cybersecurity terminology," in The Twelfth International Conference on Advances in Semantic Processing (SEMAPRO 2018), U. o. A. S. G. P. L. U. o. H. A. F. Michael Spranger, Hochschule Mittweida, Ed., Athens, Greece, November 2018.

[54] M. van Assem, M. R. Menken, G. Schreiber, J. Wielemaker, and B. Wielinga, "A method for converting thesauri to rdf/owl," in *The Semantic Web – ISWC 2004*, S. A. McIlraith, D. Plexousakis, and F. van Harmelen, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 17–31.

[55] M. Nowroozi, M. Mirzabeigi, and H. Sotudeh, "The comparison of thesaurus and ontology: Case of asist web-based thesaurus and designed ontology," Library Hi Tech, vol. 36, 01 2018.

[56] J. L. D. Kless, L. Jansen and J. Wiebensohn, "A method for re-engineering a thesaurus into an ontology," in Proceedings of International Conference on Formal Ontology in Information Systems (FOIS 2012), 2012, pp. 133–146.

[57] J. Qin and S. Paling, "Converting a controlled vocabulary into an ontology: the case of gem," *Inf. Res.*, vol. 6, 2001.

[58] D. Adams, L. Jansen, and S. Milton, "A content-focused method for re-engineering thesauri into semantically adequate ontologies," *Semantic Web*, 09 2015.

[59] I. Rösiger, J. Bettinger, J. Schäfer, M. Dorna, and U. Heid, "Acquisition of semantic relations between terms: how far can we get with standard nlp tools?" in *Proceedings of Coling 2016: 5th International Workshop on Computational Terminology (CompuTerm)*, Osaka, Japan, 2016.

[60] A. Condamines, "Taking genre into account when analyzing conceptual relation patterns," *Corpora*, vol. 8, pp. 115–140, 2008. [Online]. Available: https://hal-univ-tlse2.archives-ouvertes.fr/hal-00606250

[61] A. Blumauer and T. Pellegrini, Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 9–25.