Exploring Rank Aggregation for Cross-Lingual Ontology Alignments

Juliana M. Destro [a,*], Javier A. Vargas [a], Julio Cesar dos Reis [a] and Ricardo da S. Torres [a]
[a] *Institute of Computing, University of Campinas, SP, Brazil*
E-mails: juliana.destro@ic.unicamp.br, jalvarm.acm@gmail.com, jreis@ic.unicamp.br, rtorres@ic.unicamp.br

Abstract.
Cross-language ontology alignments are of paramount importance in several applications. A common approach to define proper alignments relies on identifying the relationships among concepts from different ontologies by performing multiple entity-based searches. In this strategy, the most suitable matching is defined by the top-ranked concept found. Often, multiple similarity rankers, defined in terms of different similarity criteria, are considered to define candidate entities. In this case, their complementary view could be exploited in the definition of the best possible matching. In this paper, we explore the use of rank aggregation functions, under both unsupervised and supervised settings, in the task of defining suitable matches among entities belonging to ontologies encoded in distinct languages. We conducted a comprehensive set of experiments with standard datasets from the OAEI competition, using ontologies in the Conference domain, and mappings among 36 language pairs. Experimental results show that the use of rank aggregation approaches leads to better f-measure results when compared with state-of-the-art techniques in cross-language ontology matching.

Keywords:
ontology matching, cross-lingual, rank aggregation, information retrieval

1. Introduction

Nowadays, software systems are more and more data intensive applications heavily relying on automatic method to integrate and retrieve data among several systems. The globalization requires that systems interoperate with others regardless the natural language encoding the data. In this context, semantic-enabled proposals demand means to describe the meaning of data with a smooth data integration.

Ontologies define a common vocabulary in a knowledge domain and are used to represent semantics in computational systems. The use of ontologies is a key element to represent knowledge by describing concepts and interrelationships among them in a domain. Ontologies are often employed as means to represent data in systems that benefit from explicit semantics, such as information retrieval systems, semantic search engines and semantic query systems. In this context, ontology mappings facilitate interconnectivity, enabling refined techniques for data integration, data mining, and analytics by correlating concepts from different and heterogeneous data sources. We use ontology in a general sense, including taxonomies and thesauri [1].

Ontology entities vary among authors, but they mostly include instances (the objects), concepts (representing classes or types of objects), attributes (features, properties, and characteristics of concepts) and relations (representing relationships between entities). Ontology concepts explore strings written in natural language to denote labels [2]. An example of ontology entity is a concept labeled "*Diabetes mellitus type 1*" with attribute "*Synonym:Type I Diabetes*," and an *is-a* ($\sqsubseteq$) relation with another concept named "*Diabetes Mellitus*."

Generation of correspondences between concepts from two different ontologies [3] is known as matching, whereas the result of this process is known as a mapping set or alignment. An example of a mapping

*Corresponding author. E-mail: juliana.destro@ic.unicamp.br.

would be a correspondence found between the concept "*Cardiopathy*" in one ontology and the concept "*Arrhythmias*" in another. In a matching process, the correspondence between these two concepts is found and added to a mapping set. When the ontologies are labeled in different natural languages, the process of creating correspondences is called cross-lingual ontology matching. Methods for ontology matching across languages are crucial for enabling end-users to access data in multiple languages.

Cross-language ontology matching is an open research problem. Ontologies are usually created for diversified purposes, even in the same domain, thus making it difficult to find correspondences between ontology elements. The matching problem is heightened when ontologies are written in different natural languages or different symbolic systems, making a simple comparison between strings a difficult process. The applicability of existing monolingual ontology matching methods is limited, due to the complexity of dealing with distinct languages [4]. Current techniques for cross-language matching do not provide results comparable to existing monolingual approaches [5].

Rank aggregation refers to the process of combining a set of ranked lists to obtain a final combination that is potentially better than any individual ranking. This process combines results from different rankers with the aim of obtaining more effective retrieval results. This is a technique applied in several tasks of information retrieval. The use of rank aggregation in similarity search [6] is of particular interest for ontology matching, as this problem can be modeled as a similarity search problem. The expected benefit of leveraging rank aggregation in cross-language ontology matching is the combination of different similarity measures, each one offering a particular and potentially complementary view of the similarity between elements. While single ranking techniques are used in ontology matching [7], rank aggregation is yet under explored.

In this investigation, we explore supervised and unsupervised rank aggregation techniques to combine distinct ranks constructed based on similarity measures. Supervised methods exploit labeled information (*i.e.*, training data) and ground-truth relevance to boost the effectiveness of a new ranker. Unsupervised methods do not rely on labeled training data, instead it is based on data discrimination or summarization, such as clustering and retrieval score combination. Our goal is to leverage rank aggregation in cross-lingual mapping, by generating ranked lists based on distinct similarity measurements between the concepts of source and target ontologies.

Our proposal compares an entity in a source ontology with all entities in a target ontology and generate several ranks based on similarity values. Afterwards, a rank aggregation technique is applied to generate a single rank per entity in the source ontology. A correspondence is created between the entity in source ontology and the top-1 final rank. In this study, several rank aggregation techniques are applied and their results are compared.

We conduct extensive experiments with standard datasets from the OAEI competition, using ontologies in the Conference domain, and mappings in 36 language pairs. We evaluate the benefits of several rank aggregation techniques. Experimental results show that the use of rank aggregation approaches leads to better f-measure scores than the ones observed for state-of-the-art techniques in cross-lingual ontology matching. In addition, our solution adds flexibility on the use of different similarity approaches.

The remaining of this article is organized as follows: Section 2 describes basic concepts and related work; Section 3 presents the formalization of the technique and the defined algorithms; Section 4 describes the experimental evaluation, whereas Section 5 discusses our findings; Section 6 provides the conclusion remarks.

2. Background and related work

We describe the basic concepts (Section 2.1) followed by the related work analysis (Section 2.2).

2.1. Basic concepts

Ontologies define a common vocabulary in a domain [8]. They are used for semantic representation in computational systems, describing the definition of concepts and the relationship among them. An ontology $\mathcal{O}$ describes a domain in terms of entities, consisting of concepts, attributes, and relationships [8]. Formally, an ontology $\mathcal{O} = (\mathcal{C}_{\mathcal{O}}, \mathcal{R}, \mathcal{A}_{\mathcal{O}})$ consists in a set of concepts $\mathcal{C}_{\mathcal{O}}$ interrelated by a set of directed relations $\mathcal{R}$. Each concept $c \in \mathcal{C}_{\mathcal{O}}$ has an unique identifier and it is associated with a set of attributes $\mathcal{A}_{\mathcal{O}}(c) = \{a_1, a_2, ..., a_p\}$. 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 the concepts (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".

Ontology matching stands for the generation of correspondences (or mappings) between elements from two different ontologies [3]. An example of a mapping would be a correspondence found between the concept "*Cardiopathy*" in one ontology and the concept "*Arrhythmias*" in another. When the input ontologies are labeled in different natural languages, the process of creating correspondences is called cross-language ontology matching. The final result of the matching process is a set containing the mappings found between the entities from two given ontologies.

Formally, the matching is the process of identifying the relationship between entities from different ontologies. For entities $e_i \in O_X$ and $e_j \in O_Y$, the alignment is expressed by the 4-tuple $m_{ei\rightarrow ej} = (e_i, e_j, s_{i,j}, r(e_i, e_j))$, 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 entities. A cross-language ontology mapping occur when O_X and O_Y are expressed in languages α and β respectively, such that $\alpha \neq \beta$.

Figure 1 presents an example of a cross-language ontology mapping. The source ontology is O_X, with labels of concepts written in English and target ontology is O_Y, with labels of concepts written in Portuguese. Concept "Vehicle" in O_X is mapped to the concept "Carro" in O_Y, identifying an equivalence relationship between these two concepts. The same equivalence relation is applied to the mapping between "Person" and "Proprietário".

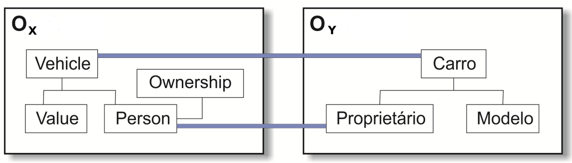

Figure 1. Example of cross-language ontology mapping.

Similarity between entities is established when given two entities e_i and e_j from an ontology (or from different ontologies) and a function to identify the similarity value between them. Formally:

$$sim(e_i, e_j) = f(e_i, e_j) \quad (1)$$

where $f(e_i, e_j)$ is the similarity measure calculated in different linguistic levels, from string-based methods to semantic techniques [9] and falls under the interval [0,1].

Problem formalization. The target problem is to determine a *mapping set* associated with equivalence relationships among entities belonging to two different ontologies described in distinct languages. In the addressed problem, given two ontologies described in distinct natural languages, our technique must obtain a mapping set through the use of results of entity-based similarity searches. Those results are defined in terms of the aggregation of ranked lists defined according to different rankers. Rank aggregation refers to the combination of a set of preference lists to produce an unique combined list. In our case, the mapping process is modeled as multiple *similarity searches*, in which entities of a target ontology are ranked according to their similarity (*cf.* Equation 1) to an input entity of a source ontology.

Figure 2 presents the rank aggregation in the context of information retrieval applied to the ontology matching problem. The input entity can be seen as the *query* pattern of a search system. The goal is to select the top-ranked entity of the target ontology as the best matching. Each similarity ranker produces a ranked list based on the similarity measure for the query $q(e_i)$, where $e_i \in O_X$. Entities are ranked according to different criteria, leading to multiple ranked lists, which can be combined by means of rank aggregation functions. In particular, the ranked lists are ordered in descendant order in respect to the similarity value calculated for each entity $e_j \in O_Y$.

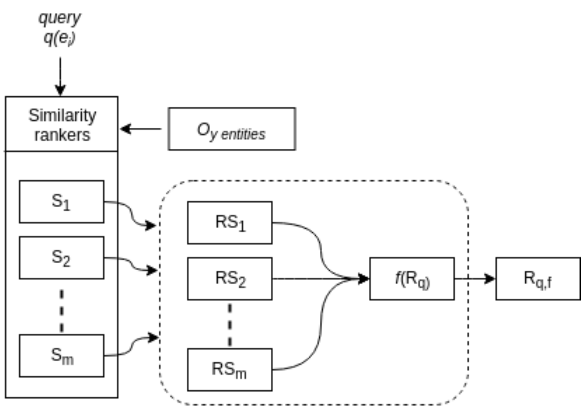

Figure 2. Rank aggregation in information retrieval mapped to the ontology matching problem.

Then, using a rank aggregation function, the lists are combined to produce a final list. Formally, let $S = \{S_1, S_2, ..., S_m\}$ be a set of similarity rankers based on which we can compute m ranked lists, $RS_1, RS_2, ..., RS_m$, given a query $q(e_i)$. Let R_q be the set formed by these ranked lists, *i.e.*, $R_q = \{RS_1, RS_2, ..., RS_m\}$. A rank aggregation function f is used to define a ranked list $R_{q,f}$ for $q(e_i)$, such that $R_{q,f} = f(R_q)$.

2.2. Related work

Several aspects and approaches of the cross-language ontology matching problem have been investigated over the years [5]. A common approach relies on adapting traditional monolingual techniques for carrying out multilingual ontology alignments [4]. However, achieved results indicate the difficulties of applying such techniques for cross-language alignments. The use of a third language in cross-language ontology matching and automatic translations were explored by Fu *et al.* [10]. Spohr *et al.* [11] studied the translation of concept labels to a third language for matching two ontologies described in different languages. Stoutenburg [12] investigated the use of ontologies combined with linguistic resources as background knowledge (*e.g.*, dictionary, thesauri, *etc.*) to enhance ontology matching processes.

The supervised rank aggregation problem has received considerable attention in recent years and a number of supervised approaches have been proposed. For instance, the work of Pujari and Kanawati [13], Subbian and Melville [14], and Wu [15].

Notably, Volkovs and Zemel [16] have shown that by applying singular-value decomposition (SVD) factorization to pairwise preference matrices can lead to effective extraction of item features. The features transform the problem into a standard Learn To Rank (L2R) one, allowing to apply any of the existing L2R methods to optimize the aggregating function for the target metric. Although the authors have shown superior empirical effectiveness of the approach to several existing aggregation methods, it also has a major drawback as it requires computing SVD factors at training or test time. For large problems with many documents per query, applying SVD at training or even test time can be prohibitively expensive, limiting the application of the method. A number of other popular supervised aggregation methods shares the same disadvantage and requires applying complex optimization procedures such as semi-definite programming [17].

In another work, Volkovs and Zemel [18] addressed the complexity of testing time by developing a flexible Conditional Random Field (CRF) framework for supervised rank aggregation. This framework uses preference matrices directly in order to avoid costly optimizations. The authors demonstrate that CRF has superior effectiveness when compared to some supervised alternatives.

The use of information retrieval techniques in ontology mapping has attracted interest in the community. Rexha *et al.* [7] explored the PageRank technique, where ontologies were indexed through an inverted index algorithm and candidate matches identified by querying such indexes. Kachroudi *et al.* [19] adopted an indexing scheme to identify which documents contain a specific word, as part of a workflow to produce ontology mapping.

Domshlak *et al.* [20] proposed a *schema metamatching* using rank aggregation to ensemble schema matchers, generating a list of best ranked schema mappings. The goal is to find a “consensus” ranking of mappings between two schemata, given the “individual” rankings provided by several schema matchers using a top-K mapping approach. This work investigated four rank aggregation algorithms.

A similar method proposed by Su and Gulla [21] uses a ranking function based on an initial similarity assertion of the concepts to identify top-K elements. Then it enhances the evaluation with other matching strategies.

A top-K mapping is explored by Wang *et al.* [22]. However, their work includes user feedback for error correction and mapping refinement. In order to minimize workload, user's feedback is only requested when mapping is considered uncertain, based on an attribute entropy measurement.

To the best of our knowledge, none of the methods or techniques in the literature proposed the use of rank aggregation for cross-language ontology matching, which demonstrates the originality of our approach. Also, our technique adds flexibility to include new similarity measures *ad-hoc*.

3. Cross-lingual Alignment Using Rank Aggregation

We propose a technique that combines syntactic and semantic similarity measures with information retrieval techniques for cross-lingual ontology matching.

We model the mapping problem as an information retrieval query. Figure 3 depicts the workflow of the proposed technique. The inputs are source and target ontologies written in Web Ontology Language (OWL). These ontologies are converted to objects.

Each entity of the source ontology is compared with all entities of the same type found in the target ontology (*i.e.*, classes are matched to classes and properties are matched to properties). In this sense, for each entity e_i in the source ontology O_X, we calculate the similarity value with each entity e_j in the target ontology O_Y (Figure 4), thus generating a ranked list $\{rank1, rank2, rank3, rank4\}$ for each similarity measure used (*cf.* Figure 5).

For similarity measures that rely on monolingual comparison (*i.e.*, syntactic and WordNet), the automatic translation of labels of entities $e_i \in O_X$ and $e_j \in O_Y$ to a pivot language is used by leveraging Google Translate API during runtime. These similarity comparisons generate k ranks, each one based on a different similarity measure. We use the measures to generate the ranks, thus adding the flexibility to the use or the addition of different similarity measures without disrupting the technique.

The ranks are then aggregated using supervised and unsupervised methods (cf. Section 3.1). Figure 6 presents that the set of multiple ranks are aggregated in a final rank. The Top-1 result of the aggregated rank $c_2 \in \mathcal{C}_{\mathcal{O}_Y}$ is mapped to the source ontology entity $c_1 \in \mathcal{C}_{\mathcal{O}_X}$, thus generating the candidate mapping $m(c_1, c_2)$ (*cf.* Figure 7). The mapping output follows the standard used by the Alignment API [23].

We present an use case to illustrate the use of rank aggregation function in a cross-lingual ontology alignment procedure. We consider two ontologies[1], O_X and O_Y (illustrated in Figure 8), where O_X is described in English language and O_Y is described in Portuguese language.

Our goal is to find a matching of concepts for "author of contribution". Because this is a concept in the ontology O_X, this element is compared with all elements of the same type in O_Y : {"*pessoa*", "*autor*", "*artigo*", "*poster*"}.

First, the similarity values are calculated between the concept "author of contribution" and all concepts in the ontology O_Y {"*pessoa*", "*autor*", "*artigo*", "*poster*"}, generating a rank based on the similarity values, per similarity measure. Figure 9 illustrates the ranks generated by each similarity measure.

After the ranks are aggregated using one of the methods described in Section 3.1, the final mapping is established between the analyzed entity "author of contribution" and the top-1 element in the final aggregated rank, "autor" (*cf.* Figure 10).

This technique ensures that all entities of the source ontology are matched to a corresponding entity of the target ontology.

3.1. Rank aggregation algorithms

Table 1 presents the set of rank aggregation techniques, supervised and unsupervised, considered in the experimental evaluation (*cf.* Section 4). Unsupervised approaches, commonly, apply simple mathematical operations over the similarity values obtained by an element in each ranked list to compute their final similarity value (*e.g.*, *sum*, *mean*, *max*, and *product*). Also, applying the same operations, some of those approaches use only the relative position of elements (rank) to compute the final scores.

In the supervised approaches, the problem of rank aggregation is mapped to the problem of learning to rank when the similarity values are available in the ranked lists. To this end, we considered for experimental purpose all the learning to rank techniques implemented in the well-known library *RankLib*.[2]

In addition, we considered a recent investigation introducing a Genetic Programming framework [24] that allows to find near-optimal combinations of the unsupervised rank aggregation techniques that improve the results of each technique individually. Finally, we took into account another Genetic-Programming-based approach [25] for learning to rank, which aims to find a mathematical function (formula), through an evolution inspired optimization process. The technique combines the similarity values obtained by an element in each ranked list, and then uses that formula to obtain the final similarity values of elements.

3.2. Similarity measures

Each one of the four generated rankings correspond to the application of a similarity measure: two syntactic (Levenshtein and Jaro, chosen by their performance reported by Christen study [39]) and two se-

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

[2]RankLib: https://sourceforge.net/p/lemur/wiki/RankLib/ (As of June 18, 2019).

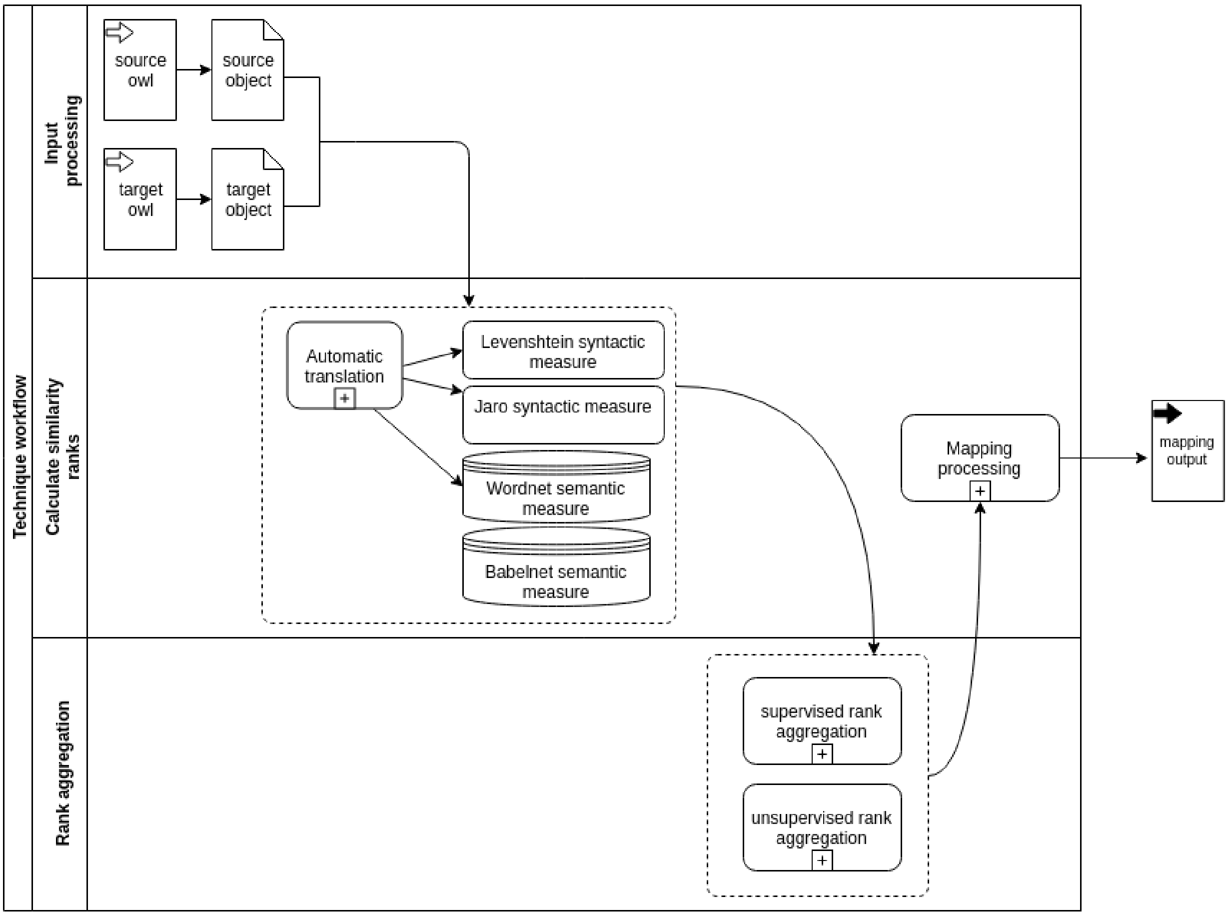

Figure 3. Technique workflow. The mapping processing stage is where the top-1 entity of the final ranking is mapped to the input concept e_1.

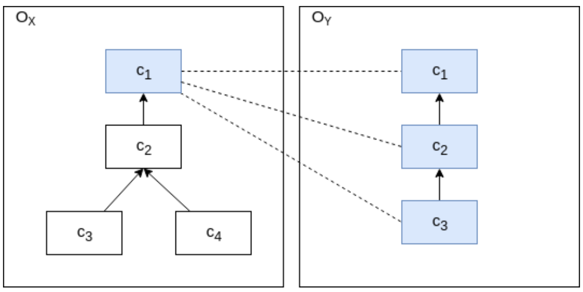

Figure 4. Concept $c_1 \in O_X$ is compared against all concepts $c_n \in O_Y$.

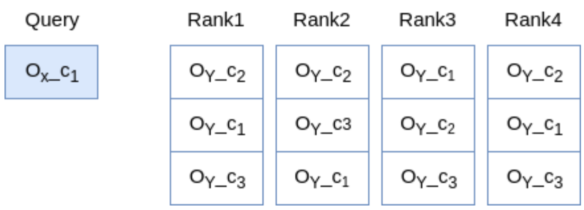

Figure 5. Ranked lists generated by each similarity measure used.

mantic similarity measures (Path-similarity for WordNet and Weighted Overlap for Babelnet). The results of the similarity measures are normalized to produce a comparable score in the range 0 to 1, where 1 represents equivalence (*i.e.*, comparing a synset with itself will return 1).

Levenshtein: also known as edit-distance [40], this technique relies on the number of single character edits (*i.e.,* insertions, deletions, substitutions) required to change one word into another.

Jaro: a similarity measure that uses the amount m of characters in common between two strings $s1$ and $s2$,

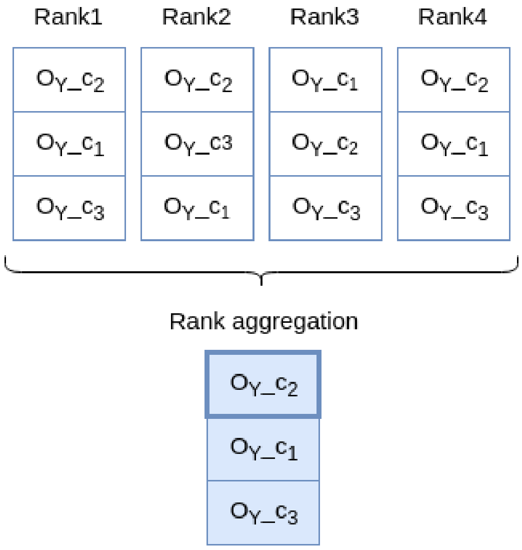

Figure 6. Rank aggregation of the ranked lists. Each rank aggregation algorithm generates a distinct final rank.

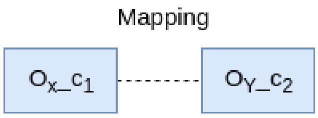

Figure 7. Mapping generated between source entity $c_1 \in O_X$ and top-1 entity of the final rank generated by the rank aggregation algorithm, $c_2 \in O_Y$.

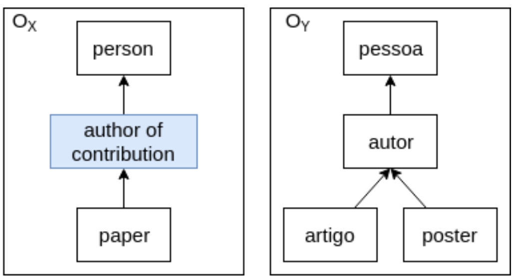

Figure 8. Entity "author of contribution" $\in O_X$ under analysis.

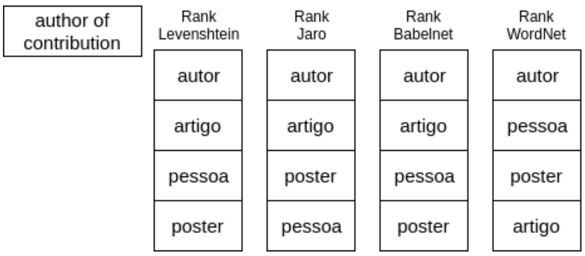

Figure 9. Ranks generated by similarity measures.

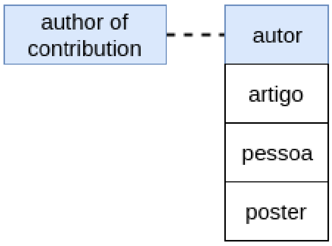

Figure 10. Mapping established between the element in O_X and the top-1 of the final aggregated rank.

Table 1
Unsupervised and Supervised Rank Aggregation (RA) and Learning to Rank (L2R) techniques considered in the experiments.

Unsupervised	Supervised
ComSUM [26] (RA)	GP-Agg [24] (RA)
CombMAX [26] (RA)	RankGP [25] (L2R)
CombMIN [26] (RA)	MART [31] (L2R)
CombMED [26] (RA)	RankNet [32] (L2R)
CombMNZ [26] (RA)	RankBoost [33] (L2R)
CombANZ [26] (RA)	AdaRank [34] (L2R)
BordaCount [27] (RA)	Coordinate Ascent [35] (L2R)
RRF [28] (RA)	LambdaMART [36] (L2R)
MRA [29] (RA)	ListNet [37] (L2R)
RL-Sim [30] (RA)	Random Forests [38] (L2R)

with consideration for transpositions (t is half the number of transpositions) [41]:

$$d_j = \frac{1}{3}(\frac{m}{|s1|} + \frac{m}{|s2|} + \frac{m-t}{m}) \quad (2)$$

Path-similarity: based on WordNet synsets (*i.e.*, set of one or more synonyms that represent the same meaning). This measure returns a score denoting how similar two synsets are, based on the shortest path that connects the senses in the *is-a* (hypernym/hyponym) taxonomy.
Weighted Overlap: allows cross-lingual similarity measurement by using NASARI vectors, based on the semantic representations of synsets in *BabelNet* [42]. The vectors are created in two steps:

- For a given concept, it collects a set of Wikipedia pages where the concept is mentioned. This contextual information collection is done by leveraging the structural information in both Wikipedia and WordNet.
- The collected contextual information is processed to extract the most representative words and synsets. A statistical measure (lexical specificity

Table 2

Example of word-based vector representation.

BabelSynsetId	WikipediaPageTitle	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

[43]) is used in this step to process the collected information. The goal is to find the most relevant words and synsets appearing in the contextual information and assign 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 2 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 contains all the synonyms which express that meaning in a range of different languages). On each row of the NASARI vector table (exemplified by two rows in Table 2), 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 *Babel synset* ID and its correspondent weight (*e.g.*, vector dimension in column *synset1_weight1*, where 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 the vectors have *Babel synset* as their dimensions, they are comparable across languages.

NASARI explores the *Weighted Overlap (WO)* method applied to the semantic vectors representations [44]. Formally:

$$Sim_{NA}(e_1, e_2) = WO(v_1, v_2), \tag{3}$$

where v_1 and v_2 refer to the word-based vector representation of the string elements e_1 and e_2, respectively. The string elements can be any string looked up for a correspondence on *BabelNet* and matched with a vector in NASARI (*e.g.*, in Table 2, e_1 can be represented by the string "*100 (number)*", and e_2 by the string "*1000 (number)*"). The similarity is computed by comparing the corresponding vectors, which results in similarity scores. The measure *WO* computes the weighted average of the two similarity scores, as follows:

$$WO(v_1, v_2) = \frac{\sum_{q \in O} (r_q^1 + r_q^2)^{-1}}{\sum_{i=1}^{|O|} (2i)^{-1}}, \tag{4}$$

where O refers to the set of overlapping dimensions between the two vectors (*i.e.*, dimensions appearing on both vectors; in the example in table 2, dimension bn:00058285n under colum *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 the *WO* equation; it is used only for ranking (*i.e.*, sorting) the dimensions.

The main advantage of NASARI is the domain neutrality offered by its base. In the lack of a multilingual domain-specific semantic network, NASARI provides a way to query and discover semantic relations among thousands of words and terms for more than 270 languages available in *BabelNet*.

4. Experimental Evaluation

The conducted evaluation aims to analyze the quality of mappings generated by our proposed technique, which considers rank aggregation in the generation of cross-lingual ontology mappings. We carried out a series of experiments relying on a set of curated mappings manually established between ontologies described in different languages.

4.1. Evaluation Protocol

We consider the *MultiFarm* dataset [45], version released in 2015, in our experiments. This dataset is used in the *OAEI MultiFarm* track and is composed of a set of 5 ontologies of the Conference domain: Cmt, Conference, ConfOf, Iasted, Sigkdd. Each ontology is translated into 10 languages: Arabic (ar), English (en), Chinese (cn), Czech (cz), Dutch (nl), French (fr), German (de), Portuguese (pt), Russian (ru), Spanish (es). Multifarm is based on the *OntoFarm* dataset, which has been successfully used for several years in the OAEI Conference track. The cross-lingual alignments of this dataset were 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 mapping between Portuguese and Spanish. Our experiments uses the subset of conference ontologies as described in Table 3

Table 3

MultiFarm ontologies and their number of entities

Language	Ontology	Classes	Object Prop.	Datatype Prop.	Total Entities
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

Our experiments built cross-language ontology mappings by using English as a pivot language for Levenshtein, Jaro, and WordNet similarity measures. The semantic similarity relying on the Babelnet does not require a translation as it can retrieve the synsets used in NASARI vectors, by using the concepts original language. The application of each similarity measure in our technique generated a rank.

The ranks were aggregated using each one of rank aggregation methods described in Subsection 3.1. Results generated mappings that were compared with the reference mappings from the *MultiFarm dataset*. The results were analyzed in terms of the metrics of precision, recall, and f-measure [46].

An evaluation protocol was established for both unsupervised and supervised rank aggregation methods. A subset of all languages was used for training and validation. The subsets are 10% of queries for training set, 15% queries for validation set, and 75% queries for testing. These subsets were generated per language and then combined, so the algorithms were trained, validated and tested using all languages at once. The comparable gold standard (*i.e.*, MultiFarm manually curated mappings) were adjusted to contain only the queries related to the testing subset. In this sense, a lower number of entities was considered in the tests, because we removed the set of queries used in training and validation from the reference mappings to assure consistency. Although the unsupervised method does not require training, the evaluation was performed with only the testing set, similarly to the supervised methods, because all the algorithms were executed simultaneously.

4.2. Results

Table 4 presents the rank aggregation method which achieved the highest precision, recall, and f-measure obtained for each language pair. The precision, recall, and f-measure scores have the same value due to the nature of the experiments. Our approach generates $n : n$ mappings, where $n = |O_X| = |O_Y|$ because the ontologies are translations of each other to different natural languages, thus every entity in the source ontology presents a correspondence in the target ontology. In this sense, both the gold standard and the generated mappings have the same size because each query (*i.e.*, each entity in the source ontology) generates a mapping between the query (source entity) and the top-1 result of the final aggregated rank.

We can observe the supervised methods standing out with the highest results (32 supervised and 04 unsupervised). Among the unsupervised methods, CombANZ is the only one appearing. The LambdaMART method stands out as the supervised method appearing the most. The best mapping results observed ($f - measure > 0.6$) refer to the following language pairs: *de-en*, *en-nl*, *cz-en*, *es-fr*, *cn-nl*, *fr-nl*, and *nl-pt*. Three out of 5 best performing mappings includes the English language. This can be an influence of the pivot language used in the syntactic similarity measures. On the other hand, most of the worst results refer to mappings involving the Russian language. For example, the mapping concerning the language pair *de-ru* led to 0.32 of f-measure.

We compared our results to the ones obtained by related work (ontology alignment systems) presented in OAEI (2018 edition). The competition conducts an

Table 4
Best f-measure results by language pair and rank aggregation algorithm.

Language pair	Precision	Recall	F-measure	Method	Type
de-en	0.75269	0.74468	0.75269	CombANZ	unsupervised
en-nl	0.67742	0.67742	0.67742	ListNet	supervised
cz-en	0.64516	0.64516	0.64516	CoordinateAscent	supervised
es-fr	0.64516	0.64516	0.64516	RandomForest	supervised
cn-nl	0.61290	0.61290	0.61290	CoordinateAscent	supervised
fr-nl	0.61290	0.61290	0.61290	LambdaMART	supervised
nl-pt	0.60215	0.60215	0.60215	GP-FFP1-p500g50B4	supervised
en-pt	0.59140	0.59140	0.59140	LambdaMART	supervised
fr-pt	0.59140	0.59140	0.59140	CoordinateAscent	supervised
cz-de	0.58065	0.58065	0.58065	CombANZ	unsupervised
cz-es	0.58065	0.58065	0.58065	RankBoost	supervised
es-nl	0.58065	0.58065	0.58065	LambdaMART	supervised
es-pt	0.56989	0.56989	0.56989	ListNet	supervised
en-es	0.54839	0.54839	0.54839	MART	supervised
cn-es	0.53763	0.53763	0.53763	RandomForest	supervised
cz-nl	0.52688	0.52688	0.52688	LambdaMART	supervised
cn-fr	0.50538	0.50538	0.50538	CombANZ	unsupervised
cn-pt	0.50538	0.50538	0.50538	LambdaMART	supervised
de-es	0.50538	0.50538	0.50538	ListNet	supervised
de-nl	0.50538	0.50538	0.50538	MART	supervised
en-fr	0.50538	0.50538	0.50538	RandomForest	supervised
cn-en	0.48387	0.48387	0.48387	RandomForest	supervised
cz-pt	0.47312	0.47312	0.47312	MART	supervised
de-fr	0.47312	0.47312	0.47312	RankNet	supervised
cz-fr	0.46237	0.46237	0.46237	RankNet	supervised
nl-ru	0.45161	0.45161	0.45161	MART	supervised
pt-ru	0.44086	0.44086	0.44086	GP-FFP1-p500g50B4	supervised
de-pt	0.43011	0.43011	0.43011	MART	supervised
cn-de	0.40860	0.40860	0.40860	RankBoost	supervised
en-ru	0.39785	0.39785	0.39785	RandomForest	supervised
es-ru	0.38710	0.38710	0.38710	LambdaMART	supervised
fr-ru	0.37634	0.37634	0.37634	MART	supervised
cn-cz	0.35484	0.35484	0.35484	CombANZ	unsupervised
cn-ru	0.32258	0.32258	0.32258	LambdaMART	supervised
cz-ru	0.32258	0.32258	0.32258	LambdaMART	supervised
de-ru	0.32258	0.32258	0.32258	LambdaMART	supervised

evaluation based on a blind dataset[3]. This dataset includes the matching tasks involving the *edas* and *ekaw* ontologies. Because the official results reported on the competition refer to a blind dataset, we were unable to use exactly the same datasets in our evaluation. In this sense, the developed comparison aims to provide some initial analyses to understand the benefits of our developed technique. Our results are compared with the average results reported by the competition based on matching tasks performed with the two ontologies in the blind dataset and our experiments obtained with conference ontology, *i.e.*, conference-edas and conference-ekaw.

Figures 11 presents the language pairs in which our proposal surpasses the state-of-the-art tools (which

[3]OAEI 2018 results for Multifarm: http://oaei.ontologymatching.org/2018/results/multifarm/index.html (As of June 18, 2019).

Table 5

Top-1 of each one of the ranked lists generated by similarity measures for the entity '*Fachgebiet des Gutachters*' for the *conference-conference-de-en* alignment. The correct match is highlighted by (*).

Similarity measure	Query	Top-1 ranked concept
Levenshtein	Fachgebiet des Gutachters (*)	expertise of reviewer (*)
Jaro	Fachgebiet des Gutachters	passive participant of conference
Babelnet	Fachgebiet des Gutachters	None
WordNet	Fachgebiet des Gutachters	None

participated in the OAEI 2018) and Figure 12 reports on the language pairs in which our results were lower. We notice that our average f-measure for all languages reaches a value of 0.50. This is comparable with the results reported by the state-of-the-art tools. Our algorithm proposal surpasses the top result of best tools placed in the competition in 2018 in 27 out of the 36 language pairs.

5. Discussion

Cross-language ontology matching remains an open research problem. The differences between languages and alphabets hamper the application of direct string or synonym comparison approaches, thus increasing the need for novel techniques. Our proposed approach added flexibility enabling *ad hoc* use of similarity measures and external resources. This investigation contributed with encouraging results obtained from experiments with several rank aggregation techniques.

The combination of different similarity measures, both semantic and syntactic, improves the results, because if one measure is not able to properly deal with that particular entity, another similarity measure can perform a balance. Several rank aggregation techniques take this balance into account, resulting in an improved f-measure and a recall better than the average reported in the state-of-the-art.

Table 5 illustrates an example in which the rank aggregation provides benefits to the mapping results. In particular, "*Fachgebiet des Gutachters*" was a difficult concept to match in the alignment *conference-conference-de-en*. It generated **no** results for the two semantic similarity measures. The rank aggregation CombANZ method was able to find the correct mapping candidate.

Although reporting a lower f-measure than the best tool in OAEI 2018 competition, the rank aggregation technique was able to improve the results of *conference-conference-pt-ru* (*cf.* Figure 12) alignment. Table 6 describes this example.

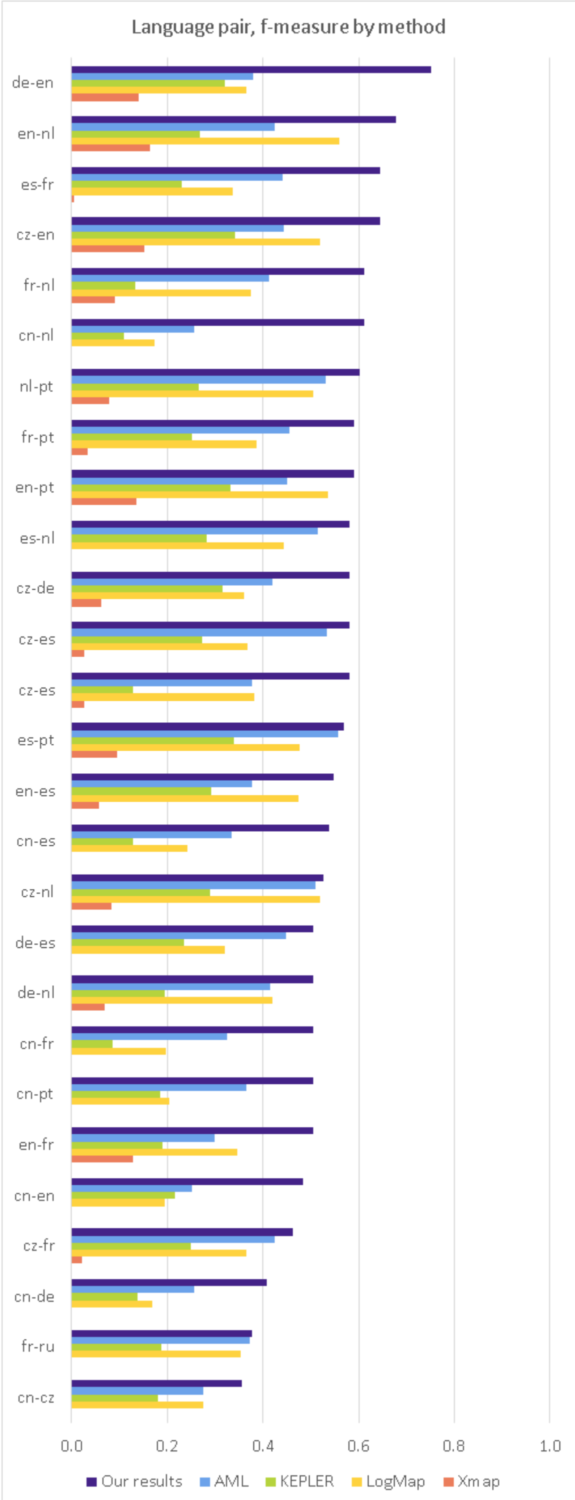

Figure 11. Results where our F-measure was higher than other matching methods.

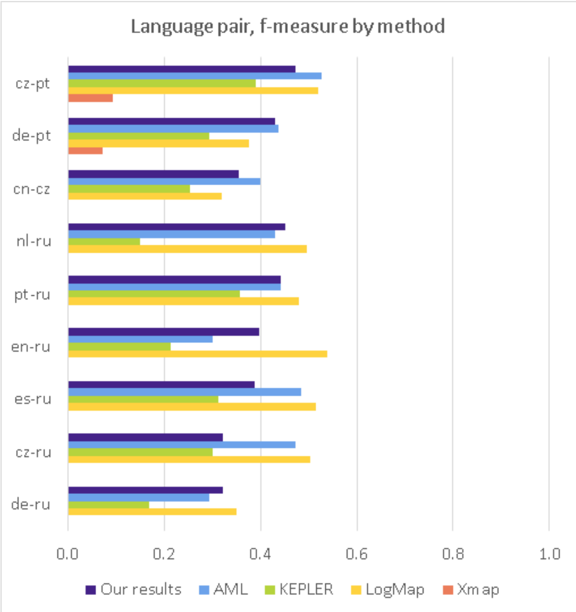

Figure 12. Results where our F-measure was lower than other matching methods.

Table 6

Top-1 of each one of the ranked lists generated by similarity measures for entity *'membro do comitê'* for the *conference-conference-pt-ru alignment*. The correct match is highlighted in the table by (*).

Similarity measure	Query	Top-1 ranked concept
Levenshtein	membro do comitê (*)	Член комиссии (*)
Jaro	membro do comitê	Член комиссии
Babelnet	membro do comitê	None
WordNet	membro do comitê	Организатор

Our conducted experiments using supervised rank aggregation techniques required training and validation set. These sets were generated using a subset of the gold standard for each language pair available. The results reported by supervised rank aggregation methods are directly impacted by the quality of the training and validation sets. The strategy of using training sets with such heterogeneous languages may not bring out the best results for these techniques. The use of larger data sets for training may lead to improved results. This investigation is left for future work.

There is a lack of similarity measures non-dependant on translations. Multilingual semantic networks, such as *Babelnet*, are better to handle the cross-language matching task when compared with string comparison similarity measures. The multilingual semantic networks enables the meaning of words be comparable across languages. This can be extensible to domain-specific networks. For instance, the UMLS could be explored in the context of biomedical ontologies. The use of improved similarity measures may lead to more effective ranked lists, potentially impacting positively both supervised and unsupervised approaches used for cross-language mapping generation.

In summary, obtained results demonstrated the benefits in exploring rank aggregation techniques combined with similarity measures to address the cross-language ontology matching problem. Ontology mappings generated by the presented technique are comparable to state-of-the-art tools. Future works include further exploration of the supervised techniques, with different training sets, crafted to handle specific language pairs instead of all languages at once, and investigate the impact of other similarity measures on the quality of mappings.

6. Conclusion

Alignment of ontologies described in different natural languages remains an open research challenge. In this investigation, we proposed an approach based on rank aggregation methods. Our approach considered four similarity measures to build ranks and used supervised and unsupervised methods of rank aggregation. 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 rank aggregation techniques in the cross-language ontology alignment problem. Future work involves to improve our cross-lingual alignment proposal by considering different combinations of similarity measures and different ways of computing the syntactic and semantic similarities taking into account additional stages in the pre-processing of entity labels. In addition, we plan to investigate the impact of training sets, for instance, training in a subset of languages and testing on a different set.

References

[1] A. Gilchrist, Thesauri, taxonomies and ontologies–an etymological note, *Journal of documentation* **59**(1) (2003), 7–18.

[2] N.F. Noy, D.L. McGuinness et al., Ontology Development 101: A Guide to Creating Your First Ontology, Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA, 2001, Accessed: 2017-07-02.

[3] J. Euzenat and P. Shvaiko, Ontology matching, 2nd edn, Springer, Heidelberg, 2013, pp. 4271–84306.

[4] C. Meilicke, C. Trojahn, O. Šváb-Zamazal and D. Ritze, Multilingual Ontology Matching Evaluation–A First Report on Using MultiFarm, in: *The Extended Semantic Web Conference(ESWC 2012): Satellite Events*, Springer, 2012, pp. 132–147.

[5] 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.

[6] R. Fagin, R. Kumar and D. Sivakumar, Efficient similarity search and classification via rank aggregation, in: *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, ACM, 2003, pp. 301–312.

[7] A. Rexha, M. Dragoni, R. Kern and M. Kröll, An Information Retrieval Based Approach for Multilingual Ontology Matching, in: *Proceedings of the 21st International Conference on Applications of Natural Language to Information Systems (NLDB 2016)*, Springer, 2016, pp. 433–439.

[8] T.R. Gruber, Toward principles for the design of ontologies used for knowledge sharing, *International Journal of Human-Computer Studies* **43** (1995), 907–928, ISSN 1071-5819.

[9] 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.

[10] 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.

[11] 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.

[12] S.K. Stoutenburg, Acquiring advanced properties in ontology mapping, in: *Proceedings of the 2nd PhD Workshop on Information and Knowledge Management (PIKM 2008)*, ACM, 2008, pp. 9–16.

[13] M. Pujari and R. Kanawati, Supervised rank aggregation approach for link prediction in complex networks, in: *Proceedings of the 21st International Conference on World Wide Web*, ACM, 2012, pp. 1189–1196.

[14] K. Subbian and P. Melville, Supervised rank aggregation for predicting influencers in twitter, in: *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, IEEE, 2011, pp. 661–665.

[15] S. Wu, The weighted Condorcet fusion in information retrieval, *Information Processing & Management* **49**(1) (2013), 108–122.

[16] M.N. Volkovs, H. Larochelle and R.S. Zemel, Learning to rank by aggregating expert preferences, in: *Proceedings of the 21st ACM international conference on Information and knowledge management*, ACM, 2012, pp. 843–851.

[17] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma and H. Li, Supervised rank aggregation, in: *Proceedings of the 16th international conference on World Wide Web*, ACM, 2007, pp. 481–490.

[18] M.N. Volkovs and R.S. Zemel, CRF framework for supervised preference aggregation, in: *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, ACM, 2013, pp. 89–98.

[19] M. Kachroudi, G. Diallo and S.B. Yahia, Initiating Cross-Lingual Ontology Alignment with Information Retrieval Techniques, in: *6émes Journées Francophones sur les Ontologies*, 2016, pp. 1–12.

[20] C. Domshlak, A. Gal and H. Roitman, Rank aggregation for automatic schema matching, *IEEE Transactions on Knowledge and Data Engineering* **19**(4) (2007), 538–553.

[21] X. Su and J.A. Gulla, An information retrieval approach to ontology mapping, *Data & Knowledge Engineering* **58**(1) (2006), 47–69.

[22] J. Wang, Q. Li and Y. Dong, A Method of Schema Matching Based on Top-K Mapping and User Feedback, in: *2009 Fifth International Joint Conference on INC, IMS and IDC*, IEEE, 2009, pp. 751–756.

[23] J. David, J. Euzenat, F. Scharffe and C. Trojahn dos Santos, The alignment API 4.0, *Semantic Web Journal* **2**(1) (2011), 3–10.

[24] J.A. Vargas Muñoz, R. da S. Torres and M.A. Gonçalves, A Soft Computing Approach for Learning to Aggregate Rankings, in: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM'2015, ACM, New York, NY, USA, 2015, pp. 83–92. ISBN 978-1-4503-3794-6.

[25] J.-Y. Yeh, J.-Y. Lin, H.-R. Ke and W.-P. Yang, Learning to rank for information retrieval using genetic programming, in: *Proceedings of SIGIR 2007 Workshop on Learning to Rank for Information Retrieval (LR4IR'2007)*, 2007.

[26] E.A. Fox and J.A. Shaw, Combination of Multiple Searches, in: *The Second Text REtrieval Conference (TREC-2)*, 1994, pp. 243–252.

[27] J.C. Borda, *Mémoire sur les élections au scrutin*, Histoire de l'Académie Royale des Sciences, 1781.

[28] G.V. Cormack, C.L.A. Clarke and S. Buettcher, Reciprocal Rank Fusion Outperforms Condorcet and Individual Rank Learning Methods, in: *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'2009, ACM, New York, NY, USA, 2009, pp. 758–759. ISBN 978-1-60558-483-6.

[29] R. Fagin, R. Kumar and D. Sivakumar, Efficient Similarity Search and Classification via Rank Aggregation, in: *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, SIGMOD'2003, ACM, New York, NY, USA, 2003, pp. 301–312. ISBN 1-58113-634-X.

[30] D.C.G. Pedronette and R. da S. Torres, Image Re-ranking and Rank Aggregation Based on Similarity of Ranked Lists, *Pattern Recognition* **46**(8) (2013), 2350–2360, ISSN 0031-3203.

[31] J.H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics* **29**(5) (2001), 1189–1232, ISSN 00905364.

[32] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton and G. Hullender, Learning to Rank Using Gradient Descent, in: *Proceedings of the 22nd International Conference on Machine Learning*, ICML'2005, ACM, New York, NY, USA, 2005, pp. 89–96. ISBN 1-59593-180-5. doi:10.1145/1102351.1102363.

[33] Y. Freund, R. Iyer, R.E. Schapire and Y. Singer, An Efficient Boosting Algorithm for Combining Preferences, *The Journal of Machine Learning Research* **4** (2003), 933–969, ISSN 1532-4435.

[34] J. Xu and H. Li, Adarank: a boosting algorithm for information retrieval, in: *Proceedings of the 30th annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, 2007, pp. 391–398.

[35] D. Metzler and W. Bruce Croft, Linear Feature-based Models for Information Retrieval, *Information Retrieval* **10**(3) (2007), 257–274, ISSN 1386-4564. doi:10.1007/s10791-006-9019-z.

[36] Q. Wu, C.J. Burges, K.M. Svore and J. Gao, Adapting Boosting for Information Retrieval Measures, *Information Retrieval* **13**(3) (2010), 254–270, ISSN 1386-4564. doi:10.1007/s10791-009-9112-1.

[37] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai and H. Li, Learning to Rank: From Pairwise Approach to Listwise Approach, in: *Proceedings of the 24th International Conference on Machine Learning*, ICML'2007, ACM, New York, NY, USA, 2007, pp. 129–136. ISBN 978-1-59593-793-3. doi:10.1145/1273496.1273513.

[38] L. Breiman, Random Forests, *Machine Learning* **45**(1) (2001), 5–32, ISSN 0885-6125. doi:10.1023/A:1010933404324.

[39] P. Christen, A comparison of personal name matching: Techniques and practical issues, in: *Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06)*, IEEE, 2006, pp. 290–294.

[40] V.I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Physics Doklady* **10** (1966), 707–710.

[41] M.A. Jaro, Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida, *Journal of the American Statistical Association* **84**(406) (1989), 414–420.

[42] 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.

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

[44] 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.

[45] C. Meilicke, R. GarcíA-Castro, F. Freitas, W.R. Van Hage, E. Montiel-Ponsoda, R.R. De Azevedo, H. Stuckenschmidt, O. ŠVáB-Zamazal, V. Svátek, A. Tamilin et al., MultiFarm: A benchmark for multilingual ontology matching, *Web Semantics: Science, Services and Agents on the World Wide Web* **15** (2012), 62–68.

[46] 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.