

## Alinhamento de Vocabulário de domínio utilizando os sistemas AML e LogMap

João Pedro Pereira Guimarães<sup>1</sup> Morgana C. Andrade<sup>2</sup> Ana Alice Baptista<sup>3</sup>

### 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 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 Initiative* 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 Initiative*. Web Semântica. Compatibilização semântica.

### Correspondência do autor

<sup>1</sup>Universidade do Minho, Minho, Portugal / e-mail: [joaoppguimaraes@gmail.com](mailto:joaoppguimaraes@gmail.com)

<sup>2</sup>Universidade Federal do Espírito Santo, Vitória, Brasil / e-mail: [morganaandrade@hotmail.com](mailto:morganaandrade@hotmail.com)

<sup>3</sup>Universidade do Minho, Minho, Portugal / e-mail: [analice@dsi.uminho.pt](mailto:analice@dsi.uminho.pt)

## Domain Vocabulary Alignment using AML and LogMap

### 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.

#### CRedit

- **Reconhecimentos:** Ao projeto IVISSEM – POCI-01-0145-FEDER-28284 e à Universidade Federal do Espírito Santo.
- **Financiamento:** Não é aplicável.
- **Conflitos de interesse:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Comitê de Ética da Universidade do Minho e aprovada sob nº CEICSH 083/2021.
- **Disponibilidade de dados e material:**  
<https://datarepositorium.uminho.pt/dataset.xhtml?persistentId=doi:10.34622/datarepositorium/6IUDJB>.
- **Contribuições dos autores:** Conceituação, Curadoria dos Dados- BAPTISTA, A.A.; Investigação, Redação, Validação - GUIMARÃES, J.P.P.; Supervisão, Redação, Revisão - ANDRADE, M.



JITA: IL. Semantic web.



Artigo submetido ao sistema de similaridade

Submetido em: 16/02/2022 – Aceito em: 01/07/2022 – Publicado em: 20/07/2022

## 1 INTRODUÇÃO

Segundo Berners-Lee et al. (2001), a Web Semântica é uma extensão da Web atual, na qual a informação tem um significado bem definido, permitindo uma melhor cooperação entre computadores e pessoas. Nesse contexto, as ontologias emergem como um meio para a implementação da Web Semântica (JACOB, 2005). Por outras palavras, uma ontologia fornece um vocabulário que descreve um domínio de interesse, e as especificações dos significados dos termos presentes nesse vocabulário (EUZENAT; SHVAIKO, 2013).

No entanto, diferentes conceitos são apresentados para o termo ontologia, seja por conta de domínios específicos, seja pelo entendimento de alguns autores. De acordo com Studer et al. (1998), uma ontologia é definida como “uma especificação formal e explícita de uma conceptualização partilhada”, e em formato legível por máquina e partilhada, dado que a conceptualização deve ser generalizada e aceita por um grupo, e não por apenas uma pessoa (GUARINO et al., 2009).

Desta forma, uma ontologia deve ser capaz de descrever uma interpretação estruturada e consentida de um determinado domínio de forma a permitir que esse conhecimento sobre o domínio seja partilhado por diversos agentes, sendo eles pessoas ou máquinas (BAÑOS-MORENO, 2017).

Devido a necessidade as ontologias apresentarem um formato legível por máquina, inúmeras linguagens foram desenvolvidas ao longo dos anos (KALIBATIENE; VASILECAS, 2011; MANIRAJ; SIVAKUMAR, 2010). Para este trabalho evidencia-se a linguagem RDF, bem como as suas extensões, OWL e SKOS.

Na obra de Euzenat e Shvaiko (2013) são identificados alguns dos muitos modelos conceituais e de dados existentes para descrever o conhecimento, nomeadamente, *tags* e folksonomias, diretórios, bases de dados relacionais, e esquemas XML. Para este trabalho foram considerados apenas dois tipos de vocabulários controlados, o tesouro e a classificação (nomenclatura).

Diante do facto de as ontologias serem conceitualizadas por diferentes pessoas para diferentes propósitos, o seu reaproveitamento acaba por ser limitado. Diferentes comunidades possuem diferentes interesses e hábitos e utilizam diferentes ferramentas e conhecimentos, na maioria das vezes com diferentes níveis de detalhamento (EUZENAT; SHVAIKO, 2013). Essas diferenças resultam em tipos distintos de heterogeneidade, que Euzenat e Shvaiko (2013) distingue da seguinte forma:

- a) heterogeneidade sintática – há quando duas ontologias não estão descritas numa mesma linguagem, ou quando duas ontologias são modeladas por diferentes representações de conhecimento (e.g., *OWL* e *F-Logic*);
- b) heterogeneidade terminológica – dá-se pela utilização de diferentes variações de nomes para descrever uma mesma entidade (e.g., diferentes línguas naturais, sinónimos etc.);
- c) heterogeneidade semântica – ocorre quando o mesmo domínio é modelado de diferente forma, geralmente através da utilização de diferentes axiomas (e.g., diferente cobertura, granularidade, perspectiva etc.); e
- d) heterogeneidade semiótica – verifica-se quando uma mesma entidade apresenta diferentes interpretações para diferentes pessoas.

Contudo, essa heterogeneidade, associada à ambiguidade e à redundância semântica, torna a interoperabilidade entre elas um desafio (GRACIA; MENA, 2012). Com o intuito de superar tais desafios, é importante a existência de métodos/algoritmos capazes de manipular as diferentes ontologias, de forma a misturar, alinhar e/ou fundir estas.

Para reduzir problemas relativos às heterogeneidades, ou possibilitar a reutilização de ontologias, adota-se o alinhamento de ontologias. De acordo com Euzenat e Shvaiko (2013), o alinhamento de ontologias visa encontrar correspondências entre entidades semanticamente relacionadas provenientes de diferentes ontologias. Como resultado do alinhamento de

ontologias surgem as denominadas correspondências que consistem num conjunto de relações entre as entidades das diferentes ontologias que podem ser utilizadas em diversas tarefas, tais como, a fusão de ontologias, pesquisas através de *queries*, tradução de dados ou para navegar por meio da conexão de dados na Web (Web Semântica) (EUZENAT; SHVAIKO, 2013).

Segundo Euzenat e Shvaiko (2013), para obtenção dessas correspondências aplicam-se técnicas de alinhamento de ontologias que podem ser categorizadas com base em duas classificações: Tipo de *input* (*kind of input*), que considera a origem da informação e o tipo de *input* que é utilizado pelas técnicas de alinhamento de ontologias; e Granularidade/Interpretação do *input* (*granularity/input interpretation classification*), que tem por base a granularidade da técnica de alinhamento. Ou seja, essa técnica considera se as entidades das ontologias são exploradas de forma isolada (*element-level*), ou através das suas relações com outras entidades (*structure-level*), e interpretação do *input* que é realizada considerando apenas a própria estrutura (*syntactic*) ou por algum formalismo semântico (*semantics*).

As limitações de espaço não nos permitem discutir cada uma das técnicas de alinhamento em detalhes, mas elas diferem em escopo, foco e estrutura. Os autores Euzenat e Shvaiko (2013) e Cheatham e Hitzler (2013) oferecem descrições detalhadas sobre esse tema. Neste artigo, é considerada a classificação de alinhamentos de acordo com a granularidade, ou seja, pela *Granularity/Input interpretation*.

O aumento do número de sistemas capazes de alinhar ontologias no início dos anos motivou a criação da *Ontology Alignment Evaluation Initiative* (OAEI), iniciativa internacional coordenada, com o objetivo de avaliar e comparar abertamente sistemas e algoritmos de alinhamento de ontologias, permitindo que qualquer pessoa seja capaz de tirar conclusões sobre as melhores estratégias.

A OAEI, em 2018 e 2019, fez uso de um conjunto de ontologias denominadas de *tracks* divididas nos seguintes grupos *Schema Matching tracks*, *Instance Matching tracks*, *Instance and Schema Matching tracks*, *Complex Matching tracks*, *Interactive Matching tracks* (ALGERGAWY, 2018, 2019).

Considerando os benefícios proporcionados pelo alinhamento de ontologias, optou-se por adotá-lo para produzir um vocabulário controlado no domínio da Ciências da Computação que atenda às necessidades do Projeto IVisSEM.

### 1.1 Âmbito do estudo: Projeto IVisSEM

Cerca de 2.5 milhões de artigos são publicados anualmente, em que quase 7 mil são publicados diariamente (IVISSEM, n.d.). Como consequência de volume, a busca de artigos com valor para os diferentes pesquisadores torna-se uma tarefa cada vez mais complicada. Com vista a facilitar tal processo, o projeto IVisSEM visa desenvolver e testar uma nova altimetria (métrica alternativa proveniente das interações dos pesquisadores) denominada de Social Scholarly Experience Metric, que resultará da aplicação de técnicas de *machine learning* e diferentes combinações de altimetrias e perfis de pesquisadores. A sua aplicação irá refletir as preferências individuais no processo de encontrar um tópico específico. Da mesma forma, as atuais listas de resultados de busca com enorme quantidade de itens serão substituídas por uma interface inovadora baseada em técnicas avançadas de visualização.

Atualmente, no terceiro ano de execução, um dos objetivos deste projeto passa pela construção de um vocabulário de domínio de base que terá termos de vocabulários controlados relacionados entre si e com *tags* inseridas pelo utilizador. Para tanto, pode-se adotar o alinhamento de ontologias para obter esse vocabulário. Tema abordado por este estudo.

O presente estudo buscou desenvolver um vocabulário de domínio, Ciência da Computação, de modo a incorporar termos de outros vocabulários controlados relacionados entre si através da implementação de um sistema automático ou semiautomático capaz de fazer o alinhamento entre eles. Foram considerados três vocabulários, 2 multidisciplinares e 1 no

domínio da Ciência e Tecnologia (*EuroSciVoc*, Domínios Científicos e Áreas Científicas da FCT e UNESCO *nomenclature for fields of science and technology*) (PUBLICATIONS OFFICE OF THE EU, 2021; PORTUGAL, 2012; SKOS, 2015).

Além dessa seção, este artigo possui mais três, são elas: Revisão sistemática da literatura, onde é apresentado o procedimento adotado e os resultados que orientaram a escolha dos sistemas a serem utilizados para o alinhamento; Enquadramento metodológico, apresenta os procedimentos metodológicos empregados para nas etapas referente a implementação e validação do alinhamento de vocabulários; Resultados, são apresentados os resultados obtidos com a implementação da técnica/ algoritmo e com o estudo Delphi para avaliação do alinhamento; Conclusão, traz os principais contributos obtidos com a realização do estudo e aponta possibilidades de trabalhos futuros; por fim, são apresentadas as referências dos artigo utilizados neste artigo.

## 2 ALINHAMENTO DE ONTOLOGIAS: REVISÃO SISTEMÁTICA DA LITERATURA

Para obter o conhecimento presente em artigos científicos sobre o tema alinhamento de ontologias foi elaborada uma revisão sistemática da literatura. A revisão sistemática foi desenvolvida com base no *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) (MOHER et al., 2009). O PRISMA consiste em uma *checklist* com 27 itens e um fluxograma com quatro etapas (identificação, seleção, elegibilidade e conclusão), e tem como principal objetivo ajudar os autores a melhorarem a elaboração de revisões sistemáticas e meta-análises (MOHER et al., 2009). Primeiramente, foi realizada uma pesquisa bibliográfica com o objetivo de identificar estudos sobre o alinhamento de ontologias. Foram considerados estudos que referenciavam algoritmos e ou métodos para alinhamento de ontologias.

Os estudos foram identificados, em dezembro de 2020, na base de dados Scopus. A opção pela Scopus deveu-se ao facto da base de dados apresentar uma cobertura de 90% das Ciências Naturais e Engenharia (AKSNES; SIVERTSEN, 2019; MONGEON; PAUL-HUS, 2015). Com o intuito de identificar possíveis métodos/algoritmos para alinhamento de ontologias foi adotada a seguinte estratégia de busca: “*Ontology Matching*” OR “*Ontology Alignment*” OR “*Ontology Mapping*”, no campo “*Article Title, Abstract, Keywords*”. Não houve limite quanto ao tipo de documento, cobertura temporal ou área de aplicação dos estudos. Artigos cujo conteúdo não se encontrava disponível não foram considerados. Também foram considerados artigos que citavam ou eram citados pelos artigos inicialmente identificados. Da execução desta pesquisa resultaram cerca de 3.658 estudos, dos quais foram considerados apenas os 2000 mais relevantes de acordo com o índice de relevância da base de dados.

Posteriormente, foi realizada uma análise dos artigos com base nos títulos e resumos. Alguns dos autores dos estudos foram também contactados para obtenção de informações complementares.

A análise prévia dos 100 primeiros artigos apontou para a relevância da competição *Ontology Alignment Evaluation Initiative* (OAEI) nesta temática. A OAEI é uma iniciativa internacional coordenada que organiza anualmente a avaliação de sistemas de alinhamento ontologias de forma controlada, através de um vasto conjunto de dados (ALGERGAWY, 2018, 2019). Diante dessa constatação, optou-se por realizar a revisão sistemática da literatura a partir dos artigos referentes às participações dos sistemas na OAEI no período de 2018 e 2019. Assim, foram considerados apenas os artigos referentes à competição e à participação dos sistemas nos dois anos, bem como os artigos externos à competição que descrevem os sistemas participantes identificados na base de dados Scopus. Desta seleção, resultaram 34 artigos.

A partir desses artigos foram extraídas informações referentes às estratégias e às técnicas adotadas por cada abordagem para o alinhamento de ontologias, assim como a

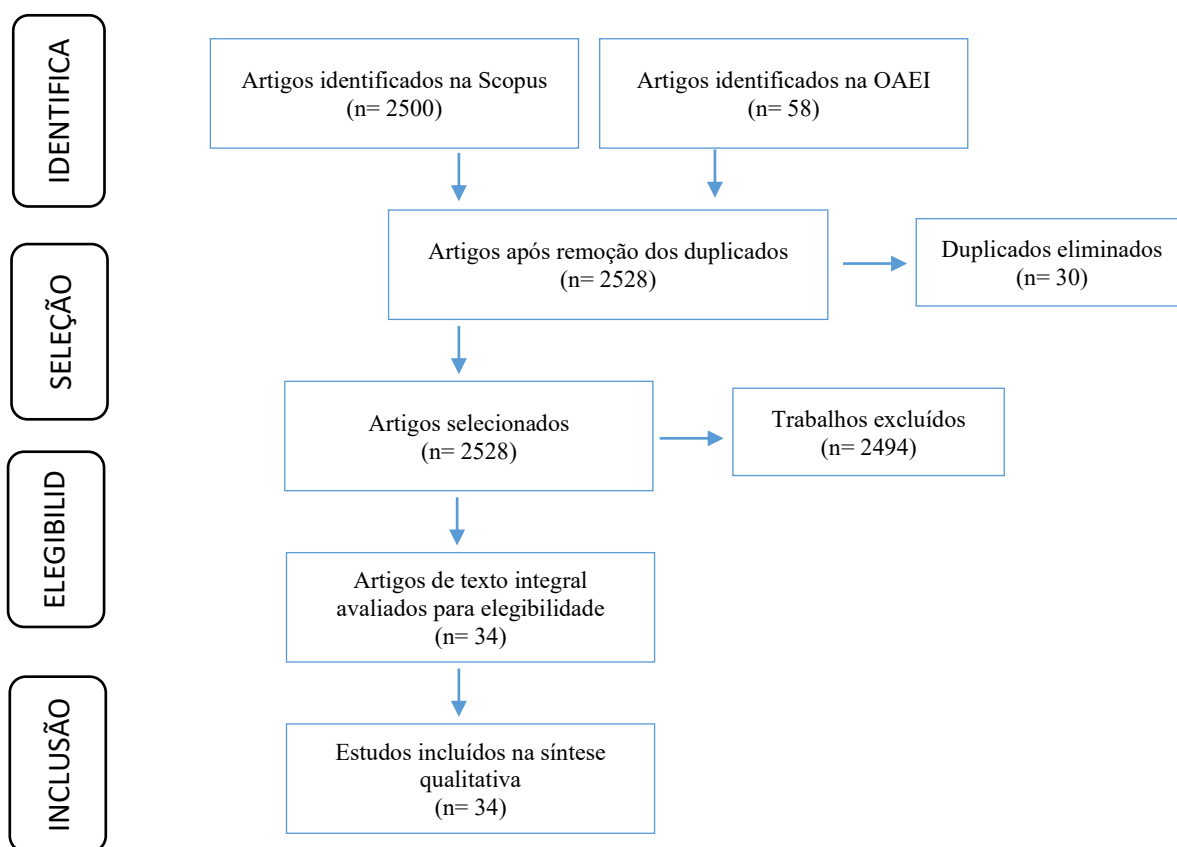
performance de cada sistema nas diferentes *tracks* da competição anual da OAEI nos anos de 2018 e 2019.

Vale destacar que, apesar de serem considerados estudos cujas performances são avaliadas no mesmo ambiente controlado (OAEI), o risco de viés diminui, porém continua a existir, uma vez que cada sistema apresenta diferentes performances em diferentes *tracks*, o que pode influenciar a análise da qualidade dos sistemas e identificação do “melhor” sistema.

A opção por utilizar os artigos apresentados na OAEI foi ratificada a partir da realização da análise bibliométrica que incluiu os 2.000 itens recuperados na Scopus. A análise bibliométrica indicou a conferência como sendo o evento mais relevante sobre o tema, assim como a identificação dos autores mais proeminentes. Os resultados da análise bibliométrica foram apresentados em outro artigo (submetido para publicação).

Na Figura 1, encontra-se ilustrado o processo de seleção dos estudos.

Figura 1. Fluxograma PRISMA



Fonte: Adaptado de Moher *et al.* (2009)

A OAEI, em 2018 e 2019, fez uso de um conjunto de *tracks* divididas nos seguintes grupos: (a) *Schema Matching tracks* que visam avaliar o alinhamento de classes e propriedades de ontologias; (b) *Instance Matching tracks* que visam avaliar o alinhamento de instâncias de ontologias; (c) *Instance and Schema Matching tracks* que visam avaliar tanto o alinhamento de classes e propriedades como instâncias de ontologias; (d) *Interactive Matching tracks* que visam avaliar a performance dos sistemas através da simulação de interação por parte do utilizador; e (e) *Complex Matching tracks* visam avaliar a obtenção de correspondências complexas (n:m) por parte dos sistemas (ALGERGAWY, 2018, 2019). Tal divisão é observada no Quadro 1.

Quadro 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 (apenas em 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

Fonte: ALGERWAY, 2018, 2019

Como referido anteriormente, as *tracks foram analisadas* considerando as técnicas de alinhamento utilizadas por cada um dos sistemas, bem como as suas performances na OAEI. No que concerne às técnicas de alinhamento utilizadas por eles, estas foram analisadas tendo por base a classificação *Granularity/Input interpretation*. Esta análise foi resultado da interpretação pessoal dos artigos dos sistemas, bem como da troca de informações, via e-mail, com os respectivos autores.

Em relação à performance dos sistemas na OAEI, esta foi avaliada tendo por base os fatores *precision*, *recall* e *F-measure* de cada um dos sistemas nas diversas *tracks* (ALGERWAY, 2018, 2019). Vale salientar que os artigos publicados sobre cada um dos sistemas, ainda que sob forma de documentos da OAEI, não foram revistos pelos seus pares da OAEI, o que pode resultar em inconsistências.

Após a análise dos 34 artigos, procedeu-se à categorização dos diferentes métodos/algoritmos utilizados por cada um dos sistemas. Em relação ao quadro, este descreve os diferentes tipos de técnicas existentes de acordo com o Euzenat e Shvaiko (2013), sendo que os diferentes sistemas exploram distintos métodos de avaliação. Desta forma, existem sistemas que, por exemplo, exploram técnicas caracterizadas como *Structure-level* e outros exploram outras técnicas caracterizadas com *Element-level*. Para cada sistema foram então identificados cada uma das técnicas utilizadas e categorizadas de acordo com os respetivos métodos descritos por Euzenat e Shvaiko (2013). No Quadro 2, encontram-se categorizados os métodos utilizados por cada um dos sistemas.

Quadro 2. Categorização dos métodos de alinhamento

Systems	Element-level		Structure-level	
	Syntactic	Semantic	Syntactic	Semantic
AGM (LÜTKE, 2019)	String-based através de Levenshtein distance, Skip-gram neural model, cosine similarity e Euclidean distance	Não analisado	Graph-based	Não analisado

Systems	Element-level		Structure-level	
	Syntactic	Semantic	Syntactic	Semantic
ALIN (SILVA, 2018, 2019)	Language-based através de tokenization e lemmatization, String-based através de Jaccard similarity, Jaro-Wrinkler similarity e n-gram, WordNet	FMA ontology	Graph-based	Não analisado
ALOD <sub>2</sub> Vec (PORTISCH, 2018)	Language-based, String-based através de neural language model, cosine similarity	WebIsALOD data set	Não analisado	Não analisado
AML & AMLC (FARIA, 2018, 2019; FARIA et al., 2013)	Language-based, String-based através de Jaccard similarity	Uber Anatomy Ontology (Uberon), Human Disease Ontology (DOID) e the Medical Subject Headings (MeSH)	Taxonomy-based	Não analisado
AROA (ZHOU, 2019)	Não analisado	Não analisado	Association rule-based with FP-growth	Não analisado
CANARD (THIÉBLIN, 2018, 2019)	String-based através de Label similarity	Não analisado	Structural similarity	Não analisado
DOMÉ (HERTLING, 2018, 2019)	String-based, Language-based, cosine similarity	Não analisado	Instance-based	Não analisado
EVOCROS (DESTRO, 2018, 2019)	Language-based, String-based através de Levenshtein e Jaro similarities, WordNet, BabelNet	Não analisado	Não analisado	Não analisado
FCAMapX & FCAMap-KG (CHANG, 2019; CHEN, 2018)	Language-based através de Token-based formal context	Não analisado	Formal Concept Analysis (Mathematical model for structuring concept hierarchies from clustering individuals)	Não analisado
FTRLIM (WANG, 2019)	FTRL model, TF-IDF, edit distance similarity e Jaccard similarity	Não analisado	Não analisado	Não analisado
Holontology (ROUSSILLE, 2018)	Exact match, Levenshtein, Jaccard e Lin similarities, constraint-based	Não analisado	Não analisado	Não analisado
KEPLER (KACHROUDI, 2018)	WordNet, Microsoft Bing Translator	Não analisado	Não analisado	Não analisado
Lily (TANG, 2018; WU, 2019)	String-based	BioPorta, UMLS Metathesaurus	Similarity propagation	Não analisado



Systems	Element-level		Structure-level	
	Syntactic	Semantic	Syntactic	Semantic
<i>LogMap</i> , <i>LogMapBio</i> , <i>LogMapIIMB</i> , <i>LogMapKG</i> & <i>LogMapLt</i> (JIMÉNEZ-RUIZ, 2018, 2019; JIMÉNEZ-RUIZ; CUENCA GRAU, 2011)	String-based	<i>BioPortal</i> , <i>UMLS</i> <i>Metathesaurus</i>	Structure comparison	Logic-based, Propositional Horn reasoning, Dowling- Gallier algorithm
<i>ONTMAP</i> <sub>1</sub> (GHERBI, 2019)	<i>n</i> -gram, lemmatization, cardinality constraints, <i>WordNet</i>	Não analisado	Structural similarity between properties	Não analisado
<i>POMap++</i> (LAADHAR, 2018, 2019)	Language-based between natural language processes	Não analisado	Structural similarity between entities	Não analisado
<i>SANOM</i> (MOHAMMADI, 2018, 2019)	Tokenization, stop word removal, stemming (lemmatization technique), Jaro-Winkler similarity, <i>WordNet</i>	Não analisado	Taxonomy- based, subsumption relation of classes	Não analisado
<i>XMAP</i> (DJEDDI, 2018)	String and linguistic similarities	<i>UMLS</i> <i>Metathesaurus</i>	Structural similarity	Não analisado
<i>Wiktionary</i> (PORTISCH, 2019)	String-based	<i>Wiktionary</i>	Não analisado	Não analisado

Fonte: Autoria própria

Ao analisar os resultados referentes às performances dos sistemas apresentados pela OAEI (ALGERWAY, 2018, 2019) foram selecionados para este estudo dois sistemas para implementação, os sistemas *AML* e *LogMap* (Cf. GUIMARÃES, 2022, p. 130-164). Esta seleção resulta da qualidade da performance dos dois sistemas nas diferentes *tracks*, uma vez que estes dois sistemas apresentavam as melhores performances em grande parte das *tracks*. De um total de 12 *tracks*, os sistemas *AML* e *LogMap*, juntamente com as suas variantes, destacaram-se em 10 e 8 delas, respetivamente.

Um outro fator que teve impacto nesta decisão foi também o número de citações dos artigos sobre os sistemas obtido pela análise bibliométrica, em que os seus autores também são bastante citados nesta temática.

### 3 PROCEDIMENTOS METODOLÓGICOS

De forma a alcançar o objetivo proposto, ou seja, criar um vocabulário para o projeto IVISSEM, foi definido um conjunto de etapas, bem como os procedimentos metodológicos utilizados em cada uma delas.

Assim foram definidas três etapas: seleção dos vocabulários controlados; implementação dos sistemas para alinhamento dos vocabulários controlados; e avaliação dos alinhamentos produzidos na implementação.

### 3.1 Seleção dos Vocabulários Controlados

Para o desenvolvimento do vocabulário para o projeto IVISSEM foram considerados três vocabulários controlados que representam as principais áreas de conhecimento: *EuroSciVoc*, *UNESCO Nomenclature for fields of Science and Technology e Domínios Científicos e Áreas Científicas* da Fundação para a Ciência e Tecnologia (FCT) do Ministério da Educação e Ciência de Portugal. A seleção dos vocabulários *EuroSciVoc* e *UNESCO Nomenclature for Fields of Science and Technology* deve-se ao facto de estes estarem associados a instituições internacionais de relevo, enquanto a seleção do vocabulário da FCT foi devido a este ter como provedor uma agência pública de Portugal no contexto da Ciência e Tecnologia, país onde o projeto está a ser desenvolvido. No Quadro 3 encontram-se detalhados os referidos vocabulários.

Quadro 3. Vocabulários controlados utilizados

Vocabulário	Provedor	Número de termos	Idiomas
<i>EuroSciVoc</i>	<i>Publications Office of the EU</i>	991	Alemão, inglês, espanhol, francês, italiano, polaco
<i>UNESCO nomenclature for fields of science and technology</i>	<i>UNESCO</i>	2504	Alemão, inglês, espanhol, francês, italiano, polaco, português
<i>Domínios Científicos e Áreas Científicas</i>	<i>FCT</i>	178	Inglês, português

Fonte: Autoria própria

10

Devido ao número de termos presentes em dois vocabulários (*UNESCO* e *EuroSciVoc*), de forma a não comprometer a validação do alinhamento final e a qualidade do vocabulário resultante, foi necessário delimitar os vocabulários, tendo em conta apenas a área da Ciência da Computação.

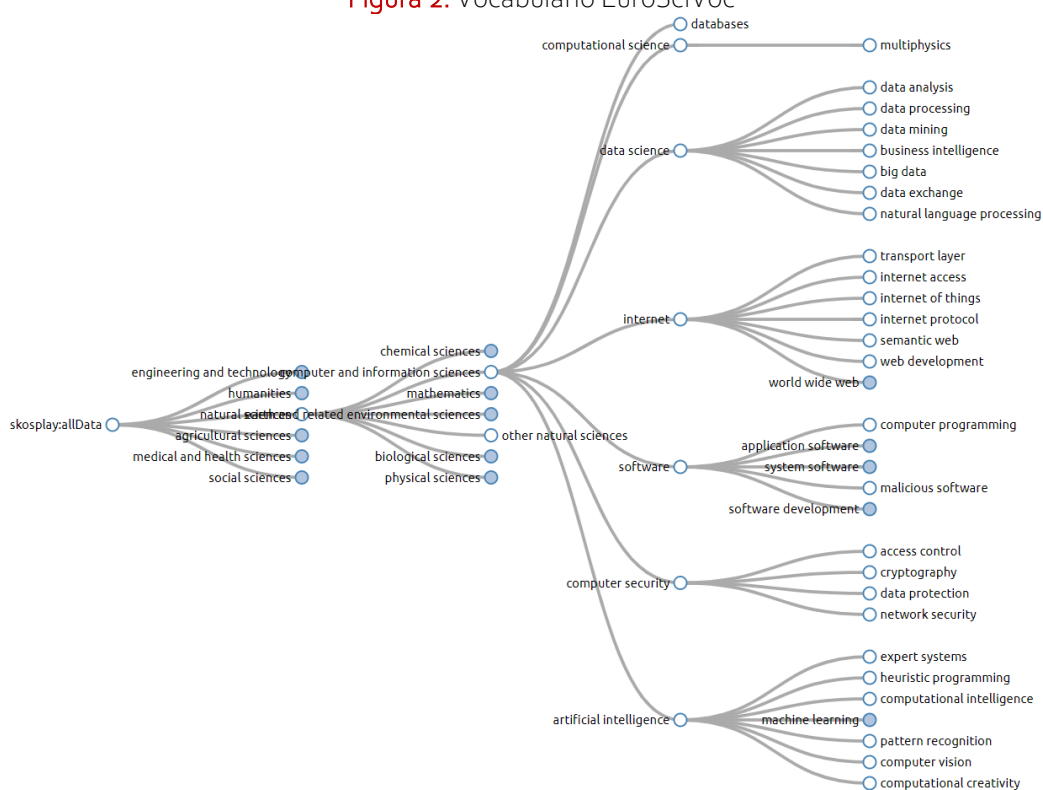
#### 3.1.1 *EuroSciVoc*

*The European Science Vocabulary (EuroSciVoc)* é uma taxonomia multilíngue baseada no manual da Frascati de 2015, da Organização para a Cooperação e Desenvolvimento Económico (OCDE), que representa todos os principais campos da Ciência. Foi ampliado com campos da ciência extraídos a partir do conteúdo do *Community Research and Development Information Service (CORDIS)* (ENCYCLOPEDIA, 2014) e organizado através de um processo semiautomático baseado em técnicas de *Natural Language Processing (NLP)*.

Uma vez que este vocabulário se apresenta descrito em SKOS-XL (extensão do SKOS que permite usar *labels* como recursos), houve a necessidade de convertê-lo manualmente em SKOS, permitindo assim efetuar o alinhamento com os outros vocabulários. Esta necessidade surgiu do facto do sistema *LogMap* não produzir alinhamentos entre ficheiros SKOS-XL e SKOS.

Após a delimitação do vocabulário, de forma a evidenciar a área da Ciência da Computação, foi tido em consideração o ramo resultante do termo “*computer and information science*”, apresentado na Figura 2. O vocabulário completo pode ser observado em Publications Office of the EU (2021).

Figura 2. Vocabulário EuroSciVoc



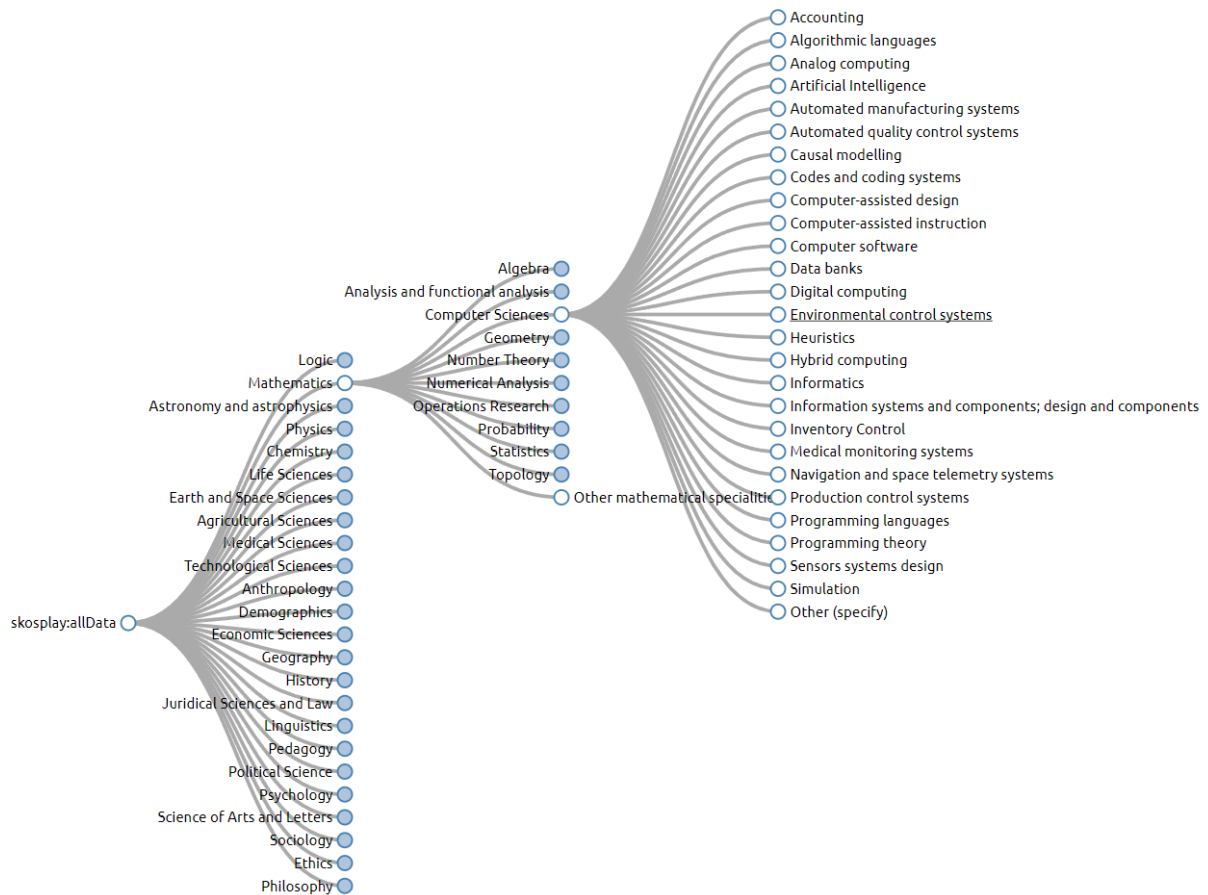
Fonte: adaptado Publications Office of the EU (2021)

### 3.1.2 UNESCO Nomenclature for Fields of Science and Technology

Proposta em 1973 e 1974 pela Divisão de Política Científica e Estatística para a Ciência e Tecnologia da UNESCO e adotada pelo Comité Científico Consultivo, a *UNESCO Nomenclature for Fields of Science and Technology* é amplamente utilizada na gestão do conhecimento de projetos de investigação e dissertações. Esta encontra-se dividida em três níveis hierárquicos: áreas, disciplinas e subdisciplinas (SKOS, 2015; PASTOR-SÁNCHEZ; RODRIGUEZ-MUÑOZ; LOPEZ-CARREÑO, 2013).

Foi efetuado um corte a partir do ramo resultante do termo “*Computer Science*”, apresentado na Figura 3. O vocabulário completo pode ser observado em SKOS... (2015).

Figura 3. Vocabulário UNESCO *Nomenclature for Fields of Science and Technology*



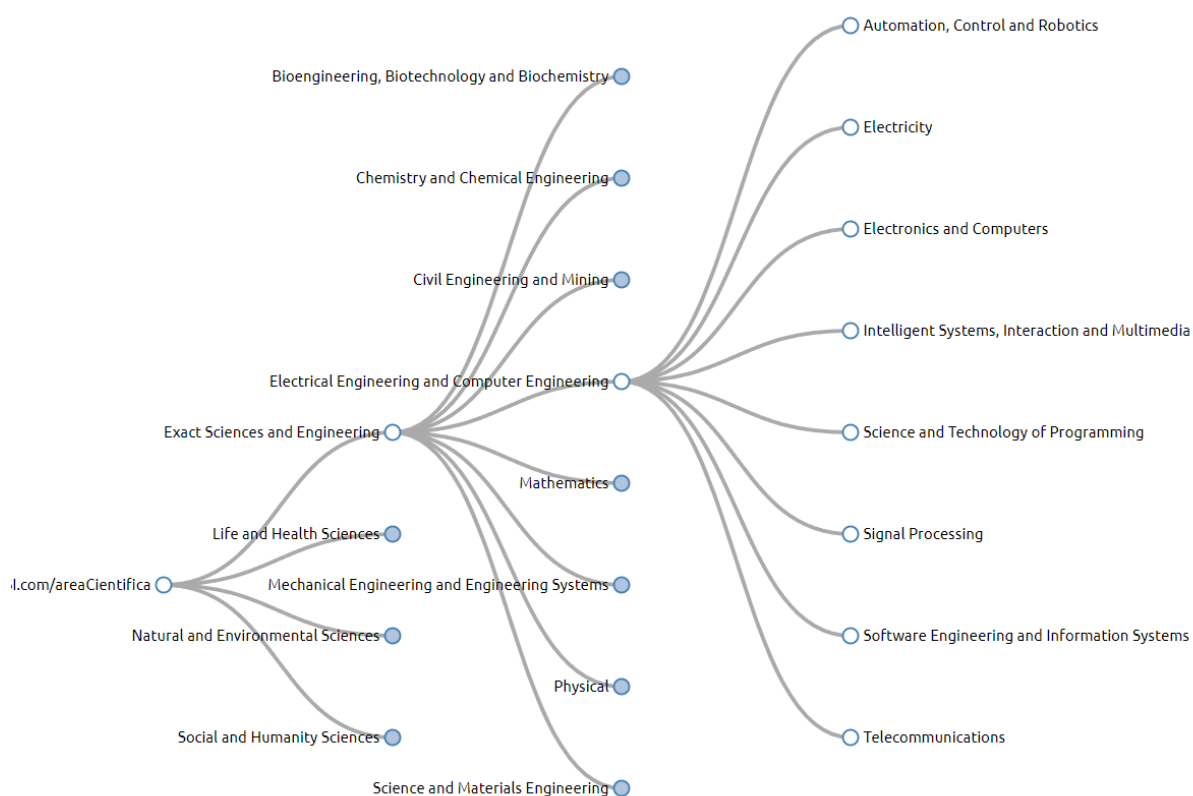
Fonte: adaptado de SKOS... (2015)

### 3.1.3 Domínios Científicos e Áreas Científicas da FCT

O vocabulário *Domínios Científicos e Áreas Científicas* utilizado pela FCT é um sistema de classificação, organizado por áreas e domínios científicos, onde cada domínio apresenta um conjunto de áreas e subáreas (MENDES, 2016). Este vocabulário é disponibilizado no formato PDF, portanto, foi utilizada uma codificação deste em RDF, desenvolvida por um grupo de alunos no âmbito da unidade curricular de Web Semântica do Curso de Sistemas de Informação da Universidade do Minho, unidade de Guimarães (Portugal), e disponibilizada na DSI Wiki (n.d.).

Para análise do vocabulário foi considerado o ramo resultante do termo “*Electrical Engineering and Computer Engineering*”, apresentado na Figura 4. O vocabulário completo encontra-se disponibilizado em Fundação para a Ciência e a Tecnologia (2012).

Figura 4. Vocabulário Domínios Científicos e Áreas Científicas



Fonte: adaptado de Fundação para a Ciência e a Tecnologia (2012)

Com adoção dos critérios de corte referidos anteriormente foi possível avançar para o alinhamento dos vocabulários, descrito na próxima seção.

### 3.2 Implementação dos sistemas para alinhamento dos vocabulários controlados

O vocabulário controlado para o projeto IVisSEM resultou do alinhamento dos três vocabulários anteriormente descritos, tendo em conta apenas os termos referentes à área da Ciência da Computação.

Assim, através da aplicação do *AML* e da interface web do *LogMap*, os vocabulários foram alinhados um a um (vocabulário A com vocabulário B e vice-versa). Estes alinhamentos apresentavam-se em formato *OAEI Alignment Format*. Neste formato, cada correspondência é composta pelos termos identificados como correspondentes (provenientes dos vocabulários *source e target*) e pelo valor de confiança. Este valor diz respeito à confiança do sistema para alinhamento na correspondência entre os termos, com variação entre 0 e 1, onde 1 é valor máximo de confiança. Para este trabalho, foram apenas considerados alinhamentos com um valor de confiança superior a 0.6. Vale salientar que, no que concerne aos valores de confiança, estes variam de sistema para sistema, resultado dos diferentes algoritmos que cada um implementa.

### 3.3 Avaliação dos alinhamentos produzidos

Para avaliar a coerência dos alinhamentos sugeridos pelos dois sistemas foi aplicado o Estudo Delphi (LINSTONE et al., 1975) com o uso de um questionário, destinado a indivíduos com conhecimentos consolidados em Ciência da Computação e Ciência da Informação.

Os estudos Delphi nos últimos anos têm sido adotados por pesquisadores nas mais diversas áreas do conhecimento e, embora, seu uso tenha se ampliado não há um consenso sobre a forma como deve ser aplicado, seja em relação ao número de participantes ou quantas rondas devem ser adotadas (BAPTISTA; CUNHA, 2007; OKOLI; PAWLOWSKI, 2004).

O número de participantes e de rondas pode ser definido mediante o perfil dos entrevistados e o objetivo do estudo (FINK et al., 1984). A seleção apropriada de painelistas ou participantes é fundamental para promover a confiabilidade e validade dos dados. Para este estudo Delphi, o painel incluiu 10 representantes das duas áreas em destaque e de perfil profissional condizente com o uso das terminologias (STEWART et al., 2017).

Quanto à quantidade de rondas, essa foi definida após a obtenção da primeira aplicação do questionário, visto que alguns autores preconizam que essa quantidade pode variar de acordo com o índice de concordância ou consenso obtido entre os participantes (DELBECQ et al., 1986). Os índices de consenso variam na literatura entre 70 e 80% (BARRIOS et al., 2021; MURRY; HAMMONS, 1995), no entanto para este estudo o índice de consenso adotado foi de 70% ao somar as concordâncias parcial e total.

Primeiramente foi desenvolvido um questionário piloto, enviado a cinco colaboradores, com o intuito de avaliar a clareza, tamanho e objetividade do questionário. Foi solicitado aos colaboradores que analisassem o questionário e apresentassem sugestões para a melhoria.

A partir das sugestões de três respondentes, foi então elaborada a versão final do questionário e enviado a dez especialistas das áreas da Ciência da Computação e as sugestões dos entrevistados foram consideradas. Esta amostra consistiu em oito indivíduos doutorados e dois mestres. Cinco eram provenientes de Portugal e cinco do Brasil. Após finalizado o questionário, foram enviados convites para os respondentes, tendo todos aceitado o convite. Os participantes dessa fase não foram incluídos no questionário aplicado na primeira ronda. A pesquisa foi submetida ao Comitê de Ética da Universidade do Minho e aprovada sob nº CEICSH 083/2021.

O questionário adotado para o estudo foi elaborado recorrendo à ferramenta Google *forms*, estruturado em duas partes. Na primeira parte, foi apresentado um termo de consentimento livre e esclarecido, que quando confirmado, os respondentes eram convidados a responderem um conjunto de questões demográficas (formação, país de residência, e atividade profissional), que permitiam caracterizar os indivíduos que respondiam ao questionário.

Na segunda parte, foram disponibilizados os alinhamentos entre os diferentes vocabulários gerados pelos sistemas, bem como as árvores dos respectivos vocabulários, a fim de facilitar a tomada de decisão por parte dos indivíduos em relação aos alinhamentos. Para obtenção das árvores dos respectivos vocabulários foi utilizado o *SKOS Play* (*SKOS Play!*, n.d.).

No que concerne aos alinhamentos gerados, para cada correspondência foi utilizada uma escala de Likert, com três opções de resposta (concordo parcialmente, concordo totalmente e discordo). A escolha do número de opções de resposta recai sobre as análises dos artigos (DALMORO; VIEIRA, 2013; RODRIGUEZ, 2005). Para que as respostas fornecidas não fossem influenciadas por possíveis preferências dos especialistas quanto aos sistemas utilizados, foi omitido o nome do sistema responsável por cada um dos alinhamentos. Assim, os sistemas *LogMap* e *AML* foram identificados como sistema 1 e 2, respectivamente. Para cada alinhamento foi ainda disponibilizado um campo para que os especialistas efetuassem observações.

Na Figura 5 é possível observar a estrutura de cada um dos alinhamentos presentes na segunda parte do questionário, tendo como exemplo o alinhamento entre o vocabulário *UNESCO Nomenclature for Fields of Science and Technology* e *EuroSciVoc*.

Figura 5. Estrutura de cada alinhamento presente na segunda parte do questionário

**Alinhamento entre Unesco e EuroSciVoc**

Descrição (opcional)

Árvore do vocabulário Unesco

Árvore do vocabulário EuroSciVoc

**A**

Unesco -> EuroSciVoc (sistema 1) \*

	Discordo	Concordo parcialmente	Concordo totalmente
navigation and space tel...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
artificial intelligence -> a...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
computer software -> so...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
computer-assisted instru...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
informatics -> hydroinfor...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
data banks -> databases	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
computer sciences -> co...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
computer sciences -> co...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
computer sciences -> co...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Observações acerca dos alinhamentos acima discriminados

Texto de resposta longa

---

Unesco -> EuroSciVoc (sistema 2) \*

	Discordo	Concordo parcialmente	Concordo totalmente
artificial intelligence -> a...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
computer software -> so...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
heuristics -> heuristics p...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
computer sciences -> co...	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Observações acerca dos alinhamentos acima discriminados

Texto de resposta longa

**B**

Fonte: Autoria própria

Os resultados obtidos a partir desses procedimentos são relatados na seção seguinte e disponibilizados no Repositório da Universidade do Minho.

## 4 RESULTADOS

Nesta seção são apresentados os principais resultados obtidos no processo de alinhamento dos vocabulários e de avaliação.

### 4.1 Alinhamento dos vocabulários

Foi produzido um total de 12 alinhamentos (6 alinhamentos por aplicação de cada um dos sistemas) disponibilizados no RepositóriUM<sup>1</sup>. Todas as análises e seus resultados podem ser consultados em Guimarães (2022, p. 44-60).

Após analisar os alinhamentos produzidos, constatou-se que o sistema *AML* gerou correspondências em todos os alinhamentos à exceção dos alinhamentos entre *EuroSciVoc-FCT*, e *Unesco-FCT*. Verificou-se também que o maior valor de confiança atribuído a um alinhamento foi de 0.92 e o menor de 0.61. Correspondências entre termos iguais apresentavam um valor de confiança maior.

<sup>1</sup> <https://doi.org/10.34622/datarepositorium/6IUDJB>

Relativamente aos alinhamentos produzidos pelo sistema *LogMap*, constatou-se que este produziu correspondências em todos os alinhamentos, exceto no alinhamento entre Unesco-FCT. O facto de nenhum dos dois sistemas ter produzido correspondências neste alinhamento pode indicar que, de facto, não existe nenhuma correspondência. Verificou-se também que o maior valor de confiança atribuído a um alinhamento foi de 1.0 e o menor de 0.77. Comparativamente com os alinhamentos produzidos pelo sistema *AML*, observou-se ainda que o sistema *LogMap* produziu um maior número de correspondências em todos os alinhamentos. Uma das causas pode ser o facto de o sistema *LogMap* apresentar correspondências com maior valor de confiança. Assim, é possível que algumas destas correspondências poderiam ter sido sugeridas também pelo sistema *AML*, porém foram descartadas pelo facto de apresentarem um valor de confiança inferior a 0,6 (estabelecido como valor mínimo de confiança). Ainda possível constatar que muitas destas correspondências se devem ao simples facto de alguns termos partilharem algumas palavras entre si, mas considerando que apresentam um contexto totalmente diferente, as torna inválidas. Como exemplo temos o alinhamento de dois termos (EuroSciVoc e FCT) realizado pelo sistema *LogMap* que não obtiveram consenso entre os especialistas: *computer and information sciences* -> *Science of communication and information (narrower of institutions, values, beliefs and behavior)*.

#### 4.2 Avaliação do alinhamento dos vocabulários

A avaliação do alinhamento foi realizada por 10 profissionais, na sua maioria formado por docentes, seguidos de pesquisadores e bibliotecários.

As respostas relativas às correspondências geradas por cada um dos sistemas foram dadas com base nas três opções de concordância disponíveis: discordo, concordo parcialmente e concordo totalmente. Para cada uma das correspondências, caso o número de discordâncias fosse superior ao número de concordâncias (totais ou parciais), a correspondência seria descartada, do contrário, a correspondência seria validada e aceita. No caso de haver mais concordâncias e o número de concordâncias parciais ser superior ou igual ao número de concordâncias totais, a correspondência foi analisada mais detalhadamente, e posteriormente, consoante o veredicto da análise, foi aceita ou não. Além das correspondências, os especialistas também puderam fazer sugestões sobre os alinhamentos e os respectivos vocabulários.

Por fim, para cada alinhamento foi ainda calculado a *precision* dos sistemas, com o propósito de analisar a performance de ambos e identificar qual o melhor. Dada a inexistência de um alinhamento de referência, o *recall* e *F-measure* dos sistemas foram calculados através de alinhamentos de referência *silver standard*, isto é, um alinhamento de referência que não é necessariamente correto e completo, que tem por base apenas as correspondências aceitas pelos especialistas em cada alinhamento. Para cada alinhamento foram obtidos os seguintes resultados (Tabelas 1-6).

**Tabela 1.** Alinhamento entre os vocabulários *EuroSciVoc* e *Unesco Nomenclature for Fields of Science and Technology*

Sistema	Correspondências sugeridas	Correspondências analisadas detalhadamente	Correspondências aceitas	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>
<i>AML</i>	5	2	5	1.0	0.83	0.91
<i>LogMap</i>	19	2	5	0.26	0.83	0.40

Fonte: Dados da pesquisa



**Tabela 2.** Alinhamento entre os vocabulários *Unesco Nomenclature for Fields of Science and Technology* e *EuroSciVoc*

Sistema	Correspondências sugeridas	Correspondências analisadas detalhadamente	Correspondências aceitas	Precision	Recall	F-measure
<i>AML</i>	4	2	4	1.0	0.67	0.80
<i>LogMap</i>	9	3	4	0.44	0.67	0.53

Fonte: Autoria própria

**Tabela 3.** Alinhamento entre os vocabulários *EuroSciVoc* e *Domínios Científicos e Áreas Científicas da Fundação para a Ciência e Tecnologia*

Sistema	Correspondências sugeridas	Correspondências analisadas detalhadamente	Correspondências aceitas	Precision	Recall	F-measure
<i>AML</i>	0	0	0	-	-	-
<i>LogMap</i>	2	0	0	0.0	-	-

Fonte: Dados da pesquisa

**Tabela 4.** Alinhamento entre os vocabulários *Domínios Científicos e Áreas Científicas da Fundação para a Ciência e Tecnologia* e *EuroSciVoc*

Sistema	Correspondências sugeridas	Correspondências analisadas detalhadamente	Correspondências aceitas	Precision	Recall	F-measure
<i>AML</i>	2	1	2	1.0	0.5	0.67
<i>LogMap</i>	7	6	4	0.57	1.0	0.73

Fonte: Dados da pesquisa

**Tabela 5.** Alinhamento entre os vocabulários *Unesco Nomenclature for Fields of Science and Technology* e *Domínios Científicos e Áreas Científicas da Fundação para a Ciência e Tecnologia*

Sistema	Correspondências sugeridas	Correspondências analisadas detalhadamente	Correspondências aceitas	Precision	Recall	F-measure
<i>AML</i>	0	0	0	-	-	-
<i>LogMap</i>	0	0	0	-	-	-

Fonte: Dados da pesquisa

**Tabela 6.** Alinhamento entre os vocabulários *Domínios Científicos e Áreas Científicas da Fundação para a Ciência e Tecnologia* e *Unesco nomenclature for fields of science and technology*

Sistema	Correspondências sugeridas	Correspondências analisadas detalhadamente	Correspondências aceitas	Precision	Recall	F-measure
<i>AML</i>	1	0	1	1.0	0.33	0.50
<i>LogMap</i>	3	2	3	1.0	1.0	1.0

Fonte: Dados da pesquisa

De um modo geral, a partir da análise dos diferentes alinhamentos, é possível observar que o sistema *LogMap*, comparativamente ao sistema *AML*, foi capaz de produzir mais correspondências em todos os alinhamentos. Porém, observou-se que parte das correspondências feitas pelo sistema *LogMap* muitas vezes eram descartadas pelos especialistas, resultando assim na maior parte das vezes numa fraca *precision* nos diferentes alinhamentos. Em contraste, observou-se que o sistema *AML*, produziu menos correspondências, porém todas elas aceitas pelos especialistas, resultando, assim, numa melhor performance (em termos de *precision*) em todos os alinhamentos em que este produziu correspondências, obtendo em todos eles uma *precision* de 1.0.

Em termos de *F-measure*, ambos os sistemas apresentaram resultados com um intervalo de valores similar, sendo que houve alinhamentos em que se destacava o sistema *AML* e outros o sistema *LogMap*.

Assim, concluiu-se que o sistema *AML* é o sistema mais consistente, e por sua vez mais indicado para ser adotado no desenvolvimento do vocabulário controlado para o projeto IVISSEM.

Em vista do resultado obtido com a aplicação do questionário (primeira ronda Delphi), em que se identificou um índice de consenso de 70% em quase todas as correspondências, concluiu-se que não haveria necessidade de uma nova ronda (DELBECQ et al., 1986; MURRY.; HAMMONS, 1995).

#### 4.3 Produto final – Vocabulário para o projeto IVISSEM

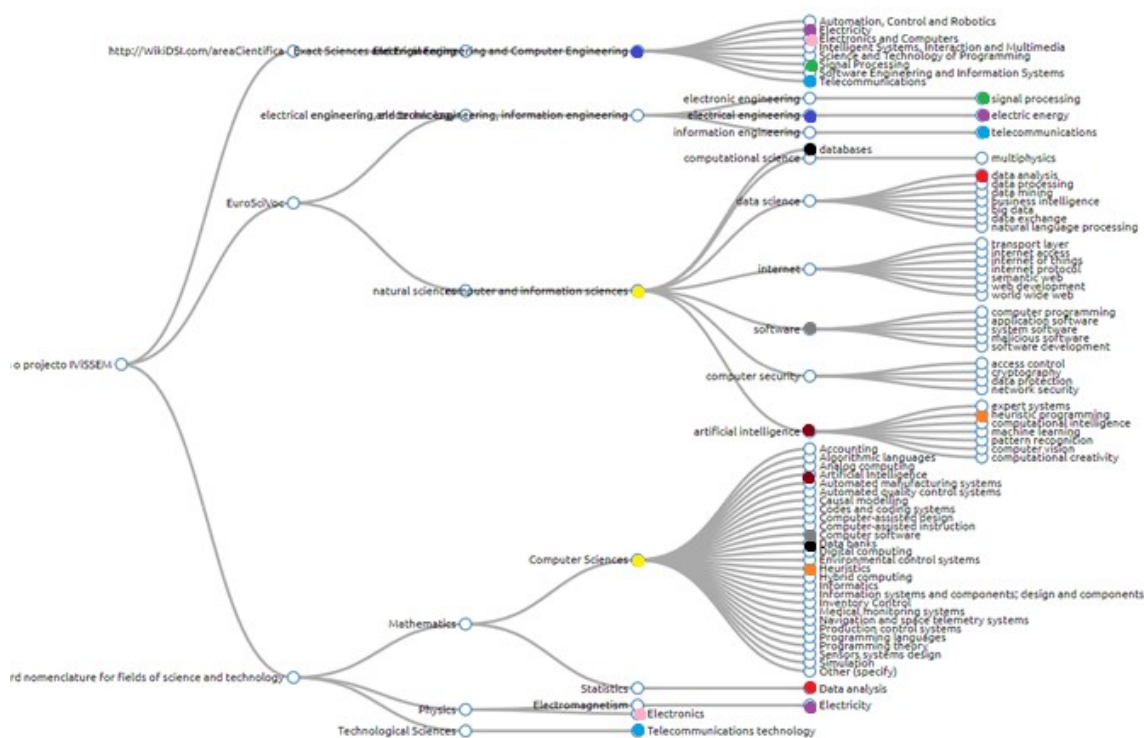
Tendo identificado o sistema *AML* como o sistema mais viável para este projeto, procedeu-se assim à criação do vocabulário controlado para o projeto IVISSEM, que tal como referido anteriormente, é resultado do alinhamento e junção dos três vocabulários controlados relativos a áreas científicas (*EuroSciVoc*, *Domínios Científicos e Áreas Científicas da FCT* e *UNESCO Nomenclature for Fields of Science and Technology*).

Para tal, foram utilizados os alinhamentos fornecidos pelo sistema de alinhamento *AML*. Uma vez que o número de alinhamentos era reduzido, foram também tidos em conta os alinhamentos fornecidos pelo sistema *LogMap* que não foram encontrados pelo sistema *AML* e que foram aceitos pelos especialistas. Caso o número de correspondências fosse superior, tal não seria possível, pelo que seriam considerados os alinhamentos resultantes do sistema *AML*.

Posto isto, recorrendo ao SKOS, foi então criado o vocabulário controlado para o projeto IVISSEM. No que concerne aos alinhamentos, estes foram transpostos para o vocabulário através da propriedade *skos:related*, permitindo assim ligar os termos que constam dos diferentes vocabulários.

O código do vocabulário gerado encontra-se disponível no RepositóriUM, <https://doi.org/10.34622/datarepositorium/T49TKX>. Na Figura 6 é possível observar a árvore do vocabulário de domínio produzido, onde que cada cor diz respeito a uma correspondência.

Figura 6. Árvore do vocabulário de domínio produzido<sup>2</sup>



Fonte: Autoria própria

## 5 CONCLUSÕES E TRABALHOS FUTUROS

A importância das ontologias como meio de implementar a Web Semântica é um facto. Da mesma forma, há a necessidade de técnicas capazes de combater a heterogeneidade e promover a interoperabilidade entre diferentes ontologias. Posto isto, o alinhamento de ontologias através de técnicas automáticas ou semiautomáticas apresenta um papel de relevo neste tópico.

Resultado do crescente número de sistemas capazes de alinhar ontologias, a OAEI surge como uma iniciativa internacional coordenada e criada para avaliar e comparar abertamente sistemas e algoritmos de alinhamento de ontologias, permitindo assim que qualquer pessoa seja capaz de tirar conclusões sobre as melhores estratégias. Foram registados resultados bastante satisfatórios em todas as *tracks* da OAEI, à exceção da *Complex Matching track*, que devido à sua complexidade e necessidade de extremas mudanças dos sistemas, esta apresenta resultados menos positivos. De todos os sistemas que participaram da OAEI em 2018 e/ou 2019, destacam-se o *AML* e *LogMap*, por serem dos poucos sistemas que participaram em quase todas as *tracks* e que obtiveram resultados bastante satisfatórios.

Não obstante ao facto de os sistemas serem desenvolvidos para alinhar ficheiros OWL, verificou-se que tanto o sistema *AML* como o *LogMap*, foram capazes de produzir alinhamentos satisfatórios entre ficheiros SKOS, sem necessidade de conversões.

Apesar dos bons resultados que as técnicas de alinhamento de ontologias têm vindo a demonstrar na OAEI, estas ainda podem ser melhoradas. Através das análises da OAEI ao longo

<sup>2</sup>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.

dos anos, é possível verificar que as técnicas de alinhamento têm vindo a demonstrar poucas melhorias em termos de qualidade. As principais causas apontadas para o pouco desenvolvimento das técnicas de alinhamento consistem na necessidade de novas tecnologias, ou na necessidade de os desenvolvedores das técnicas focarem-se na melhoria das suas técnicas, tentando alcançar melhores resultados nas *tracks* antigas e não em focarem em alcançar bons resultados em novas *tracks*.

Além disso, no que toca ao alinhamento de ontologias, a necessidade de conhecimento do domínio continua a ser uma realidade. Tal facto foi constatado neste trabalho, quando foi necessário recorrer a um conjunto de especialistas em Ciências da Computação e da Informação para avaliar as correspondências geradas.

A escassez de um alinhamento manualmente curado entre os três vocabulários utilizados, capaz de servir como um alinhamento de referência, constitui uma limitação na avaliação dos sistemas de alinhamento, dado que foi apenas utilizado um alinhamento de referência “*silver standard*” que tinha como base as correspondências geradas pelos sistemas e validadas pelos especialistas. Tal criação pode e deve ser vista como trabalho futuro.

Uma outra limitação presente neste trabalho deve-se ao facto de não ter havido disponibilidade de um número maior de especialistas que poderia ter promovido a execução do estudo Delphi com um maior número de rondas.

Com este estudo, é possível concluir que, atualmente, existem técnicas semiautomáticas que permitem de forma viável reaproveitar conhecimentos existentes na web através do alinhamento de ontologias e vocabulários controlados. Tais técnicas fazem uso de um conjunto de diferentes algoritmos, cada vez mais complexos, que permitem lidar com as ontologias de diversas formas (tendo em conta a semântica, estrutura etc.).

Para o projeto IVISSEM e eventualmente para outros projetos, este trabalho vêm permitir que, em vez de ser utilizado apenas um vocabulário, seja possível utilizar três integrados, possibilitando expandir *queries* e realizar inferências que anteriormente estavam vedadas. Este pode ser o primeiro passo para integrar a totalidade destes três vocabulários e, eventualmente, incluir outros. Este trabalho pode contribuir para trabalhos futuros através da identificação do sistema mais adequado para os integrar e, adicionalmente, permitiu desenvolver uma prova de conceito que é o vocabulário final. Vale destacar o alinhamento resultante entre o vocabulário utilizado a nível nacional pela FCT (Portugal) com um empregue pela União Europeia (*EuroSciVoc*) e outro do domínio da Ciência e Tecnologia (UNESCO), estabelecendo relações entre eles proveitosas também para outras universidades e projetos, bem como para a própria FCT.

Posto isto, como trabalho futuro, além das melhorias dos sistemas e da criação de um alinhamento manualmente curado entre os três vocabulários, também deverá ser explorado o alinhamento dos vocabulários tendo em conta outras áreas para além da Ciência da Computação.

## REFERÊNCIAS

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.

Disponível em: <https://doi.org/10.2478/JDIS-2019-0001>. Acesso em: 15 ago. 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. Disponível em: [http://ceur-ws.org/Vol-2288/oeai18\\_paper0.pdf](http://ceur-ws.org/Vol-2288/oeai18_paper0.pdf). Acesso em: 15 ago. 2021.

ALGERGAWY, A. **Results of the ontology alignment evaluation initiative** 2019. p. 46–85, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper0.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper0.pdf). Acesso em: 20 fev. 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. Disponível em: <http://hdl.handle.net/10201/56661>. Acesso em: 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. Disponível em: <https://doi.org/10.1590/s1413-99362007000200011>. Acesso em: 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. Disponível em: <https://doi.org/10.1016/J.TECHFORE.2020.120484>. Acesso em: 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. Disponível em: <https://www.scientificamerican.com/article/the-semantic-web/>. Acesso em: 10 set. 2021.

CHANG, F. **FCAMap-KG results for OAEI 2019**. p. 138–145, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper8.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper8.pdf). Acesso em: 20 fev. 2021.

CHEATHAM, M.; HITZLER, P. String similarity metrics for ontology alignment. **Lecture Notes in Computer Science: 8219 LNCS(PART 2)**, p. 294–309, 2013. Disponível em: [https://doi.org/10.1007/978-3-642-41338-4\\_19](https://doi.org/10.1007/978-3-642-41338-4_19). Acesso em: 20 fev. 2021.

CHEN, G. **FCAMapX results for OAEI 2018**, p. 160–166, 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaei18\\_paper7.pdf](http://ceur-ws.org/Vol-2288/oaei18_paper7.pdf). Acesso em: 20 fev. 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. Disponível em: <https://doi.org/10.22277/RGO.V6I3.1386>. Acesso em: 15 ago. 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. Disponível em: <https://doi.org/10.1177/002188637601200414>. Acesso em: 10 nov. 2021.

DESTRO, J. M. **EVOCROS: results for OAEI 2018**, p. 152–159, 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaei18\\_paper6.pdf](http://ceur-ws.org/Vol-2288/oaei18_paper6.pdf). Acesso em: 20 fev. 2021.

DESTRO, J. M. **EVOCROS: Results for OAEI 2019**, p. 131–137, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper7.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper7.pdf). Acesso em: 20 fev. 2021.

DJEDDI, W. E. **XMap: results for OAEI 2018**, p. 210–215, 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaei18\\_paper15.pdf](http://ceur-ws.org/Vol-2288/oaei18_paper15.pdf). Acesso em: 20 fev. 2021.

DSI. **lod:linked\_open\_data [DSI Wiki]**. Disponível em: [http://wiki.dsi.uminho.pt/doku.php?id=lod:linked\\_open\\_data#linked\\_open\\_data](http://wiki.dsi.uminho.pt/doku.php?id=lod:linked_open_data#linked_open_data). Acesso em: 30 out. 2021. Acesso em: 20 fev. 2021.

ENCYCLOPEDIA. **CORDIS**: European Commission. FP7-NMP - Specific Programme “Cooperation”: Nanosciences, Nanotechnologies, Materials and New Production Technologies. 2014. Disponível em: <https://cordis.europa.eu/en>. Acesso em: 15 ago. 2021.

EUZENAT, J.; SHVAIKO, P. Ontology matching: Second edition. *In*: \_\_\_\_\_. **Ontology matching**. 2. ed. Berlin: Springer-Verlag Berlin Heidelberg, 2013. Disponível em: <https://doi.org/10.1007/978-3-642-38721-0>. Acesso em: 20 fev. 2021.

FARIA, D. **Results of AML participation in OAEI 2018**, p. 125–131. 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaei18\\_paper2.pdf](http://ceur-ws.org/Vol-2288/oaei18_paper2.pdf). Acesso em: 20 fev. 2021.

FARIA, D. **AML and AMLC Results for OAEI 2019**, p. 101–106. 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper3.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper3.pdf). Acesso em: 20 fev. 2021.

FARIA, D. *et al.* The AgreementMakerLight ontology matching system. **Lecture Notes in Computer Science 8185 LNCS**, p. 527–541, 2013. Disponível em: [https://doi.org/10.1007/978-3-642-41030-7\\_38](https://doi.org/10.1007/978-3-642-41030-7_38). Acesso em: 20 fev. 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. Disponível em: <https://doi.org/10.2105/AJPH.74.9.979>. Acesso em: 10 nov. 2021.

| 22

FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA. **Fundação para a Ciência e a Tecnologia**. Disponível em: <https://www.fct.pt/>. Acesso em 30 mar. 2021.

GHERBI, S. **ONTMAT1**: results for OAEI 2019, p. 164–168. 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper12.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper12.pdf). Acesso em: 20 fev. 2021.

GRACIA, J.; MENA, E. Semantic heterogeneity issues on the web. **IEEE Internet Computing**, v. 16, n. 5, p. 60–67, 2012. Disponível em: <https://doi.org/10.1109/MIC.2012.116>. Acesso em: 20 fev. 2021.

GUARINO, N.; OBERLE, D.; STAAB, S. **Handbook on ontologies**. Geneve: Springer, 2009. Disponível em: <https://doi.org/10.1007/978-3-540-92673-3>. Acesso em: 5 fev. 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. **DOMÉ results for OAEI 2019**, p. 123–130. 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper6.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper6.pdf). Acesso em: 20 fev. 2021.

HERTLING, S. **DOMÉ results for OAEI 2018**, p. 144–151. 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaei18\\_paper5.pdf](http://ceur-ws.org/Vol-2288/oaei18_paper5.pdf). Acesso em: 20 fev. 2021.

IVISSEM . **Information Visualization & Social Scholarly Metric**. (n.d.). Disponível em: <http://www.ivissem.net/>. Acesso em: 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. Disponível em: <https://doi.org/10.1002/bult.283>. Acesso em: 5 fev. 2021.

JIMÉNEZ-RUIZ, E. **LogMap**: family participation in the OAEI 2018, p. 187–191, 2018. [http://ceur-ws.org/Vol-2288/oeai18\\_paper11.pdf](http://ceur-ws.org/Vol-2288/oeai18_paper11.pdf). Acesso em: 20 fev. 2021.

JIMÉNEZ-RUIZ, E. **LogMap**: family participation in the OAEI 2019, p. 160–163. 2019. Disponível em: Disponível em: [http://ceur-ws.org/Vol-2536/oeai19\\_paper11.pdf](http://ceur-ws.org/Vol-2536/oeai19_paper11.pdf). Acesso em: 20 fev. 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. Disponível em: [https://doi.org/10.1007/978-3-642-25073-6\\_18](https://doi.org/10.1007/978-3-642-25073-6_18). Acesso em: 20 fev. 2021.

KACHROUDI, M. **KEPLER at OAEI 2018**, p. 173–178. 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oeai18\\_paper9.pdf](http://ceur-ws.org/Vol-2288/oeai18_paper9.pdf). Acesso em: 20 fev. 2021.

KALIBATIENE, D.; VASILECAS, O. Survey on ontology languages. **Lecture Notes in Business Information Processing 90 LNBIP**, p. 124–141, 2011. Disponível em: [https://doi.org/10.1007/978-3-642-24511-4\\_10](https://doi.org/10.1007/978-3-642-24511-4_10). Acesso em: 18 set. 2021.

LAADHAR, A. **OAEI 2018 results of POMap++**, p. 192–196. 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oeai18\\_paper12.pdf](http://ceur-ws.org/Vol-2288/oeai18_paper12.pdf). Acesso em: 11 set. 2021.

LAADHAR, A. **POMap++ Results for OAEI 2019**: Fully Automated Machine Learning Approach for Ontology Matching. p. 169–174, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oeai19\\_paper13.pdf](http://ceur-ws.org/Vol-2536/oeai19_paper13.pdf). Acesso em: 18 set. 2021.

LINSTONE, H. A.; TUROFF, M.; HELMER, O. (ed.). **The Delphi method**. [S.l.]: Murray Turoff and Harold A. Linstone, 2002. Disponível em: <https://bit.ly/3RLQ6kF>. Acesso em: 18 set. 2021.

LÜTKE, A. AnyGraphMatcher Submission to the OAEI: knowledge graph challenge 2019. **CEUR**, v. 2536, p. 86–93, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oeai19\\_paper1.pdf](http://ceur-ws.org/Vol-2536/oeai19_paper1.pdf). Acesso em: 18 set. 2021.

MANIRAJ, V.; SIVAKUMAR, R. Ontology Languages: a review. **International Journal of Computer Theory and Engineering**, v. 2, n. 6, p. 1793-8201, 2010. Disponível em: <https://doi.org/10.7763/IJCTE.2010.V2.257>. Acesso em: 18 set. 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. Disponível em: <https://doi.org/10.12707/RIV16049>. Acesso em: 18 set. 2021.

MOHAMMADI, M. **SANOM results for OAEI 2018**, p. 205–209. 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oeai18\\_paper14.pdf](http://ceur-ws.org/Vol-2288/oeai18_paper14.pdf). Acesso em: 18 set. 2021.

MOHAMMADI, M. **SANOM results for OAEI 2019**, p. 175–180, 2018. Disponível em: [http://ceur-ws.org/Vol-2536/oeai19\\_paper14.pdf](http://ceur-ws.org/Vol-2536/oeai19_paper14.pdf). Acesso em: 18 set. 2021.

Moher, D. *et al.* Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. **PLoS Medicine**, v. 6, n. 7, 2009. Disponível em: <https://doi.org/10.1371/journal.pmed.1000097>. Acesso em: 18 set. 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. Disponível em: <https://doi.org/10.1007/S11192-015-1765-5>. Acesso em: 18 set. 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. Disponível em: <https://doi.org/10.1353/RHE.1995.0008>. Acesso em: 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. Disponível em: <https://doi.org/10.1016/J.IM.2003.11.002>. Acesso em: 10 nov. 2021.

PASTOR-SÁNCHEZ, J.A. *et al.* Publicación como Linked Open Data de la nomenclatura internacional de Ciencia y Tecnología y del Tesouro UNESCO. In: CONGRESSO ISKO ESPANHA E PORTUGAL, 1., 7-9 November 2013. **Anais do [...]**. Porto: ISKO, 2013. Disponível em: <http://eprints.rclis.org/24272/>. Acesso em: ago. 2021.

PORTISCH, J. **ALOD2Vec matcher**, p. 132–137, 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaie18\\_paper3.pdf](http://ceur-ws.org/Vol-2288/oaie18_paper3.pdf). Acesso em: 18 set. 2021.

PORTISCH, J. **Wiktionary Matcher**, p. 181–188, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaie19\\_paper15.pdf](http://ceur-ws.org/Vol-2536/oaie19_paper15.pdf). Acesso em: 18 set. 2021.

PORTUGAL. Ministério da Educação e Ciência. Fundação para a Ciência e a Tecnologia. **Domínios científicos e áreas científicas**. [Lisboa]: [FCT], 2012. Disponível em: [https://www.fct.pt/apoios/projectos/concursos/2012/docs/Dominios\\_e\\_Areas\\_Cientificas\\_C2012.pdf](https://www.fct.pt/apoios/projectos/concursos/2012/docs/Dominios_e_Areas_Cientificas_C2012.pdf). Acesso em: 18 set. 2021.

PUBLICATIONS OFFICE OF THE EU. European Science Vocabulary (EuroSciVoc). v.1.3 Luxembourg: Publications Office of the European Union, 2021. Disponível em: <http://publications.europa.eu/resource/dataset/euroscivoc>. Acesso em: 18 set. 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. Disponível em: <https://doi.org/10.1111/J.1745-3992.2005.00006.X>. Acesso em: 18 set. 2021.

ROUSSILLE, P. **Holontology**: results of the 2018 OAEI evaluation campaign, p. 167–172, 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaie18\\_paper8.pdf](http://ceur-ws.org/Vol-2288/oaie18_paper8.pdf). Acesso em: 18 set. 2021.

SILVA, J. **ALIN results for OAEI 2018**, p. 117–124, 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaie18\\_paper1.pdf](http://ceur-ws.org/Vol-2288/oaie18_paper1.pdf). Acesso em: 18 set. 2021.

SILVA, J. **ALIN results for OAEI 2019**, p. 94–100, 2019. [http://ceur-ws.org/Vol-2536/oaie19\\_paper2.pdf](http://ceur-ws.org/Vol-2536/oaie19_paper2.pdf). Acesso em: 18 set. 2021.

SKOS Play! - **Thesaurus & taxonomies**. (*n.d.*). Disponível em: <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. Disponível em: <https://skos.um.es/unesco6/>. Acesso em: 18 set. 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. Disponível em: <https://doi.org/10.1371/JOURNAL.PONE.0188348>. Acesso em: 18 set. 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. Disponível em: [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6). Acesso em: 18 set. 2021.

TANG, Y. **Lily results for OAEI 2018**, p. 179–186, 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaei18\\_paper10.pdf](http://ceur-ws.org/Vol-2288/oaei18_paper10.pdf). Acesso em: 18 set. 2021.

THIÉBLIN, É. **CANARD complex matching system**: results of the 2018 OAEI evaluation campaign, p. 138–143, 2018. Disponível em: [http://ceur-ws.org/Vol-2288/oaei18\\_paper4.pdf](http://ceur-ws.org/Vol-2288/oaei18_paper4.pdf). Acesso em: 18 set. 2021.

THIÉBLIN, É. **CANARD Complex Matching System**: results of the 2019 OAEI Evaluation Campaign, p. 114–122, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper5.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper5.pdf). Acesso em: 18 set. 2021.

WANG, X. **FTRLIM results for OAEI 2019**, p. 146–152, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper9.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper9.pdf). Acesso em: 18 set. 2021.

ZHOU, L. **AROA results for 2019 OAEI**, p. 107–113, 2019. Disponível em: [http://ceur-ws.org/Vol-2536/oaei19\\_paper4.pdf](http://ceur-ws.org/Vol-2536/oaei19_paper4.pdf). Acesso em: 18 set. 2021.