ISSN 1678-765X | doi 10.20396/rdbci.v20i00.8668437

RDBCI
Revista Digital de Biblioteconomia e Ciência da Informação
Digital Journal of Library and Information Science

Research Article

Domain Vocabulary Alignment using AML and LogMap

João Pedro Pereira Guimarães[1] Morgana C. Andrade[2] Ana Alice Baptista[3]

ABSTRACT

Introduction: In the context of the Semantic Web, interoperability among heterogeneous ontologies is a challenge due to several factors, among which semantic ambiguity and redundancy stand out. To overcome these challenges, systems and algorithms are adopted to align different ontologies. In this study, it is understood that controlled vocabularies are a particular form of ontology. Objective: to obtain a vocabulary resulting from the alignment and fusion of the Vocabularies Scientific Domains and Scientific Areas of the Foundation for Science and Technology, - FCT, European Science Vocabulary - EuroSciVoc and United Nations Educational, Scientific and Cultural Organization - UNESCO nomenclature for fields of Science and Technology, in the Computing Sciences domain, to be used in the IViSSEM project. Methodology: literature review on systems/algorithms for ontology alignment, using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses - PRISMA methodology; alignment of the three vocabularies; and validation of the resulting vocabulary by means of a Delphi study. Results: we proceeded to analyze the 25 ontology alignment systems and variants that participated in at least one track of the Ontology Alignment Evaluation Initiative competition between 2018 and 2019. From these systems, Agreement Maker Light and Log Map were selected to perform the alignment of the three vocabularies, making a cut to the area of Computer Science. Conclusion: The vocabulary was obtained from Agreement Maker Light for having presented a better performance. At the end, a vocabulary with 98 terms was obtained in the Computer Science domain to be adopted by the IViSSEM project. The alignment resulted from the vocabularies used by FCT (Portugal), with the one adopted by the European Union (EuroSciVoc) and another one from the domain of Science & Technology (UNESCO). This result is beneficial to other universities and projects, as well as to FCT itself.

KEYWORDS

Controlled vocabulary. Domain vocabulary. Ontology alignment. Ontology Alignment Evaluation Initiative. Semantic Web. Semantic compatibility.

Alinhamento de Vocabulário de domínio utilizando os sistemas AML e LogMap

RESUMO

Introdução: No contexto da Web Semântica, a interoperabilidade entre ontologias heterogêneas é um desafio devido a diversos fatores entre os quais se destacam a ambiguidade e a redundância semântica. Para superar tais desafios, adota-se sistemas e algoritmos para alinhamento de diferentes ontologias. Neste estudo, entende-se que vocabulários controlados são uma forma particular de ontologias. Objetivo: obter um vocabulário resultante do alinhamento e fusão dos vocabulários *Domínios Científicos e Áreas Científicas da Fundação para Ciência e Tecnologia, - FCT, European Science Vocabulary - EuroSciVoc e* Organização das Nações Unidas para a Educação, a Ciência e a Cultura - *UNESCO nomenclature for fields of Science and Technology*, no domínio Ciências da Computação, para ser usado no âmbito do projeto IViSSEM. Metodologia: revisão da literatura sobre sistemas/algoritmos para alinhamento de ontologias, utilizando a metodologia Preferred Reporting Items for

Author's correspondence

[1]Universidade do Minho, Minho, Portugal / e-mail: joaoppguimaraes@gmail.com

[2]Universidade Federal do Espírito Santo, Vitória, Brazil / e-mail: morganaandrade@hotmail.com

[3]Universidade do Minho, Minho, Portugal / e-mail: analice@dsi.uminho.pt

Systematic Reviews and Meta-Analyses - PRISMA; alinhamento dos três vocabulários; e validação do vocabulário resultante por meio do estudo Delphi. Resultados: procedeu-se à análise dos 25 sistemas de alinhamento de ontologias e variantes que participaram de pelo menos uma *track* da competição *Ontology Alignment Evaluation Iniciative* entre 2018 e 2019. Destes sistemas foram selecionados *Agreement Maker Light* e LogMap para realizar o alinhamento dos três vocabulários, fazendo um recorte para a área da Ciência da Computação. Conclusão: O vocabulário foi obtido a partir do *Agreement Maker Light* por ter apresentado uma melhor performance. Ao final foi obtido o vocabulário, com 98 termos, no domínio da Ciência da Computação a ser adotado pelo projeto IViSSEM. O alinhamento resultou dos vocabulários utilizados pela FCT (Portugal), com o adotado pela União Europeia (EuroSciVoc) e outro do domínio da Ciência&Tecnologia (UNESCO). Esse resultado é proveitoso para outras universidades e projetos, bem como para a própria FCT.

PALAVRAS-CHAVE
Alinhamento de ontologias. Vocabulários controlados. Vocabulários de domínio. *Ontology Alignment Evaluation Iniciative*. Web Semântica. Compatibilização semântica.

CRediT

- Recognitions: TheIViSSEM project - POCI-01-0145-FEDER-28284 and the Federal University of Espírito Santo.
- Funding: Not applicable.
- Conflicts of interest: Authors certify that they have no commercial or associative interest that represents a conflict of interest in relation to the manuscript.
- Ethical approval: Ethics Committee of the University of Minho and approved under no. CEICSH 083/2021.
- Availability of data and material: https://datarepositorium.uminho.pt/dataSep.xhtml?persistentId=doi:10.34622/datarepositorium/6IUDJB.
- Authors' contributions: Conceptualization, Data Curation- BAPTISTA, A. A.; Research, Writing, Validation - GUIMARÃES, J. P.P.; Supervision, Writing, Review - ANDRADE, M.

JITA: IL. Semantic web.

turnitin

Article submitted to the similarity system

Submitted: 16/02/2022 - **Accepted**: 01/07/2022 - **Published**: 20/07/2022

1 INTRODUCTION

According to Berners-Lee et al. (2001), the Semantic Web is an extension of the current Web, in which information has a well-defined meaning, allowing better cooperation between computers and people. In this context, ontologies emerge to implement the Semantic Web (JACOB, 2005). In other words, an ontology provides a vocabulary that describes a domain of interest, and the specifications of the meanings of the terms present in that vocabulary (EUZENAT; SHVAIKO, 2013).

However, different concepts are presented for the term ontology, either because of specific domains or because of the understanding of some authors. According to Studer et al. (1998), an ontology is defined as "a formal and explicit specification of a shared conceptualization", and in a machine-readable and shared format, since the conceptualization should be generalized and accepted by a group, not by just one person (GUARINO et al., 2009).

In this way, an ontology must be able to describe a structured and consented interpretation of a given domain to allow this knowledge about the domain to be shared by several agents, whether they are people or machines (BAÑOS-MORENO, 2017).

Due to the need for ontologies to present a machine-readable format, numerous languages have been developed over the years (KALIBATIENE; VASILECAS, 2011; MANIRAJ; SIVAKUMAR, 2010). For this work, the RDF language is highlighted, as well as its extensions, OWL and SKOS.

In the work of Euzenat and Shvaiko (2013) some of the many conceptual and data models that exist to describe knowledge are identified, namely, tags and folksonomies, directories, relational databases, and XML schemas. For this work only two types of controlled vocabularies were considered, the thesaurus and classification (nomenclature).

Faced with the fact that ontologies are conceptualized by different people for different purposes, their reuse ends up being limited. Different communities have different interests and habits and use different tools and knowledge, mostly with different levels of detail (EUZENAT; SHVAIKO, 2013). These differences result in distinct types of heterogeneity, which Euzenat and Shvaiko (2013) distinguish as follows:

a) syntactic heterogeneity - when two ontologies are not described in the same language, or when two ontologies are modeled by different knowledge representations (e.g., OWL and F-Logic).

b) terminological heterogeneity - this is due to the use of different variations of names to describe the same entity (e.g., different natural languages, synonyms, etc.).

c) semantic heterogeneity - occurs when the same domain is modeled differently, usually using different axioms (e.g., different coverage, granularity, perspective, etc.); and

d) semiotic heterogeneity - occurs when the same entity presents different interpretations to different people.

However, this heterogeneity, associated with ambiguity and semantic redundancy, makes interoperability between them a challenge (GRACIA; MENA, 2012). To overcome such challenges, it is important to have methods/algorithms capable of manipulating the different ontologies to mix, align and/or merge them.

To reduce problems related to heterogeneities, or to enable the reuse of ontologies, ontology alignment is adopted. According to Euzenat and Shvaiko (2013), ontology alignment aims to find correspondences between semantically related entities from different ontologies. As a result of ontology alignment, the so-called correspondences arise, which consist of a set of relationships between the entities of different ontologies that can be used in several tasks, such as merging ontologies, queries, data translation, or for navigating through the data connection on the Web (Semantic Web) (EUZENAT; SHVAIKO, 2013).

According to Euzenat and Shvaiko (2013), to obtain these correspondences ontology alignment techniques are applied that can be categorized based on two classifications: Type of input (kind of input), which considers the origin of the information and the type of input that is used by the ontology alignment techniques; and Granularity/Input interpretation classification (granularity/input interpretation classification), which is based on the granularity of the alignment technique. That is, this technique considers whether ontology entities are explored in isolation (element-level), or through their relationships with other entities (structure-level), and interpretation of the input that is performed considering only the structure itself (syntactic) or by some semantic formalism (semantics).

Space limitations do not allow us to discuss each of the alignment techniques in detail, but they differ in scope, focus and structure. The authors Euzenat and Shvaiko (2013) and Cheatham and Hitzler (2013) provide detailed descriptions on this topic. In this paper, the classification of alignments according to granularity, i.e., by Granularity/Input interpretation, is considered.

The increase in the number of systems capable of aligning ontologies in the early years motivated the creation of the Ontology Alignment Evaluation Initiative (OAEI), a coordinated international initiative aiming to openly evaluate and compare ontology alignment systems and algorithms, allowing anyone to be able to draw conclusions about the best strategies.

OAEI, in 2018 and 2019, made use of a set of ontologies called tracks divided into the following groups Schema Matching tracks, Instance Matching tracks, Instance and Schema Matching tracks, Complex Matching tracks, Interactive Matching tracks (ALGERGAWY, 2018, 2019).

Considering the benefits provided by ontology alignment, we chose to adopt it to produce a controlled vocabulary in the Computer Science domain that meets the needs of the IViSSEM Project.

1.1 Scope of the study: IViSSEM Project

About 2.5 million articles are published annually, in which almost 7,000 are published daily (IVISSEM, n.d.). Because of volume, finding articles of value to different researchers becomes an increasingly complicated task. To facilitate such a process, the IViSSEM project aims to develop and test a new altimetry (alternative metric coming from researchers' interactions) called the Social Scholarly Experience Metric, which will result from the application of machine learning techniques and different combinations of altimetry and researchers' profiles. Its application will reflect individual preferences in the process of finding a specific topic. Likewise, the current search result lists with huge amounts of items will be replaced by an innovative interface based on advanced visualization techniques.

Currently in its third year of execution, one of the goals of this project is to build a base domain vocabulary that will have controlled vocabulary terms related to each other and to tags inserted by the user. To this end, ontology alignment can be adopted to obtain this vocabulary. Topic addressed by this study.

The present study sought to develop a domain vocabulary, Computer Science, to incorporate terms from other controlled vocabularies related to each other through the implementation of an automatic or semi-automatic system capable of doing the alignment between them. Three vocabularies were considered, 2 multidisciplinary and 1 in the Science and Technology domain (EuroSciVoc, FCT Scientific Domains and Scientific Areas and UNESCO nomenclature for fields of science and technology) (PUBLICATIONS OFFICE OF

THE EU, 2021; PORTUGAL, 2012; SKOS, 2015).

In addition to this section, this paper has three more, they are: Systematic literature review, where the procedure adopted and the results that guided the choice of the systems to be used for the alignment are presented; Methodological framework, presents the methodological procedures employed for the steps concerning the implementation and validation of vocabulary alignment; Results, the results obtained with the implementation of the technique/algorithm and with the Delphi study to evaluate the alignment are presented; Conclusion, brings the main contributions obtained with the realization of the study and points out possibilities of future work; finally, the references of the articles used in this article are presented.

2 ONTOLOGY ALIGNMENT: A SYSTEMATIC REVIEW OF THE LITERATURE

To obtain the knowledge present in scientific articles about ontology alignment, a systematic review of the literature was developed. The systematic review was developed based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (MOHER et al., 2009). PRISMA consists of a checklist with 27 items and a flowchart with four steps (identification, selection, eligibility, and conclusion), and its main objective is to help authors improve the preparation of systematic reviews and meta-analyses (MOHER et al., 2009). First, a literature search was conducted to identify studies on ontology alignment. Studies that referenced algorithms and/or methods for ontology alignment were considered.

The studies were identified, in December 2020, in the Scopus database. The choice of Scopus was since the database has a 90% coverage of Natural Sciences and Engineering (AKSNES; SIVERTSEN, 2019; MONGEON; PAUL-HUS, 2015). To identify methods/algorithms for ontology alignment, the following search strategy was adopted: "Ontology Matching" OR "Ontology Alignment" OR "Ontology Mapping", in the field "Article Title, Abstract, Keywords". There was no limitation as to the type of document, temporal coverage, or area of application of the studies. Articles whose content was not available were not considered. Articles that cited or were cited by the articles initially identified were also considered. The execution of this search resulted in about 3,658 studies, of which only the 2000 most relevant were considered according to the database relevance index.

Subsequently, an analysis of the articles was performed based on the titles and abstracts. Some of the authors of the studies were also contacted to obtain additional information.

The previous analysis of the first 100 articles pointed to the relevance of the Ontology Alignment Evaluation Initiative (OAEI) competition in this theme. The OAEI is a coordinated international initiative that annually organizes the evaluation of ontology alignment systems in a controlled manner across a large dataset (ALGERGAWY, 2018, 2019). Given this finding, it was chosen to conduct the systematic literature review from the articles referring to the systems' participations in OAEI in the period 2018 and 2019. Thus, only articles referring to the competition and the participation of the systems in the two years were considered, as well as articles external to the competition describing the participating systems identified in the Scopus database. From this selection, 34 articles resulted.

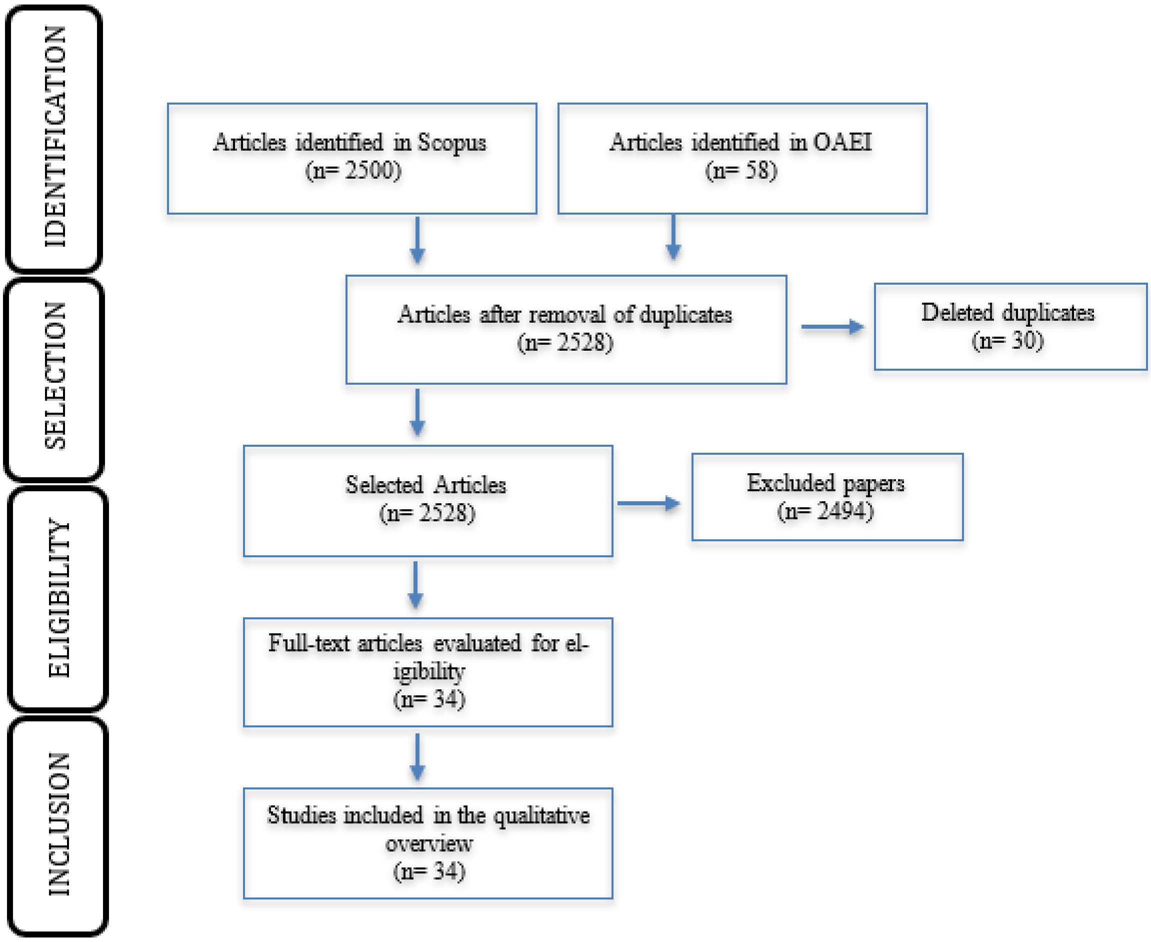
From these articles, information was extracted regarding the strategies and techniques adopted by each approach for ontology alignment, as well as the performance of each system in the different tracks of the annual OAEI competition in the years 2018 and 2019.

It is worth noting that, despite being considered studies whose performances are evaluated in the same controlled environment (OAEI), the risk of bias decreases, but it still exists, since each system presents different performances in different tracks, which can influence the analysis of the quality of the systems and identification of the "best" system.

The option to use the articles presented in the OAEI was ratified after the bibliometric analysis that included the 2,000 items retrieved in Scopus. The bibliometric analysis indicated the conference as the most relevant event on the topic, as well as the identification of the most prominent authors. The results of the bibliometric analysis were presented in another paper (submitted for publication).

Figure 1 illustrates the selection process for studies.

Figure 1. PRISMA Flowchart



Source: Adapted from Moher *et al*. (2009)

The OAEI in 2018 and 2019 made use of a set of tracks divided into the following groups: (a) Schema Matching tracks that aim to evaluate the alignment of ontology classes and properties; (b) Instance Matching tracks that aim to evaluate the alignment of ontology instances; (c) Instance and Schema Matching tracks that aim to evaluate both the alignment of classes and properties and ontology instances; (d) Interactive Matching tracks that aim to evaluate the performance of systems by simulating user interaction; and (c) Complex Matching tracks aim to evaluate the achievement of complex matches (n: m) by the systems (ALGERGAWY, 2018, 2019). Such division is observed in Frame 1.

Frame 1. OAEI *tracks*

Group	*Tracks*
Schema Matching tracks	*Anatomy track*
	Biodiversity and Ecology track
	Conference track
	Disease and Phenotype track
	Large Biomedical Ontologies track
	Multifarm track
Instance Matching tracks	*IIMB track* (only in 2018)
	Link Discovery track
	SPIMBENCH track
Instance and Schema matching tracks	*Knowledge Graph track*
Interactive Matching tracks	*Interactive Matching track*
Complex Matching tracks	*Complex Matching track*

Source: Algerway (2018, 2019)

As mentioned before, the tracks were analyzed considering the alignment techniques used by each of the systems, as well as their performances in OAEI. Regarding the alignment techniques used by them, these were analyzed based on the Granularity/Input interpretation classification. This analysis was the result of personal interpretation of the systems' articles, as well as the exchange of information, via e-mail, with the respective authors.

Regarding the performance of the systems in OAEI, this was evaluated based on the precision, recall and F-measure factors of each of the systems in the various tracks (ALGERWAY, 2018, 2019). It is worth noting that the articles published on each of the systems, although in the form of OAEI papers, were not reviewed by their OAEI peers, which may result in inconsistencies.

After analyzing the 34 articles, we proceeded to categorize the different methods/algorithms used by each of the systems. Regarding the table, it describes the different types of existing techniques according to Euzenat and Shvaiko (2013), and the different systems explore different evaluation methods. Thus, there are systems that, for example, exploit techniques characterized as Structure-level and others exploit other techniques characterized as Element-level. For each system, each of the techniques used were identified and categorized according to the respective methods described by Euzenat and Shvaiko (2013). In Frame 2, the methods used by each of the systems are categorized.

Frame 2. Categorization of alignment methods

Systems	*Element-level*		*Structure-level*	
	Syntactic	*Semantic*	*Syntactic*	*Syntactic*
AGM (LÜTKE, 2019)	*String-based* through *Levenshtein distance, Skip-gram neural model, cosine similarity* e *Euclidean distance*	Unaddressed	*AGM* (LÜTKE, 2019)	*String-based* through *Levenshtein distance, Skip-gram neural model, cosine similarity* e

Systems	Element-level		Structure-level	
	Syntactic	*Semantic*	*Syntactic*	*Syntactic*
				Euclidean distance
ALIN (SILVA, 2018, 2019)	*Language-based* through *tokenization* e *lemmatization, String-based* through *Jaccard similarity, Jaro-Wrinkler similarity* e *n-gram, WordNet*	*FMA ontology*	*ALIN* (SILVA, 2018, 2019)	*Language-based* through *tokenization* e *lemmatization, String-based* through *Jaccard similarity, Jaro-Wrinkler similarity* e *n-gram, WordNet*
ALOD2Vec (PORTISCH, 2018)	*Language-based, String-based* through de *neural language model, cosine similarity*	*WebIsALOD data set*	*ALOD2Vec* (PORTISCH, 2018)	*Language-based, String-based* through de *neural language model, cosine similarity*
AML & *AMLC* (FARIA, 2018, 2019; FARIA et al., 2013)	*Language-based, String-based* through *Jaccard similarity*	*Uber Anatomy Ontology (Uberon), Human Disease Ontology (DOID)* e *the Medical Subject Headings (MeSH)*	*AML* & *AMLC* (FARIA, 2018, 2019; FARIA et al., 2013)	*Language-based, String-based* through *Jaccard similarity*
AROA (ZHOU, 2019)	Unaddressed	Unaddressed	*AROA* (ZHOU, 2019)	Unaddressed
CANARD (THIÉBLIN, 2018, 2019)	*String-based* through de *Label similarity*	Unaddressed	*CANARD* (THIÉBLIN, 2018, 2019)	*String-based* through de *Label similarity*
DOME (HERTLING, 2018, 2019)	*String-based, Language-based, cosine similarity*	Unaddressed	*DOME* (HERTLING, 2018, 2019)	*String-based, Language-based, cosine similarity*
EVOCROS (DESTRO, 2018, 2019)	*Language-based, String-based* through de *Levenshtein* e *Jaro similarities, WordNet, BabelNet*	Unaddressed	*EVOCROS* (DESTRO, 2018, 2019)	*Language-based, String-based* through de *Levenshtein* e *Jaro similarities, WordNet, BabelNet*
FCAMapX & *FCAMap-KG* (CHANG, 2019; CHEN, 2018)	*Language-based* through *Token-based formal context*	Unaddressed	*FCAMapX* & *FCAMap-KG* (CHANG, 2019; CHEN, 2018)	*Language-based* through *Token-based formal context*

Systems	*Element-level*		*Structure-level*	
	Syntactic	*Semantic*	*Syntactic*	*Syntactic*
FTRLIM (WANG, 2019)	*FTRL model, TF-IDF, edit distance similarity* e *Jaccard similarity*	Unaddressed	*FTRLIM* (WANG, 2019)	*FTRL model, TF-IDF, edit distance similarity* e *Jaccard similarity*
Holontology (ROUSSILLE, 2018)	*Exact match, Levenshtein, Jaccard* e *Lin similarities, constraint-based*	Unaddressed	*Holontology* (ROUSSILLE, 2018)	*Exact match, Levenshtein, Jaccard* e *Lin similarities, constraint-based*
KEPLER (KACHROUDI, 2018)	*WordNet, Microsoft Bing Translator*	Unaddressed	*KEPLER* (KACHROUDI, 2018)	*WordNet, Microsoft Bing Translator*
Lily (TANG, 2018; WU, 2019)	*String-based*	*BioPortal, UMLS Metathesaurus*	*Lily* (TANG, 2018; WU, 2019)	*String-based*
LogMap, LogMapBio, LogMapIIMB, LogMapKG & *LogMapLt* (JIMÉNEZ-RUIZ, 2018, 2019; JIMÉNEZ-RUIZ; CUENCA GRAU, 2011)	*String-based*	*BioPortal, UMLS Metathesaurus*	*LogMap, LogMapBio, LogMapIIMB, LogMapKG* & *LogMapLt* (JIMÉNEZ-RUIZ, 2018, 2019; JIMÉNEZ-RUIZ; CUENCA GRAU, 2011)	*String-based*
ONTMAP1 (GHERBI, 2019)	*n-gram, lemmatization, cardinality constraints, WordNet*	Unaddressed	*ONTMAP1* (GHERBI, 2019)	*n-gram, lemmatization, cardinality constraints, WordNet*
POMap++ (LAADHAR, 2018, 2019)	*Language-based between natural language processes*	Unaddressed	*POMap++* (LAADHAR, 2018, 2019)	*Language-based between natural language processes*
SANOM (MOHAMMADI, 2018, 2019)	*Tokenization, stop word removal, stemming (lemmatization technique), Jaro-Winkler similarity, WordNet*	Unaddressed	*SANOM* (MOHAMMADI, 2018, 2019)	*Tokenization, stop word removal, stemming (lemmatization technique), Jaro-Winkler similarity, WordNet*
XMAP (DJEDDI, 2018)	*String and linguistic similarities*	*UMLS Metathesaurus*	*XMAP* (DJEDDI, 2018)	*String and linguistic similarities*

Systems	Element-level		Structure-level	
	Syntactic	*Semantic*	*Syntactic*	*Syntactic*
Wik;tionary (PORTISCH, 2019)	*String-based*	*Wiktionary*	*Wik;tionary* (PORTISCH, 2019)	*String-based*

Source: Own authorship

When analyzing the results regarding the performances of the systems presented by the OAEI (ALGERWAY, 2018, 2019) two systems were selected for this study for implementation, the AML and LogMap systems (Cf. GUIMARÃES, 2022, p. 130-164). This selection results from the quality of the performance of the two systems in the different tracks, since these two systems presented the best performances in most of the tracks. From a total of 12 tracks, AML and LogMap, along with their variants, stood out in 10 and 8 of them, respectively.

Another factor that had an impact on this decision was also the number of citations of the articles about the systems obtained by bibliometric analysis, in which their authors are also highly cited in this theme.

3 METHODOLOGICAL PROCEDURES

To achieve the proposed objective, that is, to create a vocabulary for the IViSSEM project, a set of stages was defined, as well as the methodological procedures used in each one of them.

Thus, three steps were defined: selection of controlled vocabularies; implementation of the systems for aligning the controlled vocabularies; and evaluation of the alignments produced in the implementation.

3.1 Selection of Controlled Vocabularies

To develop the vocabulary for the IViSSEM project, three controlled vocabularies representing the main areas of knowledge were considered: EuroSciVoc, UNESCO Nomenclature for fields of Science and Technology, and Scientific Domains and Scientific Areas of the Foundation for Science and Technology (FCT) of the Portuguese Ministry of Education and Science. The selection of the EuroSciVoc and UNESCO Nomenclature for Fields of Science and Technology vocabularies was since they are associated with relevant international institutions, while the selection of the FCT vocabulary was since it is a public agency of Portugal in the context of Science and Technology, the country where the project is being developed. Table 3 details these vocabularies.

Frame 3. Controlled vocabularies used

Vocabulary	Provider	Number of terms	Languages
EuroSciVoc	*Publications Office of the EU*	991	German, English, Spanish, French, Italian, Polish
UNESCO nomenclature for fields of science and technology	*UNESCO*	2504	German, English, Spanish, French, Italian, Polish
Domínios Científicos e Áreas Científicas	*FCT*	178	English, Portuguese

Source: Own authorship

Due to the number of terms present in two vocabularies (UNESCO and EuroSciVoc), in order not to compromise the validation of the final alignment and the quality of the resulting vocabulary, it was necessary to delimit the vocabularies, considering only the area of Computer Science.

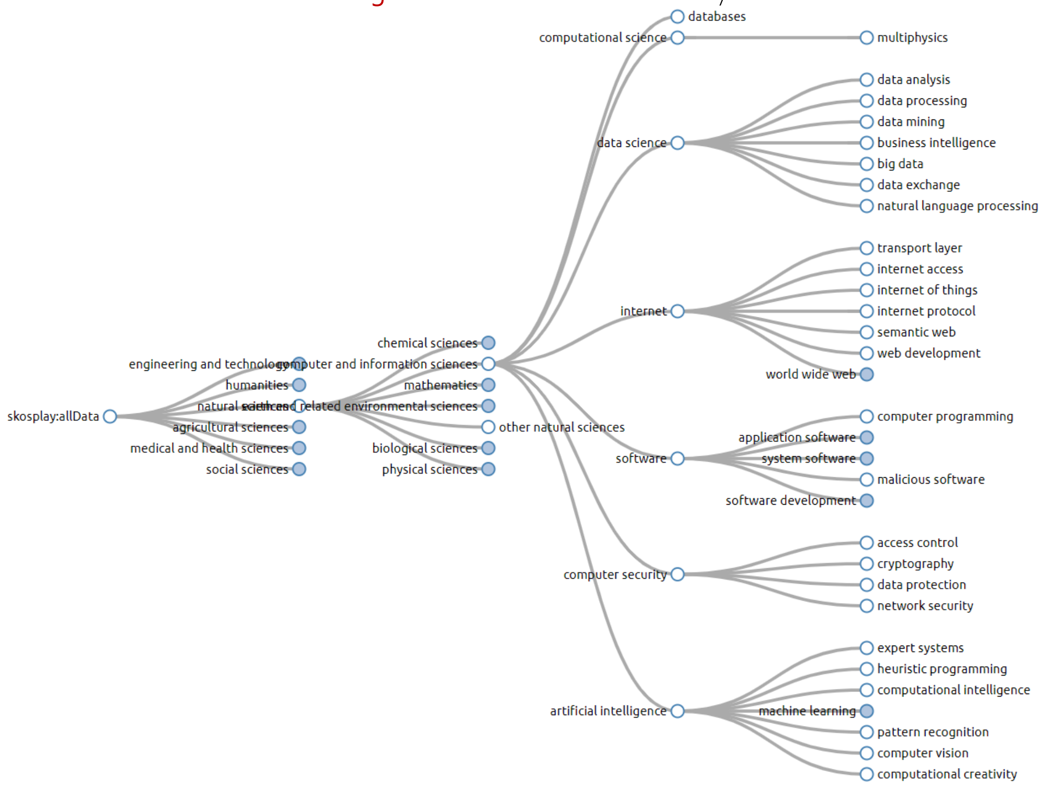
3.1.1 EuroSciVoc

The European Science Vocabulary (EuroSciVoc) is a multilingual taxonomy based on the 2015 Frascati manual of the Organization for Economic Cooperation and Development (OECD) that represents all major fields of Science. It has been extended with science fields extracted from the content of the Community Research and Development Information Service (CORDIS) (ENCYCLOPEDIA, 2014) and organized through a semi-automatic process based on Natural Language Processing (NLP) techniques.

Since this vocabulary is described in SKOS-XL (SKOS extension that allows labels to be used as resources), there was a need to manually convert it into SKOS, thus allowing alignment with other vocabularies. This need arose from the fact that the LogMap system does not produce alignments between SKOS-XL and SKOS files.

After delimiting the vocabulary, in order to highlight the area of Computer Science, the branch resulting from the term "computer and information science", presented in Figure 2, was taken into account. The complete vocabulary can be found in Publications Office of the EU (2021).

Figure 2. EuroSciVoc Vocabulary



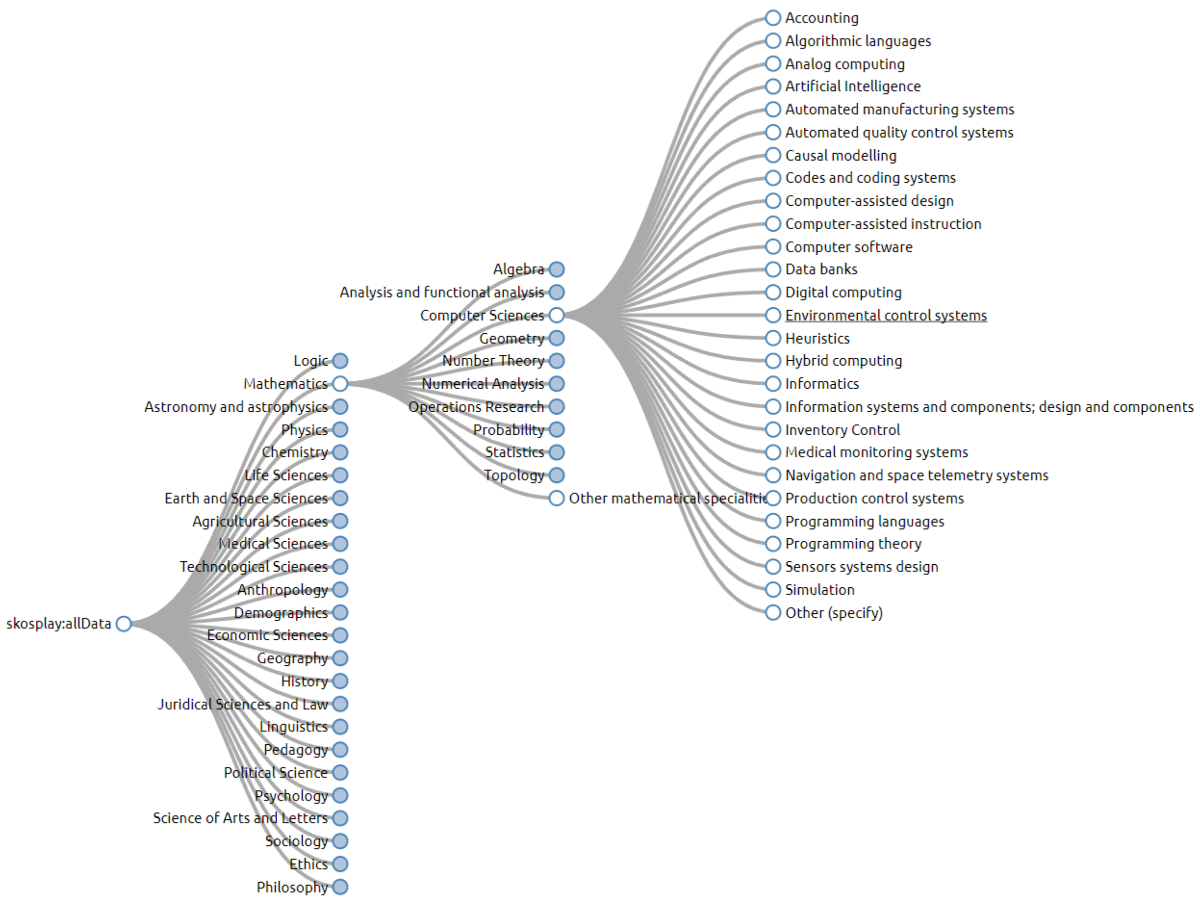
Source: adapted Publications Office of the EU (2021)

3.1.2 UNESCO Nomenclature for Fields of Science and Technology

Proposed in 1973 and 1974 by UNESCO's Science Policy and Statistics Division for Science and Technology and adopted by the Scientific Advisory Committee, the UNESCO Nomenclature for Fields of Science and Technology is widely used in the knowledge management of research projects and dissertations. It is divided into three hierarchical levels: areas, disciplines and subdisciplines (SKOS, 2015; PASTOR-SÁNCHEZ; RODRIGUEZ-MUÑOZ; LOPEZ-CARREÑO, 2013).

A cut was made from the resulting branch of the term "Computer Science", presented in Figure 3. The complete vocabulary can be observed in SKOS... (2015).

Figure 3. UNESCO Vocabulary Nomenclature for Science and Technology



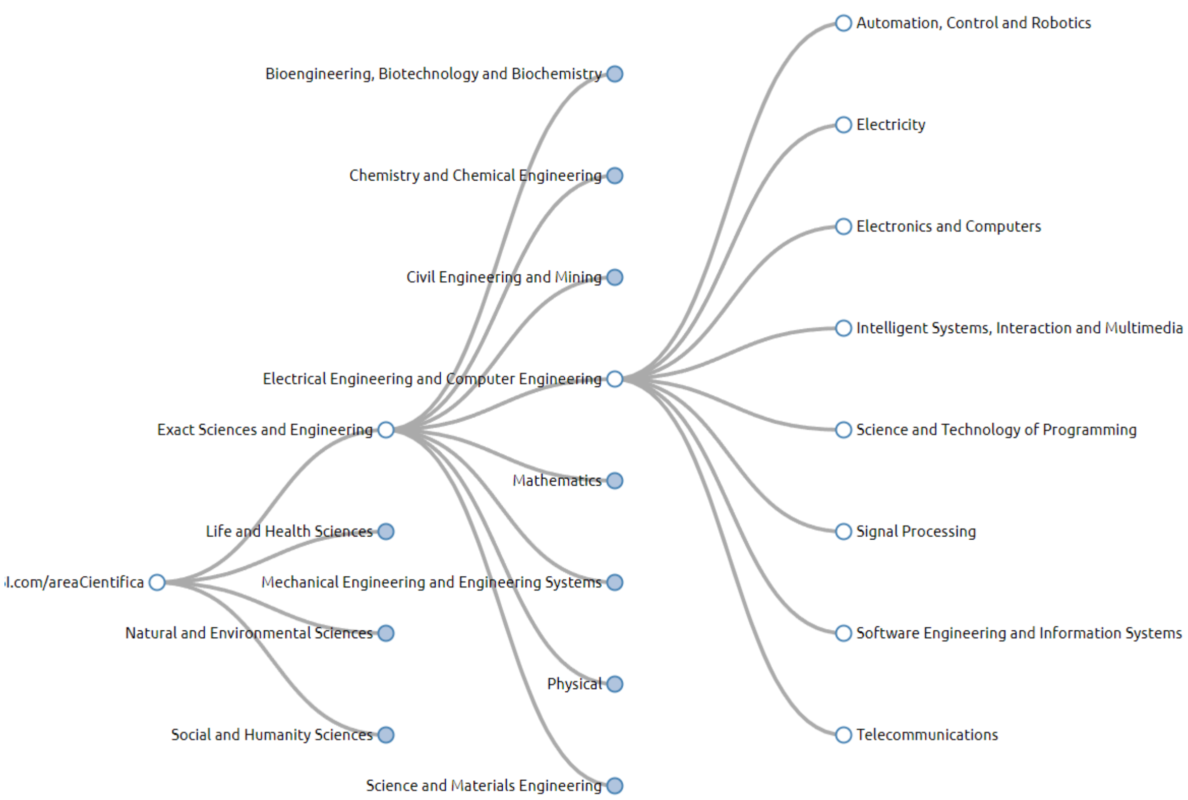
Source: adapted of SKOS... (2015)

3.1.3 FCT Scientific Domains and Scientific Areas

The vocabulary Scientific Domains and Scientific Areas used by FCT is a classification system organized by scientific areas and domains, where each domain presents a set of areas and subareas (MENDES, 2016). This vocabulary is available in PDF format; therefore, an RDF encoding was used, developed by a group of students within the Semantic Web course unit of the Information Systems Course at the University of Minho, Guimarães (Portugal), and made available at DSI Wiki (n.d.).

For vocabulary analysis, the branch resulting from the term "Electrical Engineering and Computer Engineering" was considered, presented in Figure 4. The complete vocabular is available at Fundação para a Ciência e a Tecnologia (2012).

Figure 4. Vocabulary Scientific Domains and Scientific Areas



Source: adapted from Foundation for Science and Technology (2012)

With the adoption of the previously mentioned cut-off criteria, it was possible to proceed to the alignment of the vocabularies, described in the next section.

3.2 Implementing the systems for aligning controlled vocabularies

The controlled vocabulary for the IViSSEM project resulted from the alignment of the three vocabularies previously described, considering only the terms referring to the area of Computer Science.

Thus, through the application of AML and the LogMap web interface, the vocabularies were aligned one by one (vocabulary A with vocabulary B and vice versa). These alignments were in OAEI Alignment Format. In this format, each match is composed of the terms identified as matching (coming from the source and target vocabularies) and the confidence value. This value refers to the confidence of the alignment system in the correspondence between the terms, ranging from 0 to 1, where 1 is the maximum confidence value. For this work, only alignments with a confidence value greater than 0.6 were considered. It is worth noting that, as far as confidence values are concerned, they vary from system to system, resulting from the different algorithms that each one implements.

3.3 Evaluation of the produced alignments

The number of participants and rounds can be defined by the profile of the respondents and the objective of the study (FINK et al., 1984). The appropriate selection of panelists or participants is critical to promote data reliability and validity. For this Delphi study, the panel included 10 representatives from the two highlighted areas and of a professional profile consistent with the use of the terminologies (STEWART et al., 2017).

As for the number of rounds, this was defined after the first application of the questionnaire was obtained, since some authors advocate that this quantity may vary according to the agreement or consensus index obtained among the participants (DELBECQ et al., 1986). The consensus rates vary in the literature between 70 and 80% (BARRIOS et al., 2021; MURRY; HAMMONS, 1995), however for this study the consensus rate adopted was 70% by adding the partial and total agreement.

First, a pilot questionnaire was developed and sent to five collaborators, to evaluate the clarity, size, and objectivity of the questionnaire. The employees were asked to analyze the questionnaire and to present suggestions for improvement.

Based on the suggestions of three respondents, the final version of the questionnaire was then developed and sent to ten specialists in the Computer Science fields, and the respondents' suggestions were considered. This sample consisted of eight individuals with PhDs and two with master's degrees. Five were from Portugal and five from Brazil. After the questionnaire was finalized, invitations were sent to the respondents, and all accepted the invitation. The participants of this phase were not included in the questionnaire applied in the first round. The research was submitted to the Ethics Committee of the University of Minho and approved under nº CEICSH 083/2021.

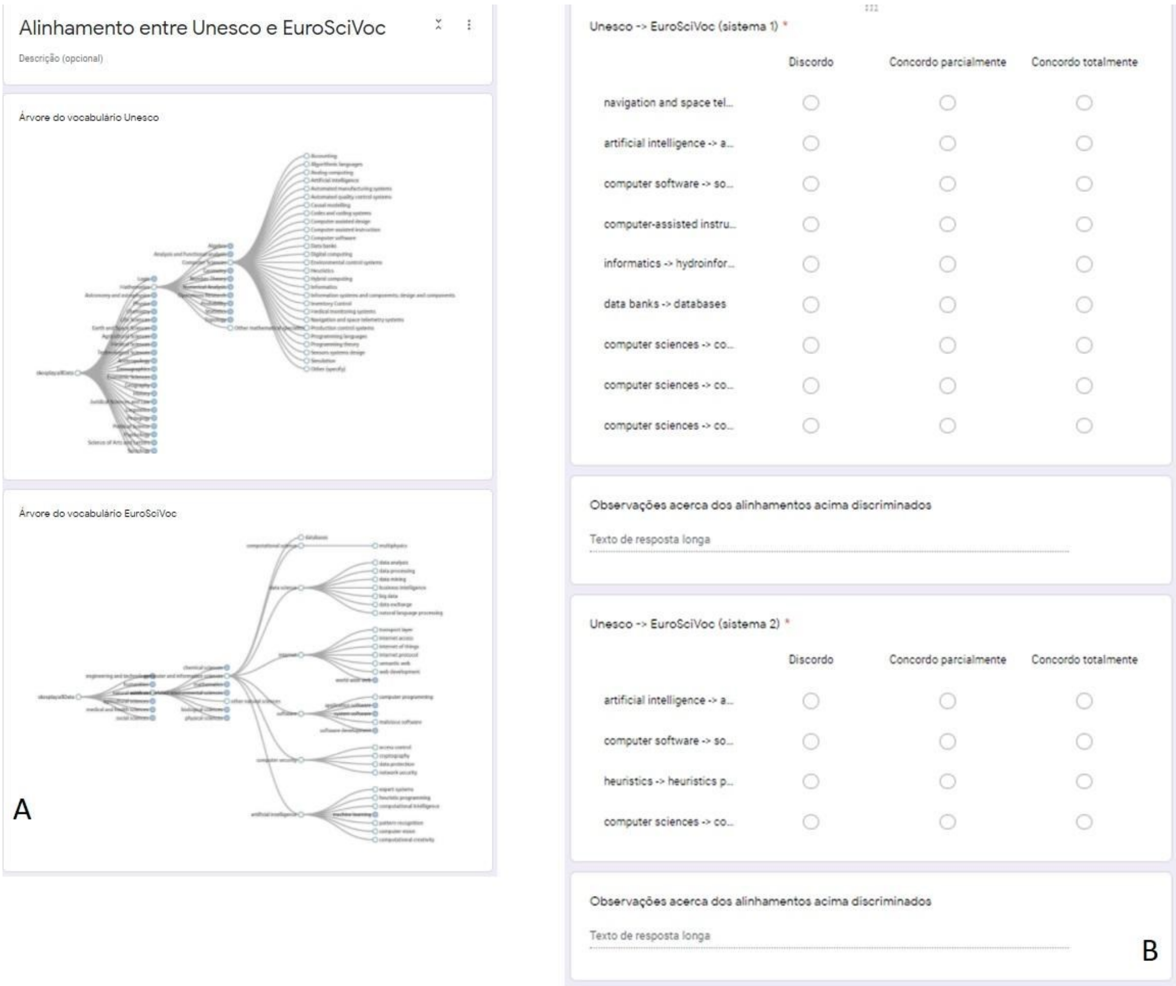
The questionnaire adopted for the study was prepared using the Google forms tool, structured in two parts. In the first part, an informed consent form was presented, which when confirmed, the respondents were asked to answer a set of demographic questions (education, country of residence, and professional activity), which allowed characterizing the individuals who answered the questionnaire.

In the second part, the alignments between the different vocabularies generated by the systems were made available, as well as the trees of the respective vocabularies, to facilitate the individuals' decision-making regarding the alignments. SKOS Play (SKOS Play! n.d.) was used to obtain the trees of the respective vocabularies.

Regarding the generated alignments, for each correspondence a Likert scale was used, with three response options (partially agree, strongly agree, and disagree). The choice of the number of response options falls on the analyses of the articles (DALMORO; VIEIRA, 2013; RODRIGUEZ, 2005). So that the answers provided would not be influenced by preferences of the experts regarding the systems used, the name of the system responsible for each of the alignments was omitted. Thus, the LogMap and AML systems were identified as system 1 and 2, respectively. For each alignment, a field was also made available for the experts to make observations.

A Figura 5 mostra a estrutura de cada um dos alinhamentos presentes na segunda parte do questionário, com o exemplo do alinhamento entre a Nomenclatura da UNESCO para Campos de Ciência e Tecnologia e o vocabulário EuroSciVoc.

Figure 5. Structure of each alignment present in the second part of the questionnaire



Source: Own authorship

This section presents the main results obtained in the vocabulary alignment and evaluation process.

4 RESULTS

This section presents the main results obtained in the vocabulary alignment and evaluation process.

4.1 Vocabulary alignment

A total of 12 alignments (6 alignments per application from each of the systems) were produced and made available at RepositóriUM, https://doi.org/10.34622/datarepositorium/6IUDJB. All the analyses and their results can be seen in Guimarães (2022, p. 44-60).

After analyzing the produced alignments, it was found that the AML system generated matches in all alignments except for the alignments between EuroSciVoc-FCT, and Unesco-FCT. It was also found that the highest confidence value assigned to an alignment was 0.92 and the lowest was 0.61. Matches between equal terms had a higher confidence value.

Regarding the alignments produced by the LogMap system, it was found that it produced matches in all alignments except for the Unesco-FCT alignment. The fact that neither system produced matches in this alignment may indicate that, in fact, there is no match. It was also found that the highest confidence value assigned to an alignment was 1.0 and the lowest was 0.77. Compared to the alignments produced by the AML system, it was also observed that the LogMap system produced a higher number of matches in all alignments. One of the causes may be the fact that the LogMap system presents correspondences with a higher confidence value. Thus, it is possible that some of these correspondences could also have been suggested by the AML system but were discarded because they presented a confidence value lower than 0.6 (established as the minimum confidence value). It is still possible to see that many of these matches are due to the simple fact that some terms share some words among themselves, but considering that they present a totally different context, it makes them invalid. As an example, we have the alignment of two terms (EuroSciVoc and FCT) performed by the LogMap system that did not obtain consensus among the experts: computer and information sciences -> Science of communication and information (narrower of institutions, values, beliefs, and behavior).

4.2 Vocabulary alignment evaluation

The evaluation of the alignment was performed by 10 professionals, mostly faculty, followed by researchers and librarians.

The answers regarding the correspondences generated by each of the systems were given based on the three available agreement options: disagree, partially agree, and totally agree. For each of the correspondences, in case the number of disagreements was higher than the number of agreements (total or partial), the correspondence would be discarded, otherwise the correspondence would be validated and accepted. In case there were more concordances, and the number of partial concordances was higher or equal to the number of total concordances, the correspondence was further analyzed, and later, depending on the verdict of the analysis, it was accepted or not. In addition to the matches, the experts were also able to make suggestions about the alignments and their vocabularies.

Finally, for each alignment the precision of the systems was also calculated, with the purpose of analyzing the performance of both and identifying which one was better. Given the inexistence of a reference alignment, the recall and F-measure of the systems were calculated using silver standard reference alignments, that is, a reference alignment that is not necessarily correct and complete, based only on the matches accepted by the experts in each alignment. For each alignment, the following results were obtained (Tables 1-6).

Table 1. Alignment between the EuroSciVoc and Unesco Nomenclature for Fields of Science and Technology vocabularies

System	Suggested Correspondences	Correspondences analyzed in detail	Accepted Correspondences	*Precision*	*Recall*	*F-measure*
AML	5	2	5	1.0	0.83	0.91
LogMap	19	2	5	0.26	0.83	0.40

Source: Research data

Table 2. Alignment between the Unesco Nomenclature for Fields of Science and Technology and EuroSciVoc vocabularies

System	Suggested Correspondences	Correspondences analyzed in detail	Accepted Correspondences	*Precision*	*Recall*	*F-measure*
AML	4	2	4	1.0	0.67	0.80
LogMap	9	3	4	0.44	0.67	0.53

Source: Research data

Table 3. Alignment between EuroSciVoc vocabularies and Scientific Domains and Scientific Areas of the Foundation for Science and Technology

Sistema	Suggested Correspondences	Correspondences analyzed in detail	Accepted Correspondences	*Precision*	*Recall*	*F-measure*
AML	0	0	0	-	-	-
LogMap	2	0	0	0.0	-	-

Source: Research data

Table 4. Alignment between the Scientific Domains and Scientific Areas vocabularies of the Foundation for Science and Technology and EuroSciVoc

Sistema	Suggested Correspondences	Correspondences analyzed in detail	Accepted Correspondences	*Precision*	*Recall*	*F-measure*
AML	2	1	2	1.0	0.5	0.67
LogMap	7	6	4	0.57	1.0	0.73

Source: Research data

Table 5. Alignment between the Unesco Nomenclature for Fields of Science and Technology and the Scientific Domains and Scientific Areas vocabularies of the Foundation for Science and Technology

Sistema	Suggested Correspondences	Correspondences analyzed in detail	Accepted Correspondences	*Precision*	*Recall*	*F-measure*
AML	0	0	0	-	-	-
LogMap	0	0	0	-	-	-

Source: Research data

Table 6. Alignment between the Scientific *Domains and Scientific Areas vocabularies of the Foundation for Science and Technology and UNESCO nomenclature for fields of science and technology*

Sistema	Suggested Correspondences	Correspondences analyzed in detail	Accepted Correspondences	*Precision*	*Recall*	*F-measure*
AML	1	0	1	1.0	0.33	0.50
LogMap	3	2	3	1.0	1.0	1.0

Source: Research data

Overall, from the analysis of the different alignments, it can be observed that the LogMap system, compared to the AML system, was able to produce more matches in all alignments. However, it was observed that part of the correspondences made by the LogMap system were often discarded by the experts, thus resulting most of the time in poor precision across the different alignments. In contrast, it was observed that the AML system produced fewer matches, but all of them were accepted by the experts, thus resulting in a better performance (in terms of precision) in all the alignments in which it produced matches, obtaining in all of them a precision of 1.0.

In terms of F-measure, both systems presented results with a similar range of values, and there were alignments where the AML system stood out and others where the LogMap system stood out.

Thus, it was concluded that the AML system is the most consistent system, and in turn more suitable to be adopted in the development of the controlled vocabulary for the IViSSEM project.

In view of the result obtained with the application of the questionnaire (first Delphi round), in which a 70% consensus index was identified in all correspondences, it was concluded that there would be no need for a new round (DELBECQ et al., 1986; MURRY.; HAMMONS, 1995).

4.3 Final product - Vocabulary for the IViSSEM project

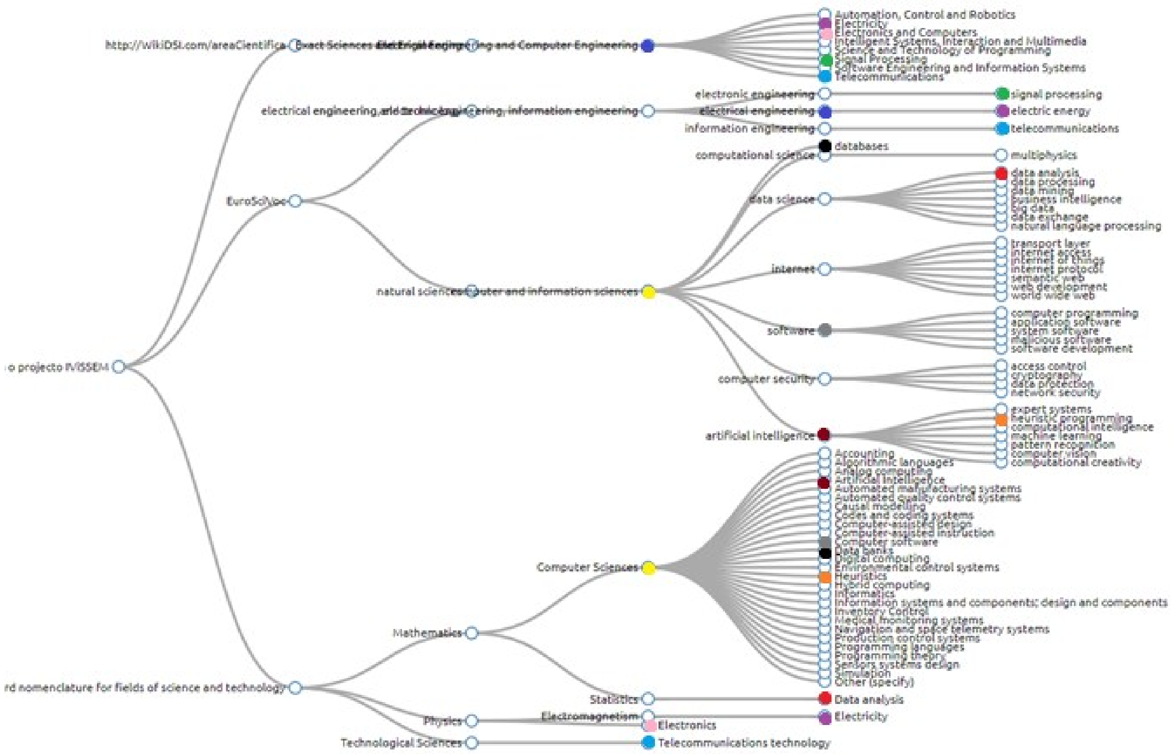
Having identified the AML system as the most viable system for this project, the controlled vocabulary for the IViSSEM project was created, which, as previously mentioned, is the result of the alignment and merging of three controlled vocabularies related to scientific areas (EuroSciVoc, FCT Scientific Domains and Scientific Areas, and UNESCO Nomenclature for Fields of Science and Technology).

For this purpose, the alignments provided by the AML alignment system were used. Since the number of alignments was small, the alignments provided by the LogMap system that were not found by the AML system and were accepted by the experts were also considered. If the number of matches was higher, this would not be possible, so the alignments resulting from the AML system would be considered.

That said, using SKOS, the controlled vocabulary for the IViSSEM project was then created. As for the alignments, these were transposed to the vocabulary through the skos: related property, thus allowing the terms from the different vocabularies to be linked.

The code for the generated vocabulary is available at RepositóriUM, https://doi.org/10.34622/datarepositorium/T49TKX. In Figure 6 it is possible to see the tree of the domain vocabulary produced, where each color refers to some correspondence.

Figure 6. Produced domain vocabulary tree[1]



Source: Own authorship

5 CONCLUSIONS AND FUTURE WORK

The importance of ontologies as a means of implementing the Semantic Web is a fact. Likewise, there is a need for techniques capable of combating heterogeneity and promoting interoperability between different ontologies. That said, ontology alignment through automatic or semi-automatic techniques plays a major role in this topic.

As a result of the growing number of systems capable of aligning ontologies, the OAEI emerges as a coordinated international initiative created to openly evaluate and compare ontology alignment systems and algorithms, thus allowing anyone to draw conclusions about the best strategies. Very satisfactory results were registered in all OAEI tracks, except for the Complex Matching track, which, due to its complexity and need for extreme system changes, shows fewer positive results. Of all the systems that participated in the OAEI in 2018 and/or 2019, AML and LogMap stand out, as they are among the few systems that participated in all the tracks and obtained very satisfactory results.

Notwithstanding the fact that the systems were developed to align OWL files, it was found that both the AML and LogMap systems were able to produce satisfactory alignments between SKOS files, without the need for conversions.

Despite the good results that ontology alignment techniques have shown at OAEI, they can still be improved. Through OAEI's analyses over the years, it is possible to see that the alignment techniques have shown little improvement in terms of quality. The main causes for

[1]Nota: Azul claro - Telecommunication; Azul escuro - Electrical engineering; Roxo - Electricity; Rosa - Electronics; Verde - Signal Processing; Vermelho - Data analysis; Preto - Databases; Cinza - Software; Laranja - Heuristics; Amarelo - Computer Sciences; Bordô - artificial intelligence.

the poor development of alignment techniques are either the need for new technology, or the need for the developers of the techniques to focus on improving their techniques, trying to achieve better results on old tracks rather than focusing on achieving good results on new tracks.

Furthermore, when it comes to ontology alignment, the need for domain knowledge is still a reality. This was seen in this work when it was necessary to use a panel of experts in Computer and Information Sciences to evaluate the generated matches.

The scarcity of a manually curated alignment between the three vocabularies used, capable of serving as a reference alignment, is a limitation in the evaluation of the alignment systems, since only a "silver standard" reference alignment was used that was based on the correspondences generated by the systems and validated by the experts. Such a creation can and should be seen as future work.

Another limitation present in this work is since a larger number of experts were not available, which could have promoted the execution of the Delphi study with a larger number of rounds.

With this study, it is possible to conclude that, currently, there are semi-automatic techniques that feasibly allow reusing existing knowledge on the web through the alignment of ontologies and controlled vocabularies. Such techniques make use of a set of different algorithms, increasingly complex, that allow dealing with ontologies in different ways (considering semantics, structure, etc.).

For the IViSSEM project and eventually for other projects, this work will allow the use of three integrated vocabularies instead of just one, making it possible to expand queries and perform inferences that were previously forbidden. This may be the first step to integrate all three vocabularies and eventually include others. This work can contribute to future works by identifying the most suitable system to integrate them and, additionally, it allowed the development of a proof of concept, which is the final vocabulary. It is worth mentioning the resulting alignment between the vocabulary used at a national level by FCT (Portugal) with one employed by the European Union (EuroSciVoc) and another from the domain of Science and Technology (UNESCO), establishing relationships between them that are also useful for other universities and projects, as well as for FCT itself.

Having said this, as future work, in addition to improving the systems and creating a manually curated alignment between the three vocabularies, the alignment of vocabularies should also be explored considering other areas besides Computer Science.

REFERENCES

AKSNES, D. W.; SIVERTSEN, G. A criteria-based assessment of the coverage of scopus and web of science. **Journal of Data and Information Science**, v. 4, n. 1, p. 1–21, 2019. Available at: https://doi.org/10.2478/JDIS-2019-0001. Access on: 15 Aug. 2021.

ALGERGAWY, A. Results of the ontology alignment evaluation initiative 2018. *In*: 13th INTERNATIONAL WORKSHOP ON ONTOLOGY MATCHING CO-LOCATED WITH THE, 13th, 17th ISWC (OM 2018). Oct 2018, Monterey, United States, 2018. **Proceedings of the** [...]. Monterey, 2018. p. 76–116. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper0.pdf. Access on: 15 Aug. 2021.

ALGERGAWY, A. **Results of the ontology alignment evaluation initiative** 2019. p. 46–85, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper0.pdf. Access on: 20 Feb. 2021.

BAÑOS-MORENO, M.J. **Propuesta de modelado de una ontología de dominio para la representación de acciones en política-economía**. 2017. 348 f. Tese (Doutorado em Documentación) – Facultad de Comunicación y Documentación, 2018. Available at: http://hdl.handle.net/10201/56661. Access on: 12 Nov. 2021.

BAPTISTA, S. G.; CUNHA, M. B. Estudo de usuários: visão global dos métodos de coleta de dados. **Perspectivas em Ciência da Informação**, v. 12, n. 2, p. 168-184, 2007. Available at: https://doi.org/10.1590/s1413-99362007000200011. Access on: 10 Nov. 2021.

BARRIOS, M. *et al*. Consensus in the delphi method: What makes a decision change? **Technological Forecasting and Social Change**, v. 163, n. C, 2021. Available at: https://doi.org/10.1016/J.TECHFORE.2020.120484. Access on: 10 Nov. 2021.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 284, n. 5, p. 1–5, 2001. Available at: https://www.scientificamerican.com/article/the-semantic-web/. Access on: 10 Sep. 2021.

CHANG, F. **FCAMap-KG results for OAEI 2019**. p. 138–145, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper8.pdf. Access on: 20 Feb. 2021.

CHEATHAM, M.; HITZLER, P. String similarity metrics for ontology alignment. **Lecture Notes in Computer Science**: 8219 LNCS(PART 2), p . 294–309, 2013. Available at: https://doi.org/10.1007/978-3-642-41338-4_19. Access on: 20 Feb. 2021.

CHEN, G. **FCAMapX results for OAEI 2018**, p. 160–166, 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper7.pdf. Access on: 20 Feb. 2021.

DALMORO, M.; VIEIRA, K. M. Dilemas na construção de escalas Tipo Likert: o número de itens e a disposição influenciam nos resultados? **Revista Gestão Organizacional**, v. 6, n. 3, p. 161-174, 2013. Available at: https://doi.org/10.22277/RGO.V6I3.1386. Access on: 15 Aug. 2021.

DELBECQ, A. L. *et al*. Group techniques for program planning: 59 a guide to nominal group and Delphi processes. **The Journal of Applied Behavioral Science**, v. 12, n. 4, p. 581-581, 1986. Available at: https://doi.org/10.1177/002188637601200414. Access on: 10 Nov. 2021.

DESTRO, J. M. **EVOCROS**: results for OAEI 2018, p. 152–159, 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper6.pdf. Access on: 20 Feb. 2021.

DESTRO, J. M. **EVOCROS**: Results for OAEI 2019, p. 131–137, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper7.pdf. Access on: 20 Feb. 2021.

DJEDDI, W. E. **XMap:** results for OAEI 2018, p. 210–215, 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper15.pdf . Access on: 20 Feb. 2021.

DSI. **lod:linked_open_data [DSI Wiki]**. Available at: http://wiki.dsi.uminho.pt/doku.php?id=lod:linked_open_data#linked_open_data. Access on: 30 out. 2021. Access on: 20 Feb. 2021.

ENCYCLOPEDIA. **CORDIS**: European Commission. FP7-NMP - Specific Programme "Cooperation": Nanosciences, Nanotechnologies, Materials and New Production Technologies. 2014. Available at: https://cordis.europa.eu/en. Access on: 15 Aug. 2021.

EUZENAT, J.; SHVAIKO, P. Ontology matching: Second edition. *In*: ______. **Ontology matching**. 2. ed. Berlin: Springer-Verlag Berlin Heidelberg, 2013. Available at: https://doi.org/10.1007/978-3-642-38721-0. Access on: 20 Feb. 2021.

FARIA, D. **Results of AML participation in OAEI 2018**, p. 125–131. 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper2.pdf. Access on: 20 Feb. 2021.

FARIA, D. **AML and AMLC Results for OAEI 2019**, p. 101–106. 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper3.pdf. Access on: 20 Feb. 2021.

FARIA, D. *et al*. The AgreementMakerLight ontology matching system. **Lecture Notes in Computer Science 8185 LNCS**, p. 527–541, 2013. Available at: https://doi.org/10.1007/978-3-642-41030-7_38. Access on: 20 Feb. 2021.

FINK, A. et al. Consensus methods: characteristics and guidelines for use. **American Journal of Public Health**, v. 74, n. 9, p. 979-983, 1984. Available at: https://doi.org/10.2105/AJPH.74.9.979 . Access on: 10 Nov. 2021.

FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA. **Fundação para a Ciência e a Tecnologia**. Available at: https://www.fct.pt/. Acesso em 30 Mar. 2021.

GHERBI, S. **ONTMAT1**: results for OAEI 2019, p. 164–168. 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper12.pdf. Access on: 20 Feb. 2021.

GRACIA, J.; MENA, E. Semantic heterogeneity issues on the web. **IEEE Internet Computing**, v. 16, n. 5, p. 60–67, 2012. Available at: https://doi.org/10.1109/MIC.2012.116. Access on: 20 Feb. 2021.

GUARINO, N.; OBERLE, D.; STAAB, S. **Handbook on ontologies**. Geneve: Springer, 2009. Available at: https://doi.org/10.1007/978-3-540-92673-3. Access on: 5 Feb. 2021.

GUIMARÃES, J. P. P. **Vocabulário de domínio para o projeto IViSSEM**. 2022. 208 f. Dissertação (Mestrado em Tecnologia e Sistemas de Informação) – Escola de Engenharia, Universidade do Minho, Guimarães, Portugal, 2022.

HERTLING, S. **DOME results for OAEI 2019**, p. 123–130. 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper6.pdf. Access on: 20 Feb. 2021.

HERTLING, S. **DOME results for OAEI 2018**, p. 144–151. 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper5.pdf. Access on: 20 Feb. 2021.

IVISSEM . **Information Visualization & Social Scholarly Metric**. (n.d.). Available at: http://www.ivissem.net/. Access on: 5 Sept. 2021.

JACOB, E. K. Ontologies and the Semantic Web. **Bulletin of the American Society for Information Science and Technology**, v. 29, n. 4, p. 19–22, 2005. Available at: https://doi.org/10.1002/bult.283. Access on: 5 Feb. 2021.

JIMÉNEZ-RUIZ, E. **LogMap**: family participation in the OAEI 2018, p. 187–191, 2018. http://ceur-ws.org/Vol-2288/oaei18_paper11.pdf. Access on: 20 Feb. 2021.

JIMÉNEZ-RUIZ, E. **LogMap**: family participation in the OAEI 2019, p. 160–163. 2019. Available at: Available at: http://ceur-ws.org/Vol-2536/oaei19_paper11.pdf. Access on: 20 Feb. 2021.

JIMÉNEZ-RUIZ, E.; CUENCA GRAU, B. **LogMap**: logic-based and scalable ontology matching. Lecture Notes in Computer Science: 7031 LNCS(PART 1), p. 273–288, 2011. Available at: https://doi.org/10.1007/978-3-642-25073-6_18. Access on: 20 Feb. 2021.

KACHROUDI, M. **KEPLER at OAEI 2018**, p. 173–178. 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper9.pdf. Access on: 20 Feb. 2021.

KALIBATIENE, D.; VASILECAS, O. Survey on ontology languages. **Lecture Notes in Business Information Processing 90 LNBIP**, p. 124–141, 2011. Available at: https://doi.org/10.1007/978-3-642-24511-4_10. Access on: 18 Sep. 2021.

LAADHAR, A. **OAEI 2018 results of POMap++,** p. 192–196. 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper12.pdf. Access on: 11 Sep. 2021.

LAADHAR, A. **POMap++ Results for OAEI 2019**: Fully Automated Machine Learning Approach for Ontology Matching. p. 169–174, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper13.pdf. Access on: 18 Sep. 2021.

LINSTONE, H. A.; TUROFF, M.; HELMER, O. (ed.). **The Delphi method**. [*S.l.*]: Murray Turoff and Harold A. Linstone, 2002. Available at: https://bit.ly/3RLQ6kF. Access on: 18 Sep. 2021.

LÜTKE, A. AnyGraphMatcher Submission to the OAEI: knowledge graph challenge 2019. **CEUR**, v. 2536, p. 86–93, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper1.pdf. Access on: 18 Sep. 2021.

MANIRAJ, V.; SIVAKUMAR, R. Ontology Languages: a review. **International Journal of Computer Theory and Engineering**, v. 2, n. 6, p. 1793-8201, 2010. Available at: https://doi.org/10.7763/IJCTE.2010.V2.257. Access on: 18 Sep. 2021.

MENDES, A. Science classification, visibility of the different scientific domains and impact on scientific development. **Revista de Enfermagem Referência**, v. 4, n. 10, p. 143-152,2016. Available at: https://doi.org/10.12707/RIV16049. Access on: 18 Sep. 2021.

MOHAMMADI, M. **SANOM results for OAEI 2018,** p. 205–209. 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper14.pdf. Access on: 18 Sep. 2021.

MOHAMMADI, M. **SANOM results for OAEI 2019,** p. 175–180, 2018. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper14.pdf. Access on: 18 Sep. 2021.

Moher, D. *et al*. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. **PLoS Medicine**, v. 6, n. 7, 2009. Available at: https://doi.org/10.1371/journal.pmed.1000097. Access on: 18 Sep. 2021.

MONGEON, P.; PAUL-HUS, A. The journal coverage of Web of Science and Scopus: a comparative analysis. **Scientometrics**, v. 106, n. 1, p. 213–228, 2015. Available at: https://doi.org/10.1007/S11192-015-1765-5. Access on: 18 Sep. 2021.

MURRY JR., J. W.; HAMMONS, J. O. Delphi: A versatile methodology for conducting qualitative research. **The Review of Higher Education**, v. 18, n. 4, p. 423-436, 1995. Available at: https://doi.org/10.1353/RHE.1995.0008. Access on: 10 Nov. 2021.

OKOLI, C.; PAWLOWSKI, S. D. The Delphi method as a research tool: an example, design considerations and applications. **Information & Management**, v. 42, n. 1, p. 15-29, 2004. Available at: https://doi.org/10.1016/J.IM.2003.11.002. Access on: 10 Nov. 2021.

PASTOR-SÁNCHEZ, J.A. *et al*. Publicación como Linked Open Data de la nomenclatura internacional de Ciencia y Tecnología y del Tesauro UNESCO. *In*: CONGRESSO ISKO ESPANHA E PORTUGAL, 1., 7-9 November 2013. **Anais do** [...]. Porto: ISKO, 2013. Available at: http://eprints.rclis.org/24272/. Access on: Aug. 2021.

PORTISCH, J. **ALOD2Vec matcher**, p. 132–137, 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper3.pdf. Access on: 18 Sep. 2021.

PORTISCH, J. **Wiktionary Matcher**, p. 181–188, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper15.pdf. Access on: 18 Sep. 2021.

PORTUGAL. Ministério da Educação e Ciência. Fundação para a Ciência e a Tecnologia. **Domínios científicos e áreas científicas**. [Lisboa]: [FCT], 2012. Available at: https://www.fct.pt/apoios/projectos/concursos/2012/docs/Dominios_e_Areas_Cientificas_C2012.pdf. Access on: 18 Sep. 2021.

PUBLICATIONS OFFICE OF THE EU. European Science Vocabulary (EuroSciVoc). v.1.3 Luxembourg: Publications Office of the European Union, 2021. Available at: http://publications.europa.eu/resource/dataset/euroscivoc. Access on: 18 Sep. 2021.

RODRIGUEZ, M. C. Three options are optimal for multiple-choice items: a meta-analysis of 80 years of research. **Educational Measurement**: Issues and Practice**,** v. 24, n. 2, p. 3–13, 2005. Available at: https://doi.org/10.1111/J.1745-3992.2005.00006.X. Access on: 18 Sep. 2021.

ROUSSILLE, P. **Holontology**: results of the 2018 OAEI evaluation campaign, p. 167–172, 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper8.pdf. Access on: 18 Sep. 2021.
SILVA, J. **ALIN results for OAEI 2018**, p. 117–124, 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper1.pdf. Access on: 18 Sep. 2021.

SILVA, J. **ALIN results for OAEI 2019**, p. 94–100, 2019. http://ceur-ws.org/Vol-2536/oaei19_paper2.pdf. Access on: 18 Sep. 2021.

SKOS Play! - **Thesaurus & taxonomies**. (*n.d.*). Available at: Available at: https://skos-play.sparna.fr/play/. Acesso em 7 Nov. 2021.
SKOS. Simple Knowledge Organization System. **UNESCO nomenclature for fields of science and technology**. Skos.Um.Es. 2015. Available at: https://skos.um.es/unesco6/. Access on: 18 Sep. 2021.

STEWART, D. *et al*. A modified Delphi study to determine the level of consensus across the European Union on the structures, processes and desired outcomes of the management of polypharmacy in older people. **PLOS One**, v. 12, n. 11, 2017. Available at: https://doi.org/10.1371/JOURNAL.PONE.0188348. Access on: 18 Sep. 2021.

STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: principles and methods. **Data and Knowledge Engineering**, v. 25, n. 1–2, p. 161–197, 1998. Available at: https://doi.org/10.1016/S0169-023X(97)00056-6. Access on: 18 Sep. 2021.

TANG, Y. **Lily results for OAEI 2018**, p. 179–186, 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper10.pdf. Access on: 18 Sep. 2021.

THIÉBLIN, É. **CANARD complex matching system**: results of the 2018 OAEI evaluation campaign, p. 138–143, 2018. Available at: http://ceur-ws.org/Vol-2288/oaei18_paper4.pdf. Access on: 18 Sep. 2021.

THIÉBLIN, É. **CANARD Complex Matching System**: results of the 2019 OAEI Evaluation Campaign, p. 114–122, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper5.pdf. Access on: 18 Sep. 2021.

WANG, X. **FTRLIM results for OAEI 2019**, p. 146–152, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper9.pdf. Access on: 18 Sep. 2021.

ZHOU, L. **AROA results for 2019 OAEI**, p. 107–113, 2019. Available at: http://ceur-ws.org/Vol-2536/oaei19_paper4.pdf. Access on: 18 Sep. 2021.