Neighbourhood-based Cross-Language Ontology Matching

Juliana M. Destro [a,*], Gabriel de Oliveira dos Santos [a], Julio Cesar dos Reis [a],
Ariadne Maria Brito Rizzoni Carvalho [a], Ivan Ricarte [c] and Ricardo da S. Torres [b]
[a] *Institute of Computing, University of Campinas, SP, Brazil*
E-mails: juliana.destro@ic.unicamp.br, mailgabriel@ic.unicamp.br, jreis@ic.unicamp.br, ariadne@ic.unicamp.br
[b] *Department of ICT and Natural Sciences, Norwegian University of Science and Technology, Ålesund, Norway*
E-mail: ricardo.torres@ntnu.no
[c] *School of Technology, University of Campinas, Limeira-SP, Brazil*
E-mail: ricarte@unicamp.br

Abstract. Cross-language ontology alignments play a key role for the semantic integration of data described in different languages. The task of automatically identifying ontology mappings in this context requires exploring similarity measures as well as ontology structural information. Such measures compute the degree of relatedness between two given terms from ontology's entities. The structural information in the ontologies may provide valuable insights about the concept alignments. Although the literature has extensively studied these measures for monolingual ontology alignments, the use of similarity measures and structural information for the creation of cross-language ontology mappings still requires further research. In this article, we define a novel technique for automatic cross-language ontology matching based on the combination of a composed similarity approach with the analysis of neighbour concepts to improve the effectiveness of the alignment results. Our composed similarity considers lexical, semantic and structural aspects based on background knowledge to calculate the degree of similarity between contents of ontology entities in different languages. Experimental results with MultiFarm indicate a good effectiveness of our approach including neighbour concepts for mapping identification.

Keywords: ontology mapping, ontology cross-language alignment

1. Introduction

Ontologies[1] are used on a multitude of applications in computer science in the role of a specification mechanism or definition of a common vocabulary. Mapping establishes correspondences between different ontology entities and are relevant for the integration of heterogeneous data sources. There is a growing number of ontologies described in different natural languages. The challenge of generating correspondences between different ontologies, created for diversified purposes, is aggravated when concepts are labeled in different natural languages, even in the same domain. Although automatic monolingual ontology matching has been extensively investigated [1], cross-language ontology matching still demands further investigations aiming to automatically identify correspondences between ontologies described in different languages [2].

In this context, accurate automatic methods are essential for ensuring the quality of the generated mappings. Current ontologies have highly grown in size. As differences between the used alphabets hamper the use of simple string comparison techniques, similarity measures play a key role to obtain well-defined ontology mappings because they allow calculating the level of lexical and semantic similarity between concepts [3]. Cross-language ontology matching approaches in the literature have not yet thoroughly investigated the

*Corresponding author. E-mail: juliana.destro@ic.unicamp.br.
[1] *What is an ontology?* http://www-ksl.stanford.edu/kst/what-is-an-ontology.html (As of April 2019).

influence of similarity calculation neither have they analyzed the influence of neighbour concepts in the matching process.

In this article, we propose an original cross-language ontology alignment technique based on the analysis of neighbour concepts relying on composed similarity measure, by combining both syntactic and semantic similarity techniques. Syntactic similarity computes a score calculated based on string analysis (extracted from labels of entities), whereas the semantic similarity is computed taking into account background knowledge, such as synonyms and the context in which terms appear (*e.g.*, use of external dictionaries and vocabularies). Our investigation explores a *Weighted Overlap* measure [4] relying on the neutral-domain semantic network *BabelNet* [5] and computes a weighted mean of semantic and syntactic similarities. The proposed technique also takes into account the similarity of those concepts immediately related to a given entity (the neighbours), both on source and target ontologies. The method finds the highest value of similarities among these concepts. In this investigation, we name such value as neighbourhood similarity. The neighbourhood similarity is used to improve the correctness of mappings and it is thus combined with the composed similarity whenever the initial value of composed similarity is in a doubtful range, that is, between a default and minimum threshold (set as parameters before the processing begins).

We carried out a series of experiments to investigate the quality of mappings generated by our technique. Our experiments explored conference-domain ontologies in 45 language pairs from the *MultiFarm*[2] dataset [6]. *MultiFarm* provides curated mappings between multilanguage ontologies. This dataset has been extensively used to assess cross-language ontology matching methods. The obtained results indicate that syntactic and semantic similarities may have different weights in order to obtain a good accuracy. Our experiments suggest that the threshold, language in which the ontologies are described and translation tool play an important role in the quality of generated alignments.

The remaining of this paper is organized as follows: Section 2 describes the related work; Section 3 formalizes the fundamental concepts of our proposal; Section 4 reports on our proposed technique; Section 5 describes the experimental results whereas Section 6 discusses our findings; Section 7 provides the conclusion remarks.

[2]https://www.irit.fr/recherches/MELODI/multifarm (As of April 2019).

2. Background

There has been a number of investigations on specific aspects of cross-language for ontology matching. Meilicke *et al.* [7] studied the effectiveness of a set of matching systems based on a dataset defined to evaluate ontology alignment. Their results indicated the difficulties of traditional ontology matching algorithms for carrying out multilingual ontology alignment. Trojahn *et al.* [8] described an extensive survey of matching systems and strategies for accomplishing multilingual and cross-language ontology matching. More recently, Ivanova [2] provided a classification of the available approaches and strategies used by current cross-language mapping systems.

Several approaches have explored the translation effects and the use of a third language in cross-language ontology alignment. In particular, Fu *et al.* [9] analyzed the impact of automatic translations on multilingual ontology alignment, highlighting the translation's relevance for achieving adequate matching quality. Spohr *et al.* [10] studied the translation of concept labels to a third language for matching two ontologies described in different languages.

Ontology alignment techniques have considered the use of similarity methods, which aim to calculate the degree of relatedness between concepts exploring different sources (*e.g.* dictionary, thesauri). Annane *et al.* [11] proposed a method to build a customized background knowledge resource to improve recall of generated mappings without sacrificing precision. Stoutenburg [12] argued that the use of ontologies combined with linguistic resources as background knowledge might enhance ontology matching processes. This appears as an alternative to syntactic similarity measures relying only on string comparison to determine the similarity value.

The use of multiple similarity measures for the ontology alignment task has been investigated in the literature. Nguyen and Conrad [13] proposed an ontology matching method based on the combination of lexical-based, structure-based, and semantic-based techniques. After obtaining the structural correspondences among the concepts, the method explores a semantic similarity based on WordNet dictionary and the results are combined. Their approach was evaluated with monolingual ontology alignments. Further investigations are necessary to understand whether a combination and use of semantic similarity can be relevant for cross-language ontology alignment.

Experimental studies have analyzed the influence of syntactic and semantic similarity methods and the structure of terms denoting concepts in ontologies in the context of cross-language alignment [14]. These studies highlight the potential influence of similarity measures.

Structural information (*i.e.* neighbourhood) and multiple similarity measures were used by Essayeh and Abed [15] to generate a similarity matrix. Lin *et al.* [16] also used structural information combined with other similarity methods but neither tackled the cross-language ontology alignment problem.

We investigated the most commonly used approaches for existing cross-language ontology matching methods in participants of *OAEI* (Ontology Alignment Evaluation Initiative)[3] *MultiFarm* track:

– **Translation-based:** the translator is used to overcome the natural language barrier and to enable applying monolingual methods to perform cross-language matching. This approach consists of translating the labels and other elements of concepts (such as synonyms) to the same language of the other ontology, or to a pivot language. Several systems use such approach with English as a pivot language, such as *CroLOM* [17], *AML* [18], and *YAM++* [19].
  The proposal of *CroLOM* is based on natural languages processing techniques (such as lemmatization, stopwords elimination and stemming) to normalize labels extracted from ontologies. These entities are translated into English, as a pivot language, and the technique computes a Cartesian product among the concepts that compose the ontologies. They apply semantic and syntactic similarity measures in a hybrid way to identify potential mappings. The syntactic similarity is calculated from the *Levenshtein distance* [20], whereas the semantic similarity considers the category of words in WordNet. At this stage, an initial filter is applied to select candidate correspondences containing the maximum similarity value. Then, a second filter is applied to identify the correspondences that contain similarity value greater than a given threshold.
  The *SOCOM++* [21] approach considers several setups with different parameters. In contrast to *CroLOM*, it translates concept labels of the source ontology to the same language of the target ontology, thus no pivot languages are considered. Afterwards, both ontologies are described in the same language and monolingual matching methods are applied. In this process, the context of a given concept is analyzed considering all immediate neighbour concepts to improve the quality of the obtained alignment. This approach was designed to support user's influence on adjustments in the translation of the selected labels, and thus users can analyze the resulting mappings and propose changes.
  The *AML (AgreementMakerLight)* is a general purpose ontology matching system based on the design principles of *AgreementMaker* [22]. *AML* relies primarily on lexical matching and structural algorithms for both matching and filtering. It makes use of external biomedical ontologies and the WordNet as sources of background knowledge.
  In *YAM++*, concept labels of both ontologies (source and target) are translated into the English language. The concepts are filtered in a stage named candidate filtering. In this stage, heuristic filters are applied to selected candidate correspondences, reducing the search space. In the following stage, the method analyzes the neighbourhood of previously selected concepts to discover as many as possible high accurate mappings. Finally, the selected mappings go through a process of semantic verification [23], in which those correspondences considered inconsistent are removed.

– **Information retrieval**: this approach explores information retrieval techniques, for instance, PageRank and indexing, to define matchings. *LogMap* [24] and *KEPLER* [25] explores such techniques.
  LogMap considers a Lexical indexing, which is an inverted index used to store the lexical information. It exploits ontology modularization techniques to reduce the size of the problem. The relevant modules in the input ontologies together with (a subset of) the candidate mappings are encoded in *LogMap* using a Horn clause propositional representation. This approach extends Dowling-GallierâĂŹs algorithm [26] to track all mappings that may be involved in the unsatisfiability of a

[3]http://oaei.ontologymatching.org (As of April 2019).

class and performs a greedy local repair; that is, it repairs unsatisfiabilities on-the-fly and only looks for the first available repair plan. It considers a Semantic Indexing, which allows to answer many entailment queries as an index lookup operation over the input ontologies and the mappings computed. The semantic index complements the use of the propositional encoding to detect and repair unsatisfiable classes in the input ontologies.

The *KEPLER*'s approach relies on divide and conquer strategy; first it splits up the ontology into small blocks, maximizing the relationship inside the block, and minimizing the relationship between the blocks themselves. On the following step, it translates the ontologies to English as the pivot language, and uses the indexing strategy to reduce the searching space. It considers Candidate Mappings Identification, which queries documents in a vector space that contains a set of ontological entities and their synonyms obtained via WordNet for each ontology. Finally, the algorithm filters the candidate mappings by using two filters: the first filter eliminates the redundancy between these candidates by eliminating possible duplicates; the second filter eliminates false positive candidates.

Our approach differs from the above-mentioned proposals because we combine both semantic and syntactic similarities by computing the composed similarity assigning weights to each similarity measure. In addition, we define a similarity value to the neighbourhood with the aim of improving the correctness of the generated mappings.

3. Formalization

This section formalizes the fundamental concepts in this investigation.

3.1. Ontologies

Ontologies define a common vocabulary in a domain [27]. They are used for semantic representation in computational systems, describing the definition of concepts and the relationship among them.

Definition 3.1 (Ontology)**.** *An ontology $\mathcal{O}$ describes a domain in terms of concepts, attributes and relationships. Formally, an ontology $\mathcal{O} = (\mathcal{C}_\mathcal{O}, \mathcal{R}, \mathcal{A}_\mathcal{O})$ consists in a set of classes or concepts $\mathcal{C}_\mathcal{O}$ interrelated by a set of directed relations $\mathcal{R}$. Each concept $c \in \mathcal{C}_\mathcal{O}$ has a unique identifier and it is associated with a set of attributes $\mathcal{A}_\mathcal{O}(c) = \{a_1, a_2, ..., a_p\}$. Concepts are ontology entities represented by owl:Class construct in OWL[4]. Each relation $r(c_1, c_2) \in \mathcal{R}$ can be described as a tuple $(c_1, c_2, r(c_1, c_2))$, where $r(c_1, c_2)$ is a function returning the type of relationship between (c_1, c_2) (e.g., "$\equiv$", "$\sqsubseteq$", etc.). The symbols "$\equiv$" and "$\sqsubseteq$" represent relationships "equivalence" and "is-a", respectively. Furthermore, the relationships can express domain-related relations. For instance, considering the biomedical domain, the concepts c_1 :"Insulin" and c_2 :"Diabetes" may be related by the following function: $r(c_1, c_2)$ = "Treats". Relations are entities represented by owl:ObjectProperty or owl:DatatypeProperty constructs in OWL [28].*

Definition 3.2 (Neighbour Concepts)**.** *We define neighbour concepts of a given entity $e \in \mathcal{C}_\mathcal{O}$ or $e \in \mathcal{R}$ the set of concepts with a direct relation to e. Formally, the neighbourhood of e is the set $nbh = \{cpt | cpt \in \mathcal{C}_\mathcal{O} \wedge dist(e, cpt) = 1\}$, where $dist(e, cpt)$ is the distance (in terms of the number of edges) between 'e' and 'cpt'.*

Figure 1 presents an illustrative example of neighbour concepts. The neighbourhood of "*Pancreas*" is composed of "*Endocrine System*", "*Digestive System*", "*Insulin*" and "*Glucagon*", because all of them are directly related to "*Pancreas*". Because the distance between "*Kidney*" and "*Pancreas*" is equal to two, it is not considered a neighbour concept of "*Pancreas*".

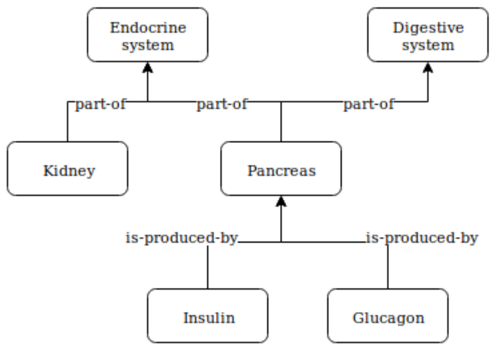

Fig. 1. Example[scale=0.50] of neighbourhood.

[4]The W3C Web Ontology Language (OWL) is a Semantic Web language https://www.w3.org/OWL/ (As of April 2019).

3.2. Cross-Language Ontology Alignment

The cross-language ontology problem is formally defined as follows:

Definition 3.3 (Cross-language ontology alignment)**.** *Let O_X and O_Y be ontologies described in different natural language "X" and "Y", respectively; and entities $e_i \in O_X$ and $e_j \in O_Y$. The problem relies on automatically identifying the adequate set of 4-tuples $m_{ei \to ej} = (e_i, e_j, s_{i,j}, r(e_i, e_j))$, where $s_{i,j}$ is the similarity value between (e_i, e_j) and falls under the interval [0,1] and $r(e_i, e_j) \in \mathcal{R}$ is the relationship between these elements. For instance, considering the concepts $c_1 \in \mathcal{C}_{\mathcal{O}pt}$ and $c_2 \in \mathcal{C}_{\mathcal{O}en}$, from ontologies described in Portuguese and English, respectively, such that c_1 = "Cabeça" and c_2 = "Head", the alignment between these concepts is $m_{c1 \to c2} = (c_1, c_2, 1, \equiv)$.*

Definition 3.4 (Mappings)**.** *The final result of the alignment process is a set containing the mappings found between the entities (classes, object properties and datatype properties) from two given ontologies. Formally, the mapping between the ontologies O_X and O_Y is given by each element of $\mathcal{M}_{\mathcal{O}_\mathcal{X} \to \mathcal{O}_\mathcal{Y}}(\lambda) = \{m_{ei \to ej} | e_i \in \mathcal{O}_\mathcal{X} \wedge e_j \in \mathcal{O}_\mathcal{Y} \wedge s(e_i, e_j) \geqslant \lambda\}$, where "$\lambda$" is the threshold (minimum value to consider similar) and $sim(e_i, e_j)$ is the similarity between a string denoting e_i and e_j.*

3.3. Similarity Measures

Definition 3.5 (Similarity between entities)**.** *Given two entities e_i and e_j from an ontology (or from different ontologies), the similarity value between them is defined as the maximum similarity value among the attributes (e.g. labels, synonyms, etc) of e_i and e_j. Formally:*

$$syn(e_i, e_j) = \arg\max \ sim(a_{ix}, a_{jy}) \tag{1}$$

where $sim(a_{ix}, a_{jy})$ is the relatedness degree between the pair of attributes a_{ix} and a_{jy} from e_i and e_j, respectively. The similarity may be calculated in different linguistic levels, from string-based methods to semantic techniques [29].

Syntactic Similarity Measure. Levenshtein Distance [20] is an algorithm that computes a syntactic or string-based similarity, which can be understood as the minimum number of single-character editions (insertions, deletions or substitutions) needed to change a string s into s'. This algorithm has been chosen to compute the syntactic similarity in this investigation because Levenshtein Distance has been well-studied and has been extensively used to spelling correction, being considered a good alternative to syntactic analysis [30].

Semantic Similarity Measure. Semantic similarity between concepts is a metric to evaluate how similar two given concepts are, considering their meanings in a certain context. For instance, the words "lead" and "iron" are much more similar considering the metal context than "lead" and "leader". On the other hand, when we consider the organizational context "lead" and "leader" may be more similar than "lead" and "iron".

There are algorithms to calculate semantic similarity. Usually, these algorithms explore an external resource such as vocabulary, dictionaries, and thesauri. In this work, we use Weighted Overlap applied to NASARI vectors, together with the neutral-domain semantic network *BabelNet* [5].

This choice relies on the studies of the influence of semantic similarity in neutral-domain context, using Weighted Overlap [14].

NASARI helps us to compute the similarity value in multilingual contexts because it uses vectors based on "*synsets*" (set of synonyms) used by *Babelnet* [31]. The vectors are created in two steps: first, for a given concept, it collects a set of Wikipedia pages where the concept is mentioned. The second step consists in processing the collected contextual information using a statistical measure (lexical specificity [32]), aiming at finding the most relevant words and synsets appearing in the contextual information and assigning to each one of them a weight (based on the statistical measure). Each of the these words and synsets are used as dimensions in the vector-based representation.

Table 1 shows the semantic vector-based representation of two *Babel synsets* (*i.e.*, the identification used in BabelNet to represent a given meaning of a word and all the synonyms expressing that meaning in a range of different languages). On each row of the NASARI vector table (exemplified by two rows in Table 1), the first column is the *Babel synsets* ID and the second column is the textual description of the synset (*e.g.*, the synsetID bn:00000009n represents the synset "*100 (number)*"). The vector dimensions are described from column three onwards, and are represented by a *Babelnet synset* ID and its correspondent weight (*e.g.*, vector dimension in column *synset1_weight1*, where

Table 1: Example of NASARI vector representation, where synset_weight represents dimensions $1 : n \wedge n <= 300$.

Babelnet SynsetId	Wikipedia PageTitle	synset1_weight1	...	synset$n$_weight$n$
bn:00000009n	100 (number)	bn:00058285n_332.33	...	bn:00031261n_9.35
bn:00000010n	1000 (number)	bn:00058285n_347.11	...	bn:00024261n_2.11

bn:00058285n is the dimension and 332.33 is the weight). Vectors are truncated to the non-zero dimensions only (*i.e.*, all dimensions present weight above zero). Because vectors present *Babelnet synset* as their dimensions, they are comparable across languages.

NASARI leverages *Weighted Overlap (WO)* method applied to the semantic vectors representations [33] to calculate the semantic similarity between two elements e_1 and e_2 (*cf.* Equation (2)):

$$sem(e_1, e_2) = WO(v_1, v_2) \tag{2}$$

Weighted Overlap calculates the similarity between the meanings of two given lexical items. Formally:

$$WO(v_1, v_2) = \frac{\sum_{i=1}^{|S|} (r_i^1 + r_i^2)^{-1}}{\sum_{i=1}^{|S|} (2i)^{-1}} \tag{3}$$

In Equation 3, S refers to the set of overlapping dimensions between the two vectors (*i.e.*, dimensions appearing on both vectors; in the example in Table 1, dimension bn:00058285n under column *synset1_weight1*). The r_q^j is the rank of dimension q in the vector v_j. Note that the weight is not used in WO equation; it is only used for ranking (*i.e.*, sorting) the dimensions.

Definition 3.6 (Composed Similarity)**.** *We define the composed similarity by combining syntactic and semantic measures. Let sem(e_1, e_2) (Equation (2)) be the semantic similarity and syn(e_1, e_2) the syntactic one between the entities e_1 and e_2, respectively. Formally:*

$$simC(e_1, e_2) = \frac{\alpha syn(e_1, e_2) + \beta sem(e_1, e_2)}{\alpha + \beta} \tag{4}$$

where α and β are constants.

Note that both semantic and syntactic similarities are a particular case of the composed similarity, when α and β are equal to zero, respectively.

We explore the composed similarity together with Neighbourhood Analysis (*cf.* Section 4) in our cross-language ontology alignment technique.

4. Cross-Language Ontology Alignment Relying on Neighbourhood Analysis

Our technique for cross-language ontology matching is based on a composed similarity measure relying on both syntactic and semantic similarity techniques, leveraging the similarity of local neighbour concepts (*cf.* Definition 3.2) to settle doubtful mappings.

Figure 2 presents the workflow of the proposed technique. The inputs are a source and target ontologies written in OWL (Web Ontology Language) format. These ontologies are converted to an object, preserving the relations and neighborhood relationship between concepts. Each entity of the source ontology is compared with all entities of the same type (*i.e.*, concepts are compared only with concepts, relations are compared only with relations) of the target ontology and their composed similarity is calculated. If the similarity surpasses the threshold, the pair is mapped. If not, the calculated similarity is compared with a minimum threshold to verify if the composed similarity is in a doubtful range. If the similarity value is above minimum a threshold, the similarity between the neighbour concepts is taken into account in a new validation against the threshold.

Algorithm 1 defines a cross-language alignment between two distinct ontologies $\mathcal{O}_\mathcal{X}$ and $\mathcal{O}_\mathcal{Y}$ expressed in different natural languages. The algorithm considers the following input arguments:

- Input ontologies $\mathcal{O}_X, \mathcal{O}_Y$
- $\lambda \in (0, 1]$ - default threshold
- $min_\lambda \in [0, \lambda)$ - minimum threshold
- α - Syntactic weight
- β - Semantic weight
- *pivot* - The pivot language

The algorithm starts with mapping set $\mathcal{M}_{\mathcal{O}_\mathcal{X} \to \mathcal{O}_\mathcal{Y}} \leftarrow \emptyset$ (line 1) and the similarity variables with zero. It calculates the cartesian product from the set of entities $\mathcal{C}_{\mathcal{O}_\mathcal{X}}$ and $\mathcal{C}_{\mathcal{O}_\mathcal{Y}}$, and $\mathcal{R}_{\mathcal{O}_\mathcal{X}}$ and $\mathcal{R}_{\mathcal{O}_\mathcal{Y}}$ from ontologies $\mathcal{O}_\mathcal{X}$ and $\mathcal{O}_\mathcal{Y}$, respectively. It considers automatic translation of labels of entities e_1 and e_2 to a pivot language, providing (w_1, w_2), where $w_1 = translated(e_1)$ and $w_2 = translated(e_2)$ (line 9), leveraging Google Translate API during runtime. The algorithm computes the similarity value based on a syntactic measure (line 10). The syntactic similarity is calculated relying on the strings (w_1, w_2). The semantic similarity value is also computed. To this end, for each tuple $(e_1, ling_{e_1}, e_2, ling_{e_2})$, composed of the entities e_1 and e_2, and their respective natu-

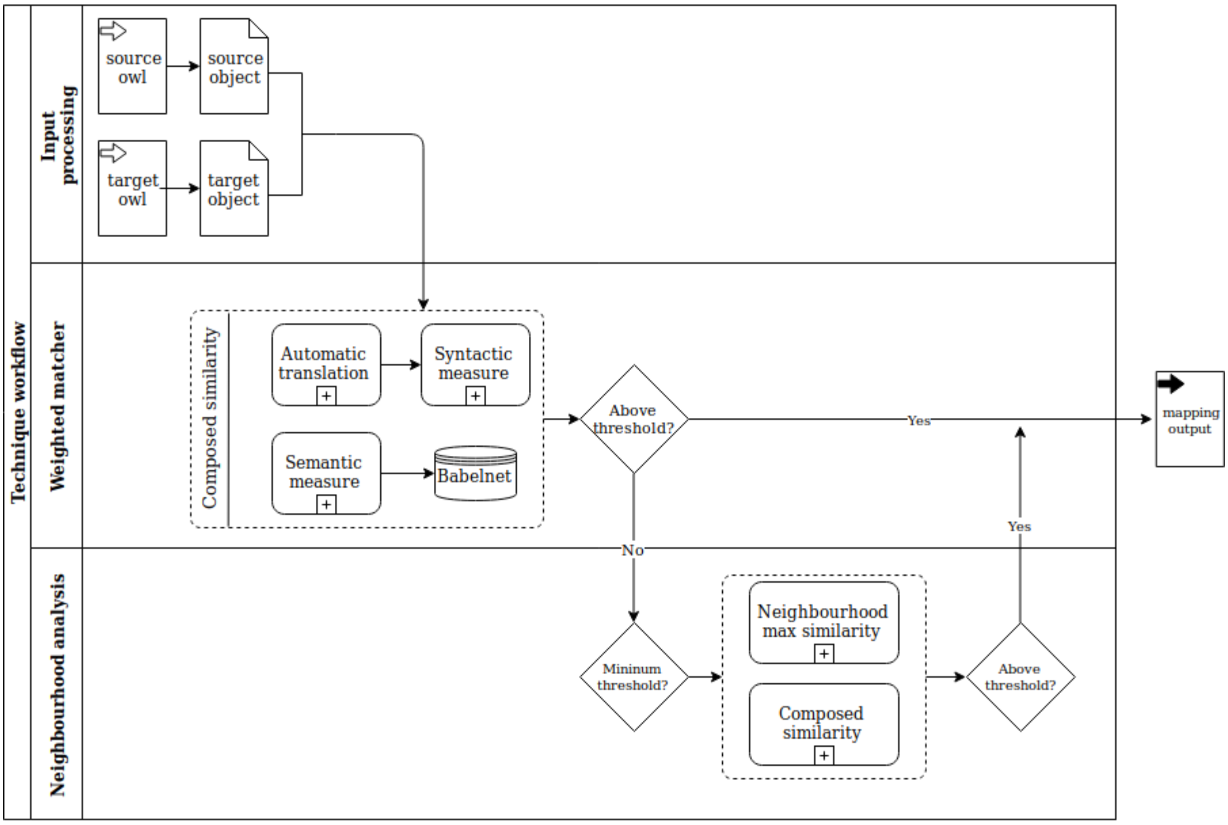

Fig. 2. Schematic overview of the proposed cross-language ontology alignment based on neighbourhood analysis.

ral languages $ling_{e_1}$ and $ling_{e_2}$, the algorithm calls the function $babelnet(e_1, ling_{e_1}, e_2, ling_{e_2})$ (line 13). This function uses *Babelnet synsets* and NASARI semantic vectors (*cf.* Section 3.3) to calculate the *Weighted Overlap* (Equation (3)).

The algorithm calculates the weighted average, assigning weights previously defined by α and β to the syntactic syn_{sim} and semantic sem_{sim} similarities, respectively. It results on the composed similarity $composed_{sim}$ (line 15). If the $composed_{sim}$ value is lower than the default threshold λ and is greater than or equal to min_{λ}, the similarity is considered to be a in a doubtful range and the algorithm verifies the neighbourhood of the involved concepts to ensure the quality of mappings.

The neighbourhood analysis computes the maximum similarity among the neighbour concepts of the considered entities (source and target). Algorithm 2 computes the similarity among the neighbours of the entities e_1 and e_2 (source and target entities given as input). Only concepts in the neighborhood are retrieved, for entities either in $\mathcal{C}_O$ or $\mathcal{R}$. First, it extracts the concepts neighboring e_1 and e_2 to nbh_1 and nbh_2, respectively (line 1 and 2 in Algorithm 2). The algorithm aims to find the pair of neighbour concepts (one from the source ontology and the other one from the target one) with the maximum similarity value based on the composed similarity measure.

Figure 3 presents an example to illustrate the technique of neighbourhood analysis. We consider two ontologies[5], O_X and O_Y, where O_X is described in Portuguese language and O_Y is described in English language.

[5]These ontologies are considered only for the purpose of this example. They were not extracted from real-world ontologies.

Algorithm 1: Cross-language ontology alignment based on composed similarity measure considering neighbourhood analysis

Require: $\mathcal{O}_X, \mathcal{O}_Y, \lambda, min_\lambda \in [0,1], \alpha, \beta, pivot$

$\mathcal{M}_{\mathcal{O}_\mathcal{X} \rightarrow \mathcal{O}_\mathcal{Y}} \leftarrow \emptyset$ {Initialize the mapping as an empty set}
$syn_{sim} \leftarrow 0$
$sem_{sim} \leftarrow 0$
$composed_{sim} \leftarrow 0$
$nbh_{sim} \leftarrow 0$
for all $e_1 \in \mathcal{O}_\mathcal{X}$ **do**
  for all $e_2 \in \mathcal{O}_\mathcal{Y}$ **do**
    if $\alpha > 0$ **then**
      $w_1 \leftarrow translate(e_1, \text{pivot})$, $w_2 \leftarrow translate(e_2, \text{pivot})$
      $syn_{sim} \leftarrow syntactic_{sim}(w_1, w_2)$
    end if
    if $\beta > 0$ **then**
      $sem_{sim} \leftarrow semantic_{sim}(e_1, ling_{e_1}, e_2, ling_{e_2})$
    end if
    $composed_{sim} = \frac{\alpha syn_{sim} + \beta sem_{sim}}{\alpha + \beta}$ {Compute the composed similarity value} {Analyze the neighbourhood of concepts if in a doubtful range}
    if $composed_{sim} < \lambda$ **and** $composed_{sim} \geqslant min_\lambda$ **then**
      $nbh_{sim} \leftarrow neighbourhood_{sim}(e_1, e_2)$ {Algorithm 2}
      $similarity \leftarrow composed_{sim}^{(1-nbh_{sim})}$
    else
      $similarity \leftarrow composed_{sim}$
    end if
    if $similarity \geqslant \lambda$ **then**
      $m_{e1 \rightarrow e2} \leftarrow (e_1, e_2, similarity, \equiv)$
      $\mathcal{M}_{\mathcal{O}_\mathcal{X} \rightarrow \mathcal{O}_\mathcal{Y}} \leftarrow \mathcal{M}_{\mathcal{O}_\mathcal{X} \rightarrow \mathcal{O}_\mathcal{Y}} \cup \{m_{e1 \rightarrow e2}\}$
    end if
  end for
end for
return $\mathcal{M}_{\mathcal{O}_\mathcal{X} \rightarrow \mathcal{O}_\mathcal{Y}}$ {Generated mappings}

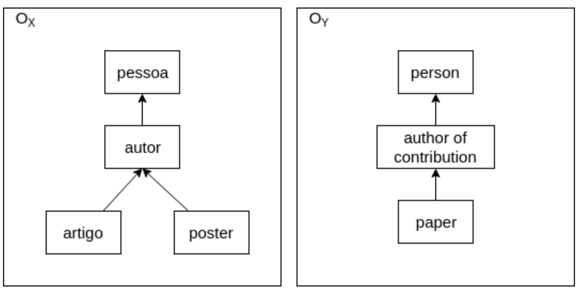

Fig. 3. Ontologies O_X and O_Y under analysis.

Algorithm 2: Neighbourhood analysis.

Require: e_1, e_2 {Given the entities e_1 and e_2 from the source and target ontologies respectively} {Extract the concept neighbourhood of entities e_1 and e_2 to nbh_1 and nbh_2}

$nbh_1 \leftarrow neighbourhood(e_1)$
$nbh_2 \leftarrow neighbourhood(e_2)$
$maxSim \leftarrow 0$
for all $n_1 \in nbh_1$ **do**
  for all $n_2 \in nbh_2$ **do**
    $sim \leftarrow composed_similarity(n_1, n_2)$ {Compute the composed similarity value}
    if $sim > maxSim$ **then**
      $maxSim \leftarrow sim$
    end if
  end for
end for
return $maxSim$

In the example of Figure 4, the entities "*autor*" and "*author of contribution*" are under analysis. We first calculate the syntactic and semantic similarity value between the two entities, and find a composed similarity value of 0.80.

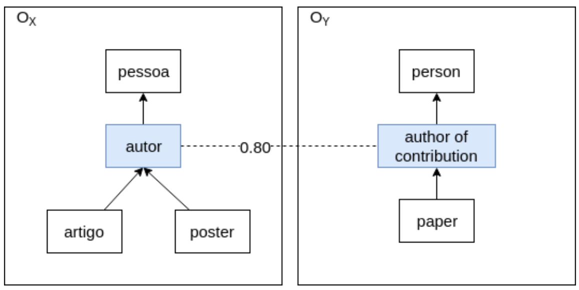

Fig. 4. Pair of entities considered for mapping.

In our example, the minimum threshold to consider a mapping doubtful is 0.33 and default threshold is 0.95. Thus, by our Algorithm 1, it is necessary go through the neighbourhood analysis. The neighbours of "*autor*" are {"*pessoa*", "*artigo*", "*poster*"}, and the neighbours of "*author of contribution*" are {"*person*", "*paper*"}. We apply a cartesian product to these sets to evaluate the composed similarity (*i.e.*, syntactic and semantic similarity combined) between the set of neighbours and retain maximum similarity value found. In this illustration, the maximum similarity is found for "*pessoa*" and "*person*", with a 1.0 measure as depicted in Figure 5.

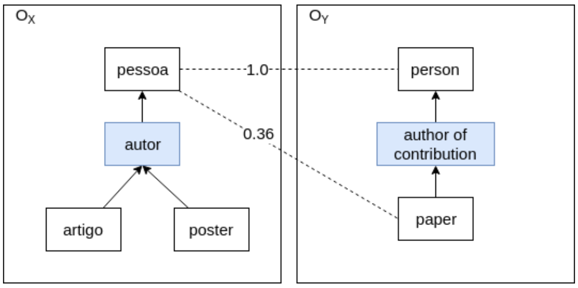

Fig. 5. "*pessoa*" and "*person*" are the pair of neighbouring concepts with maximum similarity.

The neighbourhood similarity value returned by Algorithm 2 updates the similarity value considering $composed_{sim}^{(1-nbh_{sim})}$ in Algorithm 1 (line 19). Therefore, after the neighbourhood analysis process, the final similarity value is equal to $0.80^{(1-1.0)} = 0.80^{0.0} = 1.0$, thus these concepts under analysis are influenced by the neighbourhood analysis. Because the final similarity is greater than the default threshold then a mapping is created between "autor" and "author of contribution". Note, when the neighbourhood similarity is high, close to 1, the resulting similarity also approaches 1, therefore it is likely to surpass the default threshold, and then be considered a candidate mapping.

Our method assumes as correct mappings when the neighbour concepts are quite similar even if the pair of concepts under analysis itself is not so similar. Finally, Algorithm 1 verifies whether the similarity value computed is greater than or equals to a beforehand input threshold λ (line 21 in Algorithm 1). If such condition is satisfied, the algorithm inserts the mapping $(e_1, e_2, 1, \equiv)$ into the set $\mathcal{M}_{\mathcal{O}_\mathcal{X} \to \mathcal{O}_\mathcal{Y}}$ indicating a cross-language correspondence between the entities. The output mapping set file follows the general alignment format as the same used by the Alignment API[6]. The implementation of the defined algorithms can be obtained in our project code repository[7].

[6]http://alignapi.gforge.inria.fr/index.html (As of April 2019).
[7]https://gitlab.ic.unicamp.br/jreis/evocros

Table 2: MultiFarm ontologies and statistics

Languages	Ontologies	Classes	Object Properties	Datatype Properties	Total Entities
Arabic	conference-ar	61	46	18	125
Chinese	conference-cn	61	46	18	125
Czech	conference-cz	61	46	18	125
German	conference-de	61	46	18	125
English	conference-en	61	46	18	125
Spanish	conference-es	61	46	18	125
French	conference-fr	61	46	18	125
Dutch	conference-nl	61	46	18	125
Portuguese	conference-pt	61	46	18	125
Russian	conference-ru	61	46	18	125

5. Experimental Evaluation

This evaluation aims to analyze the quality of mappings generated by our proposed technique which considers the structure of ontologies in the alignment of ontologies described in different natural languages. We conducted a series of 1260 experiments relying on a set of curated mappings manually established between ontologies described in distinct languages.

5.1. Datasets and Procedure

MultiFarm[6], version released in 2015, is the considered dataset in our experiments. This datase is used in the *OAEI* and it is composed of a set of 5 ontologies of the Conference domain[8], translated into 10 languages: Arabic (ar), English (en), Chinese (cn), Czech (cz), Dutch (nl), French (fr), German (de), Portuguese (pt), Russian (ru), Spanish (es), and the corresponding cross-language mappings between them. This dataset is based on the *OntoFarm* dataset, which has been successfully used for several years in the OAEI Conference track. Our experiments uses only Conference ontologies described in Table 2

This dataset was manually curated and may be used as a reference to assess algorithms that build automatic cross-lingual ontology mappings. For instance, the pair pt-es refers to the ontology mappings between Portuguese and Spanish conference ontologies. We consider 45 set of mappings with different pairs of language as follows: ar-cn, ar-cz, ar-de, ar-en, ar-es, ar-fr, ar-nl, ar-pt, ar-ru, cn-cz, cn-de, cn-en, cn-es, cn-fr, cn-nl, cn-pt, cn-ru, cz-de, cz-en, cz-es, cz-fr, cz-nl, cz-pt, cz-ru, de-en, de-es, de-fr, de-nl, de-pt, de-ru, en-es, en-fr, en-nl, en-pt, en-ru, es-fr, es-nl, es-pt, es-ru, fr-nl, fr-pt, fr-ru, nl-pt, nl-ru and pt-ru. Table 3 presents the used sets of mappings in our experiments.

[8]Cmt, Conference, ConfOf, Iasted, Sigkdd

Our experiments built cross-language ontology mappings by using English as a pivot language for syntactic similarity measurement. Babelnet is used for semantic similarity measurement and does not need a translation as it can retrieve the synsets used in NASARI vectors, by using the concepts original language. The results obtained by executing Algorithm 1 in different scenarios were compared with the reference mappings from the *MultiFarm dataset*, then metrics of precision, recall, and f-measure [34] were calculated.

We executed Algorithm 1 setting different weights and thresholds for similarity, but considering the minimum threshold equals to 0.33, based on fraction $\{\frac{1}{3}\}$ of the similarity spectrum analyzed $[0, 1]$. We used the mappings Conference-Conference of all 45 pair of languages in Multifarm as reference (*cf.* Table 3).

The weights of syntactic and semantic similarities in the composed similarity measure followed the fractions $\{\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}\}$, considering the constraint $\alpha + \beta = 1$. We present results varying the threshold level to comprehend its role in the studied scenarios. We vary the threshold in $\{0.66, 0.75, 0.80, 0.95\}$, which were selected based on the fractions $\{\frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{19}{20}\}$. The threshold 0.95 was chosen to evaluate the behaviour of the algorithm in contrast to the high level of threshold. Table 4 shows the experiments configuration applied for each pair of language.

5.2. Experimental Results

Table 5 presents the achieved results with the highest f-measure obtained for each pair of language. Results vary by language pair, but they have an average threshold of 0.73, average syntactic weight of 0.73 and average semantic weight of 0.27.

The highest f-measure was found between West Germanic languages (English and German), 0.64591. The majority of the results with the lowest f-measure values involve the Arabic language, present in seven pairs of the lowest ten f-measures: ar-ru, de-ar, ar-fr, cn-ar, cz-ar, ar-pt, ar-es, cz-ru, de-ru, pt-ru.

An example of how our defined technique helps improving the results can be observed in conference-conference-en-pt set of mappings. The concept "*invited speaker*" in conference-en should be mapped to concept "*Palestrante convidado*" in conference-pt, according to the gold standard. Using similarity threshold equals to 0.66, $\alpha = 0.67$ and $\beta = 0.33$, the composed similarity value calculated between them was 0.4307. This similarity value is inside the doubtful

Table 3: Conference-conference alignments used in the experiments and number of mappings in each alignment.

Alignment	# of Mappings
conference-conference-ar-es	121
conference-conference-ar-fr	121
conference-conference-ar-nl	121
conference-conference-ar-pt	121
conference-conference-ar-ru	121
conference-conference-cn-ar	121
conference-conference-cn-cz	123
conference-conference-cn-de	123
conference-conference-cn-en	123
conference-conference-cn-es	123
conference-conference-cn-fr	123
conference-conference-cn-nl	123
conference-conference-cn-pt	123
conference-conference-cn-ru	123
conference-conference-cz-ar	121
conference-conference-cz-de	123
conference-conference-cz-en	123
conference-conference-cz-es	123
conference-conference-cz-fr	123
conference-conference-cz-nl	123
conference-conference-cz-pt	123
conference-conference-cz-ru	123
conference-conference-de-ar	120
conference-conference-de-en	123
conference-conference-de-es	123
conference-conference-de-fr	123
conference-conference-de-nl	123
conference-conference-de-pt	123
conference-conference-de-ru	123
conference-conference-en-ar	121
conference-conference-en-es	123
conference-conference-en-fr	123
conference-conference-en-nl	123
conference-conference-en-pt	123
conference-conference-en-ru	123
conference-conference-es-fr	123
conference-conference-es-nl	123
conference-conference-es-pt	123
conference-conference-es-ru	123
conference-conference-fr-nl	123
conference-conference-fr-pt	123
conference-conference-fr-ru	123
conference-conference-nl-pt	123
conference-conference-nl-ru	123
conference-conference-pt-ru	123

Table 4: Experiments configurations. Different weights for syntactic and semantic similarity are applied to each threshold. Each configuration was applied to Conference-Conference mappings for each of the 45 pairs of languages obtaining a total of 1260 experiments.

Similarity threshold	Syntactic measure	Semantic measure
0.66	0.20	0.80
	0.25	0.75
	0.33	0.67
	0.50	0.50
	0.67	0.33
	0.75	0.25
	0.80	0.20
0.75	0.20	0.80
	0.25	0.75
	0.33	0.67
	0.50	0.50
	0.67	0.33
	0.75	0.25
	0.80	0.20
0.80	0.20	0.80
	0.25	0.75
	0.33	0.67
	0.50	0.50
	0.67	0.33
	0.75	0.25
	0.80	0.20
0.95	0.20	0.80
	0.25	0.75
	0.33	0.67
	0.50	0.50
	0.67	0.33
	0.75	0.25
	0.80	0.20

range and thus the neighborhood similarity was verified to confirm the candidate mapping. The calculated neighborhood composed similarity value was 0.5847. The similarity value was then recalculated using the formula $composed_{sim}^{(1-nbh_{sim})}$, thus $0.4307^{(1-0.5847)} = 0.7048$, surpassing the 0.66 threshold. Therefore, the mapping is included in the generated ontology set of mappings.

Another example is the concept *"chair of workshop track"* in conference-en and "*coordenador de trilha de workshop*" in conference-pt. Using a similarity threshold equals to 0.66, $\alpha = 0.67$ and $\beta = 0.33$, the composed similarity value calculated between them was 0.3828 and the neighborhood similarity was 1.0. The recalculated similarity value was 1.0 and then the mapping was included in the generated ontology alignment.

Table 5: Results with highest f-measure for each pair of language.

Language pair	Similarity threshold	Syntactic measure	Semantic measure	Precision	Recall	F-Measure
ar-es	0.75	0.80	0.20	0.41071	0.35659	0.38174
ar-fr	0.75	0.80	0.20	0.44444	0.27273	0.33803
ar-nl	0.66	0.80	0.20	0.37989	0.47552	0.42236
ar-pt	0.66	0.75	0.25	0.35220	0.40288	0.37584
ar-ru	0.66	0.80	0.20	0.31356	0.28030	0.29600
cn-ar	0.66	0.80	0.20	0.52381	0.25191	0.34021
cn-cz	0.75	0.80	0.20	0.63492	0.30075	0.40816
cn-de	0.66	0.80	0.20	0.50476	0.37589	0.43089
cn-en	0.66	0.80	0.20	0.54717	0.40845	0.46774
cn-es	0.66	0.75	0.25	0.45082	0.39568	0.42146
cn-fr	0.75	0.80	0.20	0.58140	0.37037	0.45249
cn-nl	0.66	0.75	0.25	0.46667	0.48611	0.47619
cn-pt	0.66	0.80	0.20	0.41111	0.49664	0.44985
cn-ru	0.66	0.80	0.20	0.50980	0.37681	0.43333
cz-ar	0.66	0.80	0.20	0.30189	0.45070	0.36158
cz-de	0.75	0.80	0.20	0.46809	0.45205	0.45993
cz-en	0.80	0.67	0.33	0.68182	0.41667	0.51724
cz-es	0.80	0.75	0.25	0.64894	0.42657	0.51477
cz-fr	0.75	0.80	0.20	0.42636	0.39007	0.40741
cz-nl	0.95	0.80	0.20	0.70000	0.42282	0.52720
cz-pt	0.75	0.67	0.33	0.58654	0.42069	0.48996
cz-ru	0.66	0.75	0.25	0.41129	0.36429	0.38636
de-ar	0.66	0.75	0.25	0.28421	0.40602	0.33437
de-en	0.80	0.67	0.33	0.77570	0.55333	0.64591
de-es	0.80	0.67	0.33	0.82258	0.36429	0.50495
de-fr	0.66	0.80	0.20	0.36111	0.59091	0.44828
de-nl	0.80	0.75	0.25	0.59406	0.41667	0.48980
de-pt	0.75	0.80	0.20	0.45963	0.49664	0.47742
de-ru	0.66	0.80	0.20	0.40146	0.38732	0.39427
en-ar	0.75	0.80	0.20	0.42029	0.45313	0.43609
en-es	0.80	0.67	0.33	0.66667	0.40845	0.50655
en-fr	0.75	0.67	0.33	0.64516	0.30075	0.41026
en-nl	0.80	0.20	0.80	0.86111	0.40260	0.54867
en-pt	0.80	0.67	0.33	0.68293	0.38889	0.49558
en-ru	0.66	0.80	0.20	0.43972	0.42759	0.43357
es-fr	0.75	0.67	0.33	0.65672	0.33333	0.44221
es-nl	0.80	0.67	0.33	0.78351	0.50000	0.61044
es-pt	0.80	0.50	0.50	0.74545	0.51572	0.60967
es-ru	0.66	0.75	0.25	0.45600	0.39583	0.42379
fr-nl	0.75	0.67	0.33	0.62500	0.40441	0.49107
fr-pt	0.75	0.67	0.33	0.63636	0.36567	0.46445
fr-ru	0.66	0.80	0.20	0.36508	0.46309	0.40828
nl-pt	0.75	0.50	0.50	0.80682	0.46104	0.58678
nl-ru	0.75	0.80	0.20	0.56075	0.42254	0.48193
pt-ru	0.66	0.80	0.20	0.34597	0.48993	0.40556

6. Discussion

Cross-language ontology matching relies on several different approaches to obtain mappings that interrelate ontologies described in distinct languages. Cross-language ontology matching requires adequate techniques relying on similarity measures to overcome the matching task barrier. Ontological structure and similarity measures might help in matching algorithms to determinate the adequate mappings. Existing techniques can favor from the understanding of the benefits and limitations of syntactic and semantic similarity approaches to develop a better combination of them.

In this context, this investigation contributed with several experiments to determinate the impact of ontological structure and the similarity measures to be considered in the alignment of ontologies described in different languages. Our experiments were designed to help us understanding their effects in the matching process and the quality of the generated cross-lingual ontology mappings.

Our proposal concerns the influence of the ontological structure and similarity computation on cross-language ontology matching. Our goal was to understand how to combine them aiming to build accurate cross-lingual ontology mappings. To this end, we took into account the weighted average between syntactic and semantic similarities. Our approach considered neighbour concepts directed related to concepts under analysis in candidate mappings.

The choice of weights assigned to each similarity measure played an important role in the results. As we showed empirically, semantic and syntactic similarities might not have the same relevance, *i.e.*, the same weight. Considering the syntactic weight close to 0.70 generated the best mapping results, *i.e.*, it resulted in ontology mappings with the overall highest f-measure value. Thus, our technique may be understood as a good alternative to syntactic or semantic only methods. It might perform even better taking into account the correct parameters.

We found that the gain of effectiveness may vary according to the language describing the content of the ontologies. Comparing the results in Table 5, we observe that the results for the arabic language are generally in the lowest tier. A possible explanation for this behaviour might be the use of Google Translate for automatic translation to the pivot language. Although Google Translate has largely improved over the years[35], there are still some incorrect or mistypes in the translations that hinder syntactic measures. An example is the word أَيستِقبَل, incorrectly translated to *“recepion"*. The correct translation is *“reception"*. Another example is أَليَوم مَهَتَ, incorrectly translated to *“today's station"* by Google Translate, when the correct translation would be *“terminal date"*.

The characteristics of the entity labels of the Conference ontology restrict the use of the semantic similarity measurement Weighted Overlap, because the entity labels are mostly complex sentences instead of words. Babelnet, the external source used in semantic measurement, is a dictionary, not a translation tool and therefore only able to identify synsets in words. This also explains the average semantic similarity being approximately $\frac{1}{4}$ of the overall weight. Thus, it might be useful considering semantic algorithms such as stop-words elimination and stemming, etc. to break the complex sentences into simple structures.

The results showed an influence of threshold; as the threshold rises, the precision also increases. It may be explained by considering equivalence of only those concepts with a high level of similarity. However, the f-measure value reduces as the threshold increases. This happens because higher values assigned to threshold leads to the algorithm disregarding entities that are equivalent, but somehow were assigned a lower level of similarity than expected by the threshold. For instance, in en-es ontology mappings, the similarity between “strange" and “estranho" was equal to 0.89, but the given threshold is 0.95, thus “estranho" is not mapped to “strange". As a result, the recall drops substantially, because many correct correspondences are ignored, and thus f-measure decreases. Empirically, we concluded that the thresholds generating the more accurate mappings were around $\lambda = 0.75$ and $\lambda = 0.80$.

Table 6 describes the results obtained by related work (ontology alignment systems) presented in OAEI (the version of 2018) with the same dataset in which our experiments were conducted. By comparing our obtained results to the presented systems, our average f-measure for all languages value of 0.45 surpasses three of the four assessed systems, AML [18], whose average f-measure is equal to 0.27, XMAP [36], which presented 0.14 of average f-measure and LogMap [24]. With these results, our algorithm gets closer to KEPLER [25], the best tool placed in the competition in 2018 when considering the f-measure metric and even surpasses KEPLER in 12 out of the 45 language pairs.

Our obtained findings support the hypothesis that composing different types of similarity measures and

Table 6: Results obtained with existing ontology alignment systems in OAEI (Multifarm Track) in 2018, considering the alignments between conference ontologies in different languages.

Tool	Precision	Recall	F-measure
KEPLER [25]	0.85	0.36	0.49
LogMap [24]	0.95	0.28	0.41
AML [18]	0.96	0.16	0.2
XMAP [36]	0.13	0.19	0.14

taking into account the ontology structure, by considering the similarity of neighbour concepts, can reveal satisfactory generated ontology mappings for cross-language ontology alignment.

7. Conclusion

Alignment of large ontologies described in different natural languages remains an open research challenge. In this investigation, we proposed an approach based on the weighted mean of syntactic and semantic similarities for this task. Our approach considered the influence of neighbour concepts on the cross-lingual alignment method, combining it with the composed similarity. The defined algorithms were implemented and we carried out a series of experiments to evaluate the effectiveness of this approach. Our findings based on experiments with standard datasets revealed the effectiveness of combining similarity measures and considering the neighbourhood of concepts in the cross-language ontology alignment problem. Future work involves to improve our cross-lingual alignment proposal by considering different combinations of background knowledge, such as specific-domain thesauri to evaluate the semantic similarity. In addition, we plan to investigate different ways of computing the syntactic and semantic similarities considering additional stages in the pre-processing of entity labels.

Acknowledgements

This work has the financial support of CNPq (grant #307560/2016-3), São Paulo Research Foundations (FAPESP) (grants #2017/02325-5, #2014/12236-1, #2015/24494-8, #2016/50250-1, and #2017/20945-0) and the FAPESP-Microsoft Virtual Institute (grants #2013/50155-0 and #2014/50715-9). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001[9].

References

[1] P. Shvaiko and J. Euzenat, in: *Ontology matching: State of the art and future challenges*, Vol. 25, 2013, pp. 158–176.

[2] T. Ivanova, Cross-lingual and Multilingual Ontology Mapping - Survey, in: *Proceedings of the 19th International Conference on Computer Systems and Technologies*, CompSysTech'18, ACM, New York, NY, USA, 2018, pp. 50–57. ISBN 978-1-4503-6425-6. doi:10.1145/3274005.3274034. http://doi.acm.org/10.1145/3274005.3274034.

[3] C. Pesquita, D. Faria, A.O. Falcão, P.W. Lord and F.M. Couto, Semantic Similarity in Biomedical Ontologies, *PLoS Computational Biology* **5**(7) (2009). doi:10.1371/journal.pcbi.1000443. https://doi.org/10.1371/journal.pcbi.1000443.

[4] M.T. Pilehvar, D. Jurgens and R. Navigli, Align, disambiguate and walk: A unified approach for measuring semantic similarity, in: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1, 2013, pp. 1341–1351.

[5] R. Navigli and S.P. Ponzetto, BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, *Artificial Intelligence* **193** (2012), 217–250.

[6] C. Meilicke, R. Garcia-Castro, F. Freitas, W.R. van Hage, E. Montiel-Ponsoda, R.R. de Azevedo, H. Stuckenschmidt, O. Sváb-Zamazal, V. Svátek, A. Tamilin, C.T. dos Santos and S. Wang, MultiFarm: A Benchmark for Multilingual Ontology Matching, *Web Semantics: Science, Services and Agents on the World Wide Web* **15** (2012), 62–68.

[7] C. Meilicke, C. Trojahn, O. Šváb-Zamazal and D. Ritze, Multilingual ontology matching evaluation–a first report on using multifarm, in: *Extended Semantic Web Conference*, Springer, 2012, pp. 132–147.

[8] C. Trojahn, B. Fu, O. Zamazal and D. Ritze, State-of-the-art in multilingual and cross-lingual ontology matching, in: *Towards the Multilingual Semantic Web*, Springer, 2014, pp. 119–135.

[9] B. Fu, R. Brennan and D. O 'sullivan, Cross-lingual Ontology Mapping - An Investigation of the Impact of Machine Translation, in: *Proceedings of the 4th Annual Asian Semantic Web Conference (ASWC 2009)*, Springer, 2009.

[10] D. Spohr, L. Hollink and P. Cimiano, A machine learning approach to multilingual and cross-lingual ontology matching, in: *Proceedings of the 10th International Semantic Web Conference (ISWC 2011)*, Springer, 2011, pp. 665–680.

[11] A. Annane, Z. Bellahsene, F. Azouaou and C. Jonquet, Building an effective and efficient background knowledge resource to enhance ontology matching, *Journal of Web Semantics* **51** (2018), 51–68.

[12] S.K. Stoutenburg, Acquiring advanced properties in ontology mapping, in: *Proceedings of the 2nd PhD workshop on Information and knowledge management*, 2008, pp. 9–16.

[9]The opinions expressed in here are not necessarily shared by the financial support agencies.

[13] T.T.A. Nguyen and S. Conrad, Ontology Matching using multiple similarity measures, in: *7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, Vol. 01, 2015, pp. 603–611.

[14] J.M. Destro, J.C.d. Reis, A.M. Brito, R. Carvalho and I.L.M. Ricarte, Influence of Semantic Similarity Measures on Ontology Cross-language Mappings, *Proceedings of the Symposium on Applied Computing* (2017), 323–329. ISBN 978-1-4503-4486-9. doi:10.1145/3019612.3019836. http://doi.acm.org/10.1145/3019612.3019836.

[15] A. Essayeh and M. Abed, Towards ontology matching based system through terminological, structural and semantic level, *Procedia computer science* **60** (2015), 403–412.

[16] H. Lin, Y. Wang, Y. Jia, J. Xiong, P. Zhang and X. Cheng, An ensemble matchers based rank aggregation method for taxonomy matching, in: *Asia-Pacific Web Conference*, Springer, 2015, pp. 190–202.

[17] A. Khiat, CroLOM results for OAEI 2017: summary of cross-lingual ontology matching systems results at OAEI, *CEUR Workshop Proceedings* **2032** (2017), 129–134.

[18] D. Faria, B.S. Balasubramani, V.R. Shivaprabhu, I. Mott, C. Pesquita, F.M. Couto and I.F. Cruz, Results of AML participation in OAEI 2018, *CEUR Workshop Proceedings* **2288** (2018).

[19] D. Ngo, Z. Bellahsene and R. Coletta, Yam++-a combination of graph matching and machine learning approach to ontology alignment task, *Journal of Web Semantics* **16** (2012), 16.

[20] V.I. Levenshtein, Binary codes capable of correcting deletions, insertions and reversals., *Soviet Physics Doklady* **10**(8) (1966), 707–710, Doklady Akademii Nauk SSSR, V163 No4 845-848 1965.

[21] B. Fu, R. Brennan and D. OâĂŹSullivan, A configurable translation-based cross-lingual ontology mapping system to adjust mapping outcomes, *Web Semantics: Science, Services and Agents on the World Wide Web* **15** (2012), 15–36.

[22] I.F. Cruz, F.P. Antonelli and C. Stroe, AgreementMaker: efficient matching for large real-world schemas and ontologies, *Proceedings of the VLDB Endowment* **2**(2) (2009), 1586–1589.

[23] D. Ngo and Z. Bellahsene, Efficient Semantic Verification of Ontology Alignment, *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* **1** (2015), 141–148.

[24] E. Jimenez-Ruiz, B.C. Grau and V. Cross, LogMap family participation in the OAEI 2018, *CEUR Workshop Proceedings* **2288** (2018).

[25] M. Kachroudi, G. Diallo and S.B. Yahia, KEPLER at OAEI 2018, *CEUR Workshop Proceedings* **2288** (2018).

[26] W.F. Dowling and J.H. Gallier, Linear-time algorithms for testing the satisfiability of propositional Horn formulae, *The Journal of Logic Programming* **1**(3) (1984), 267–284.

[27] N.F. Noy and D.L. McGuinness, Ontology Development 101: A Guide to Creating Your First Ontology, (accessed: 04.Feb.2021).

[28] P.S. JÃľrÃťme Euzenat, in: *Ontology Matching, 2nd Edition*, Springer, Heidelberg, 2013, pp. 85–90.

[29] D. Dinh, J.C. Dos Reis, C. Pruski, M. Da Silveira and C. Reynaud-Delaître, Identifying relevant concept attributes to support mapping maintenance under ontology evolution, *Web semantics: Science, services and agents on the world wide web* **29** (2014), 53–66.

[30] G. Navarro, A Guided Tour to Approximate String Matching, *ACM Comput. Surv.* **33**(1) (2001), 31–88, ISSN 0360-0300. doi:10.1145/375360.375365. http://doi.acm.org/10.1145/375360.375365.

[31] J. Camacho-Collados, M.T. Pilehvar and R. Navigli, Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities, *Artificial Intelligence* **240** (2016), 36–64.

[32] P. Lafon, Sur la variabilité de la fréquence des formes dans un corpus, *Mots. Les langages du politique* **1**(1) (1980), 127–165.

[33] M.T.P. José Camacho-Collados and R. Navigli, NASARI: a Novel Approach to a Semantically-Aware Representation of Items, in: *Proceedings of the North American Chapter of the Association of Computational Linguistics (NAACL 2015)*, Denver,USA, 2015, pp. 567–577.

[34] D.M.W. Powers, Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation, *Journal of Machine Learning Technologies.* **2** (2011), 37–63.

[35] Y. Wu, M. Schuster, Z. Chen, Q.V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey et al., Google's neural machine translation system: Bridging the gap between human and machine translation, *arXiv preprint arXiv:1609.08144* (2016).

[36] W.E. Djeddi, S.B. Yahia and M.T. Khadir, XMap : Results for OAEI 2018, *CEUR Workshop Proceedings* **2288** (2018).