Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 270 (2025) 505–514

Procedia Computer Science

www.elsevier.com/locate/procedia

29th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2025)

SIS: Leveraging Semantically-Informed Similarity of Text Embeddings for Enhanced Ontology Alignment

Giulio Macilenti[a,*], Armando Stellato[a], Manuel Fiorelli[a]

[a]Tor Vergata University of Rome, Via del Politecnico 1, Rome 00133, Italy

Abstract

In recent years, a long-standing task such as Ontology Alignment (OA) has been approached using Large Language Models (LLMs). When taking advantage of the extensive linguistic capabilities of LLMs for OA, a critical challenge arises with respect to the length of text that the model has to process. This issue is significant both because of the computational cost and because of the difficulty of the LLM in maintaining specificity and precision when processing very long prompts. To address these challenges, many existing LLM-based systems for OA use a high-recall filter to reduce the search space before applying the model as a high accuracy evaluation function. These high-recall filters are often constructed using embeddings of the textual information contained within the ontology.

In this context, we propose Semantically-Informed Similarity (SIS), a novel method for extracting and comparing such textual information, leveraging all semantic relations defined as triples. Our method consists of separately embedding list of words that, for each concept, represent the objects of a given predicate. We then compute the SIS similarity as the sum of the cosine similarities of these vectors. We evaluated our method on the OAEI conference track using both SBERT and GloVe embeddings, and compared it against a baseline approach similar to those employed in existing systems. Our results show a significant performance improvement using the SIS method. In the case of SBERT, the high-recall filter achieves remarkable results, with a recall exceeding 90 percent for reasonable parameter settings. Furthermore, we show that the out-performance of our method correlates positively with the level of structuring in the ontologies to be matched.

Keywords: Ontology alignment; Text embeddings; Large Language Models

1. Introduction

Ontology Alignment (OA) aims to identify semantic correspondences between concepts in different ontologies[1, 2]. Although the prospect of adopting a single ontology may offer certain advantages[3], it is important to recognise

* Corresponding author.
E-mail address: giulio.macilenti@students.uniroma2.eu

10.1016/j.procs.2025.09.169

that in a multi-thematic and multi-perspective context, the reality is that different ontologies may still be encountered for the same or overlapping domains [4]. Meaningful alignments are therefore necessary to facilitate interoperability, reusability and modularity of semantic resources. Especially when dealing with large ontologies, a complete or at least partial automation of the alignment process is needed, in order to spare resources and improve precision. However, after decades of research and development, OA remains a challenging task [2, 5].

The first generation of OA systems was rule-based, relying primarily on lexical matching combined with structural information and logical filters[6]. In recent years, several new strategies based on Machine Learning (ML) have emerged. The latest step in this line of research is the use of fine-tuned transformers[7] or instruction-tuned large language models (LLMs)[8] for the OA task.

One key challenge for LLM-based OA systems is the need to process large amounts of textual information with high specificity. However, LLMs often handle long inputs unevenly, potentially overlooking relevant content and reducing performance [9, 10]. To address this, most OA tools using LLMs apply high-recall filters to narrow the candidate space before applying high-precision methods.

Our contribution is situated precisely in this passage: we propose a semantically-informed way to use the information present in the ontologies for the first step of search space reduction. The idea behind our method is simple. The vast majority of alignment systems leverage embeddings, i.e. they represent text related to the semantic resources in vector spaces that are usually high-dimensional, so that proximity in this vector space reflects semantic similarity. However, most existing tools use only the information provided by specific predicates, such as label or comment, or all text related to a resource without any semantic distinction.

Our proposal instead is to separately embed, for each semantic resource, chunks of text that are objects of the same predicate. A similarity score between any pair of concepts can then be computed by summing the cosine similarities of these two sets of vectors, considering only pairs associated with the same predicate. Finally, candidate matches can be selected using a k-nearest neighbors search based on the resulting scores.

Despite its simplicity, we believe that our method can improve the performance of existing LLM-based systems for OA, increasing recall in the first step and allowing a downstream high-precision filter to operate on a significantly smaller search space, that nevertheless contains virtually every target match, thereby preserving overall precision. Furthermore, from a more general point of view, our approach takes an additional step towards the mutual integration of linguistic and semantic resources, using the second level as a higher abstraction layer that can drive the vast linguistic capabilities of LLMs. Beyond the first step of search space reduction, the similarity between the embedding vectors constructed in this way could also be used to compute the final alignment.

We evaluated our method on the shared tasks in the OAEI[1] (Ontology Alignment Evaluation Initiative) conference track, comparing it with a filter that uses the text related to the concepts in the ontologies without considering the semantic relations, which is more similar to those currently used in existing systems. We applied our method to both the GloVe [11] and SBERT [12] embeddings. GloVe (Global Vectors for Word Representation) is an unsupervised learning algorithm for obtaining vector representations for words, which is trained on aggregated global word-word co-occurrence statistics, while Sentence-BERT (SBERT) is a modification of the pretrained BERT network that derives semantically meaningful sentence embeddings relying on the transformer architecture. In our experiments a significant and consistent improvement in performance is observed for both types of embeddings, which demonstrates the validity of the approach and provides a basis for further evaluation and development.

We also investigated whether the performance improvement of our method correlates with the level of specification in the ontologies to be matched. Our hypothesis is that the method performs better when concepts are described in greater detail—i.e., when more semantic relations are present. To test this, we measured the correlation between the performance gain and the average number of relations in the ontology pairs. The results show a positive correlation, supporting our initial assumption.

The rest of the paper is organized as follows. In Section 2 we briefly review related work, in Section 3 we present the details of our method, while the evaluation on the OAEI conference track, along with the study of the correlation with the level of specification of the ontology is presented in Section 4. We conclude in Section 5 with some final remarks and an outlook on future research.

[1] https://oaei.ontologymatching.org/

2. Related work

The objective of our method is to enhance LLM-based alignment systems by incorporating a semantically-informed use of text embeddings as a first step. Consequently in this section on related work we first provide an overview of previous work that uses text embeddings for the OA task. Following that, we discuss recent approaches and studies that use LLMs to perform semantic matching.

2.1. Text embeddings for ontology alignment

To the best of our knowledge, Zhang et al. [13] were the first to introduce word embeddings into the field of ontology alignment. They addressed the limitations of WordNet-based similarity measures [14] due to WordNet's limited coverage of ontology elements. To do this, they trained word2vec embeddings [15] on Wikipedia and used the cosine distance between entity names, entity labels, and entity comments to produce the matches. Their evaluation on the OAEI 2013 benchmark and conference track, as well as on three real-world ontologies, showed that the matcher based on the embeddings outperformed the WordNet-based matchers in all test cases. Dhoulib et al. [16] proposed a new ontology alignment approach that combines FastText [17] word embeddings and the radius measure: the vector representation associated with each concept is computed by averaging the vectors of its labels. This method achieved state-of-the-art performance on an OAEI conference complex alignment benchmark [18] and has been tested on a real-world use case, namely the alignment between the Silex ontology – an ontology describing skills, occupations, and business sectors – and other ontologies in the same domain. Among the other alignment systems, it is worth mentioning DeepAlignment [19] and OntoEmma [20]: both use text embeddings and machine learning to produce matches between concepts. In DeepAlignment, the information contained in the ontologies is used together with external sources to extract synonymy/antonymy relation to improve the pre-trained embeddings for the alignment task. OntoEmma relies on a neural architecture capable of encoding external definitions and context information to enrich entities in an ontology. The evaluation of the system was performed on the OAEI largebio SNOMED-NCI subtask.

All of the approaches mentioned above use word2vec-based embeddings (including fasttext embeddings), which do not take context into account and are therefore unable to capture fine-grained differences within a word's meaning. The transformer architecture [7] gives the possibility to overcome this limitation by producing contextualized embeddings that can distinguish between word senses. BERT [21] is a language model that uses this architecture to produce such contextualized embeddings, which are successfully used to compute semantic similarity between concepts. Unfortunately, ontology alignment may require the representation of whole sentences, not just single words, and BERT is not designed to do this. To solve this issue, Reimers and Gurevych adapted pre-trained BERT to allow for semantically meaningful sentence embeddings that can be compared using cosine similarity. Their Sentence BERT (SBERT) model [12] outperformed other sentence embeddings methods – including GloVe ones[11] – in most of the Semantic Text Similarity (STS) and SentEval tasks, which encompasses binary and multi-class classification and natural language inference.

Regarding the use of transformer-based embeddings for OA, the first contribution to mention is [22]. In this paper, Beutel and Boer proposed and tested several alignment methods, based both on BERT and SBERT embeddings and on the more traditional word2vec ones, and applied them to a real-word use case in the labor market: the automatic alignment of ESCO and O*NET ontologies. The conclusion is that BERT-based embeddings mostly outperform word2vec-based embeddings, but at the same time it is found that their performance is not yet good enough to provide a useful alignment in real-world use cases: approaches combining automatic and manual alignment are needed to improve the results. Always starting from BERT, in [23] He et al. presented BERTMap, a system that predicts mappings using a classifier based on fine-tuning BERT on text semantics corpora extracted from ontologies, and then refines the mappings by extension and repair using the ontology structure and logic. The evaluation performed on part of the OAEI LargeBio Track shows that BERTMap frequently outperforms the state of the art. Finally, it is important to mention TEXTO [24]. In this system, the labels are embedded with GloVe [11], while the description of each concept is embedded with SBERT. Then, a weighted combination of the cosine similarities of these two features is used to produce a score: if this score is above a certain threshold, the match between two entities is accepted. The results of TEXTO are evaluated against the OAEI Common Knowledge Graphs Track, augmented with the description of each class, and a new dataset based on the refreshed alignment of Schema.org and Wikidata, showing good results.

It is worth concluding this section with a general remark. While the use of textual information in ontology alignment is well established—and many existing systems therefore share similarities with SIS—our approach differs in two key respects: it leverages all available textual predicates, unlike methods such as [30] that rely only on labels or descriptions; and it introduces a lightweight, high-recall pre-filter, intended to precede more precise logic- or LLM-based matching stages.

2.2. Large language models for ontology alignment

The breakthrough of large language models is relatively recent, so the literature on their application to the OA task is not as extensive. The first papers that are worth citing are those that directly use few-shot or zero-shot prompting to match two ontologies. Norouzi et al. [25] tested seven different prompt structures on the popular ChatGPT(v4.0) model: both the source and the target ontologies are provided to the model as formatted text, and an alignment between them is requested. The method is evaluated on the OAEI conference track and achieves high recall and low precision. The same authors published another work [26] where they used LLM prompting together with the integration of rich ontology content to produce correspondences that are more complex than simple equivalence. He et al. [27] have also explored LLMs for OA in a zero-shot setting. They tested an open-source LLM, Flan-T5-XXL, and GPT-3.5-turbo on two challenging subsets of the NCIT-DOID and SNOMED-FMA (Body) equivalence matching datasets, both part of the Bio-ML OAEI track. Their results confirmed a weakness in precision, and the conclusion is that LLMs are promising for OA, but several problems need to be addressed, including prompt engineering, formulation of the overall frameworks, and incorporation of ontology contexts.

Precisely to resolve these issues, some more complex approaches have emerged that leverage the contribution of the LLM as part of a layered architecture. These systems are particularly relevant to our work because they always incorporate a high-recall filter based on text embeddings as a first step. For example, in [28], Wang et al. explore the in-context learning ability of LLMs for biomedical concept linking. They first embed concepts using different language models, and then use embedding similarity to retrieve the top candidates. These candidates' contextual information is subsequently incorporated into the prompt and processed by the LLMs to select the matches. The method is applied to the Biomedical Datasets for Equivalence and Subsumption Matching [29], and it shows a competitive performance relative to supervised learning methods. For what concerns our paper, it is important to note that this approach involves generating embeddings only for the canonical concept name string or for a combined version that includes the name string and its context. A similar architecture is used in *OLaLa* [30]. First, matching candidates are extracted from two given input ontologies O1 and O2. The system computes a vector representation of a resource by embedding all text provided by its labels and descriptions using SBERT. It then determines candidate matches for each resource in one ontology by finding the k closest resources in the other ontology based on the cosine similarity of their embeddings. These candidates are then evaluated by an LLM, which can be prompted with either individual candidate matches or a multiple choice between candidate matches for a given resource. Finally, a high-precision matcher, a cardinality filter, and a confidence filter are applied to fulfill the usual requirements for an alignment. *OLaLa* has been tested on different OAEI tracks[31] and has achieved performances comparable to the state of the art.

3. Semantically Informed Similarity (SIS) Approach

Our goal is to produce a high-recall filter to reduce the search space for OA. Improving such a preliminary step in the matching process can help enhance the overall performance of alignment systems, especially those that are LLM-based. SIS was originally conceived for aligning ontologies expressed in OWL; however, its exclusive reliance on the underlying RDF triple structure makes it readily applicable to other scenarios as well, such as instance matching in large-scale knowledge graphs. Our method consists of the following steps: we extract textual information from the matching candidates, we embed it following the predicate structure to obtain a set of vectors, and then compute a similarity function over such a set, which becomes the variable for a k-nearest neighbor search. We will apply this technique to two different embedding models, SBERT and GloVe, which we will then compare in our experiments. We will henceforth refer to this method as SIS: Semantically-Informed Similarity. The way we build our set of vectors in both cases is very similar, so we will first present the procedure with SBERT and then list the differences when using GloVe embeddings.

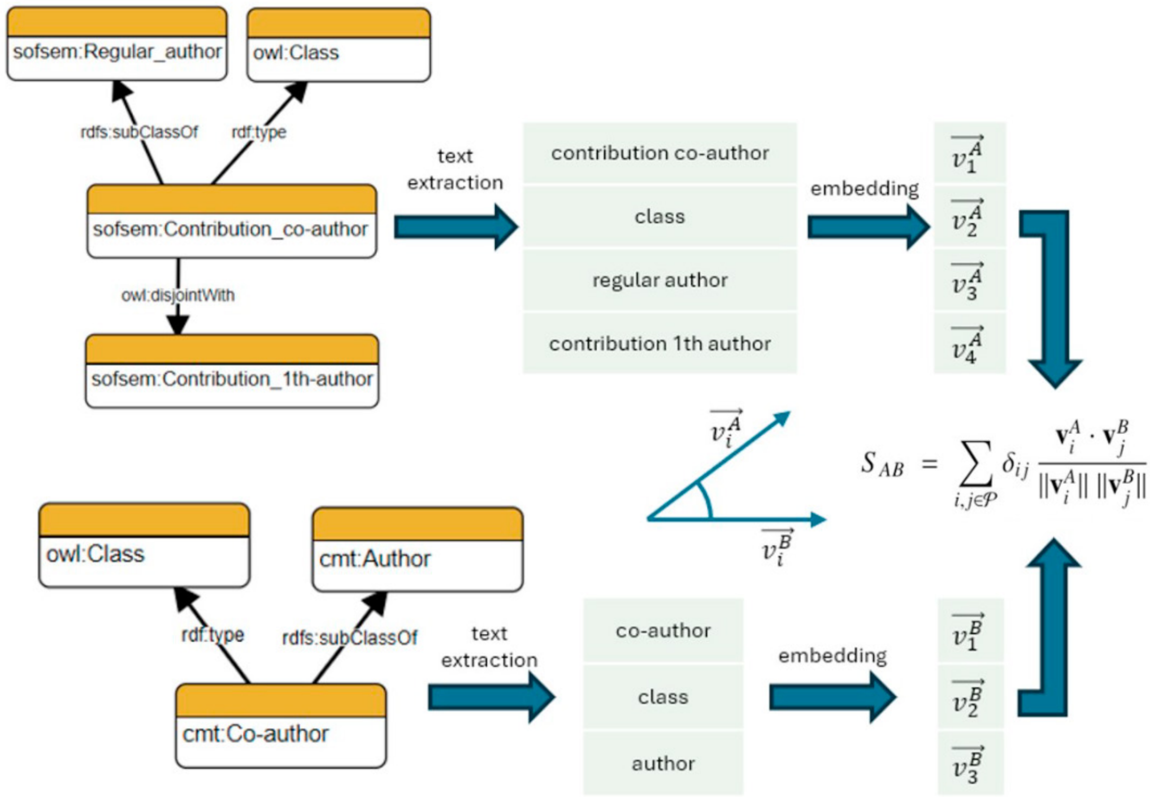

Fig. 1. Example of our alignment method applied to two matching classes from the OAEI conference track ontologies: `Contribution_co-author` from Sofsem and `Co-author` from Cmt.

By first, for each concept in the two ontologies to be aligned, we consider all the triples in which it appears as subject. From this set of triples we create sentences by gathering all the objects that are associated with the same predicate, lexicalizing them and concatenating the produced lexicalizations. The lexicalization function depends on the nature of the object: for literals, their lexical content is returned as-is, for IRIs their labels (rdfs:labels, skos or skos-xl labels, depending on the nature of the ontology) are taken if present, otherwise their local name is considered; blank nodes are ignored in any case. Proceeding in this way, we have chunks of text related to each different predicate used to describe the examined concept. We then perform a step of preprocessing which includes lowering all the text, removing special characters, transforming camelcase representations (as it often happens for IRI local names) into a distinct sets of tokens and finally we embed all the sentences we obtained from such a process with SBERT. There is a possibility that some of the local names obtained through this procedure are not meaningful, introducing noise and lowering performance. In our evaluation this was not significant, since concepts in OAEI ontologies always have local names that embody meaning.

At this point, each concept A in the two ontologies is represented by a set of vectors $\mathbf{v_i}$, labeled with the specific relation to which they are associated, that is, by indicating as $\mathcal{P}$ the set of all IRIs appearing as predicates in both ontologies:

$$\forall A \in (O_1 \cup O_2), \quad \exists \{\mathbf{v_i}\}_{i \in \mathcal{P}}. \tag{1}$$

Now, for each pair of concepts A, B from the ontologies O_1, O_2 we can compute a similarity score as the sum of the cosine similarities between vectors that share the same label. In formula, the similarity between a concept A and a concept B is given by:

$$S_{AB} = \sum_{i,j \in \mathcal{P}} \delta_{ij} \frac{\mathbf{v}_i^A \cdot \mathbf{v}_j^B}{||\mathbf{v}_i^A||\ ||\mathbf{v}_j^B||} \tag{2}$$

where δ_{ij} is the Kronecker delta. Following this procedure, we are essentially summing over the intersection between the predicates associated to the concepts A and B . Given all these S_{AB} scores, the last step is to perform a k-nearest neighbors search, selecting the k concepts with the highest similarity value for each concept.

Our initial idea was to apply our method only to the best performing sentence embedding models, such as SBERT. However, to make the experiments more meaningful, we decided to include not only transformer-based embeddings but also older models like GloVe, allowing us to explore the approach's effects across a broader range of models. We expect to see an increase in performance in both cases. Moreover, since the sentences we build are not actually meaningful sentences but just lists of words, there is a possibility that the best way to embed them is just to compute an average of the embeddings of the individual words. This is precisely what is done when sentences are embedded with GloVe.

We compared our semantically-informed similarity to a method that is more similar to what is actually done in the existing systems, for example in [28, 30]. We called this method Related TeXt (RTX) approach, and it simply consists of using all texts related to a specific semantic resource without differentiating by predicate. The objects of all the predicates are put together in a single sentence which is then transformed into a vector. All other steps, including text preprocessing, were carried out identically. Actually, in many of the cited approaches, the procedure is used only on a specific subset of predicates, but we chose to evaluate our method against a baseline that uses all of them to make the comparison fair and to highlight that the improvement is due precisely to our separate treatment of each semantic relation.

4. Experiments and analysis

Here we report our evaluation. The experiments were carried out on the OAEI conference track[32], which consists of 21 alignments, corresponding to the complete alignment space between 7 ontologies describing the same domain (conference organization). The reference alignment on which we are evaluating is the one referred to as ra1, which can be freely downloaded along with the datasets from the OAEI page. The alignment contains 305 matches over 270759 possible pairs. We chose the conference track because the ontologies within it, although of medium size, do not make extensive use of labels and comments, which are the properties most often used by existing tools for matching: in these cases our method is particularly suitable. In Subsection 4.1 we measured the recall of the SIS approach compared to the RTX baseline, while in Subsection 4.2 we studied the correlation of SIS's outperformance with the level of specification of the ontologies that are aligned.

4.1. Evaluation of SIS on the OAEI conference track

We used GloVe and SBERT embeddings in both SIS and RTX settings on all ontology pairs in the OAEI conference track. The experiments are performed for different values of k, which is the number of neighbors selected in the k-nearest neighbors search. We are analyzing a high-recall filter that precedes the full alignment algorithm; consequently, recall is the primary metric in our evaluation. Precision at this stage is, by construction, low, but the key factor is the extent to which the filter reduces the search space for each value of k. Accordingly, the table reports—for every k—the mean percentage of search-space reduction, calculated as one minus the ratio between the candidate correspondences generated by our method and the total number of entity pairs that can be formed between the two ontologies. The reduction value represents the average across the ontology pairs in the conference track, which naturally differ in size; nonetheless, is worth noting that the individual measurements remain close to that mean.

We computed the recall of SIS and RTX methods over all the alignments of the OAEI conference track: in Table 1 and Figures 2 and 3 we present the results averaged over the whole track, where the error is the standard deviation over the alignments set.

Table 1. Results on the OAEI conference track, for different values of k. Recall is averaged over the alignments in the conference track and the error is the standard deviation. Best results are in bold.

k	Average Search Space Reduction (%)	Recall (Avg. ± Std.)			
Method		**GloVe SIS**	**GloVe RTX**	**SBERT SIS**	**SBERT RTX**
1	99.2%	0.45 ± 0.17	0.12 ± 0.13	**0.65** ± **0.18**	0.12 ± 0.13
2	98.4%	0.53 ± 0.16	0.21 ± 0.15	**0.73** ± **0.18**	0.21 ± 0.15
3	97.6%	0.56 ± 0.14	0.26 ± 0.15	**0.77** ± **0.15**	0.26 ± 0.15
5	95.9%	0.62 ± 0.15	0.35 ± 0.17	**0.85** ± **0.16**	0.35 ± 0.17
8	93.5%	0.63 ± 0.14	0.49 ± 0.19	**0.90** ± **0.12**	0.50 ± 0.18
10	91.9%	0.65 ± 0.14	0.54 ± 0.18	**0.91** ± **0.12**	0.55 ± 0.17
12	90.2%	0.66 ± 0.15	0.60 ± 0.18	**0.92** ± **0.12**	0.61 ± 0.16
15	87.8%	0.69 ± 0.16	0.64 ± 0.18	**0.94** ± **0.11**	0.66 ± 0.16
18	85.4%	0.74 ± 0.16	0.65 ± 0.17	**0.95** ± **0.10**	0.68 ± 0.15
25	79.7%	0.79 ± 0.16	0.70 ± 0.15	**0.95** ± **0.09**	0.73 ± 0.13

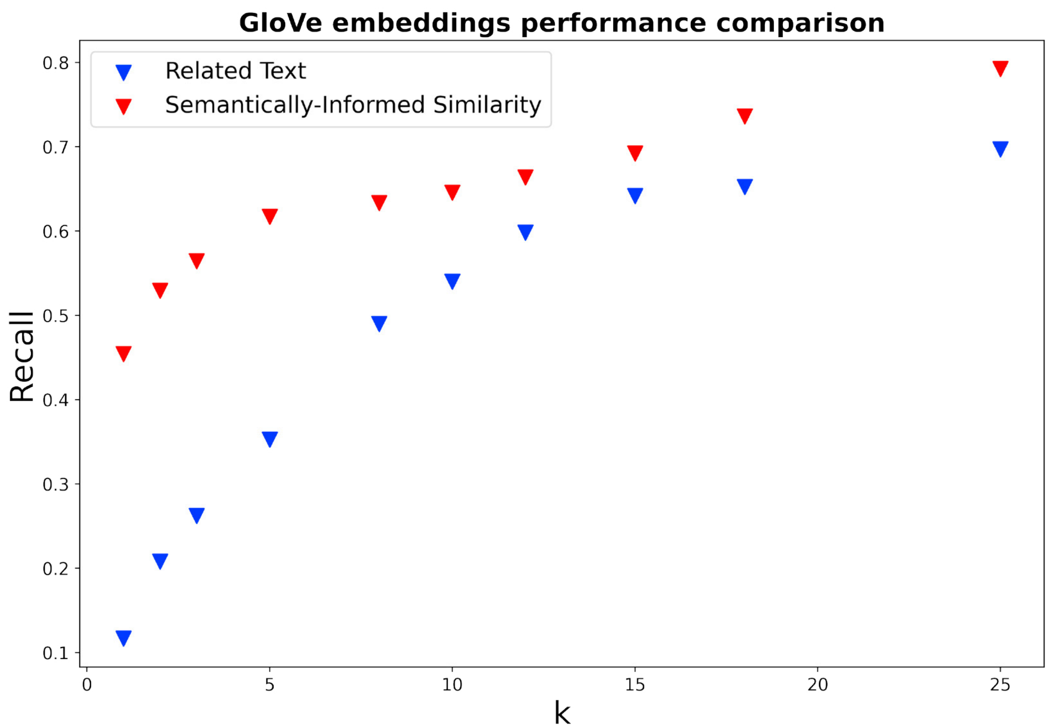

Fig. 2. The recall of GloVe embeddings in Related TeXt (RTX) and Semantically-Informed Similarity (SIS) settings, as a function of k. Data are averaged over all the OAEI conference track.

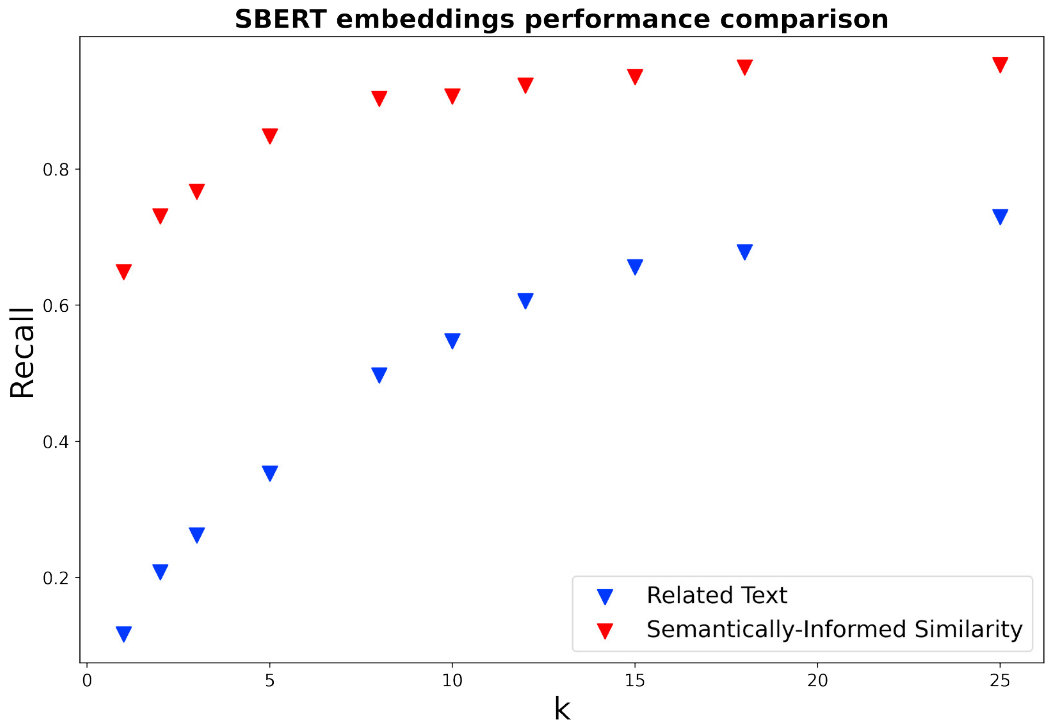

Fig. 3. The recall of SBERT embeddings in Related TeXt (RTX) and Semantically-Informed Similarity (SIS) settings, as a function of *k*. Data are averaged over all the OAEI conference track.

As clearly shown in both the table and the figures, our method improves the performance for all the values of *k* and for both embedding models. In general, the performance improves more for SBERT than for GloVe. For high values of k, with GloVe embeddings, the RTX and SIS performances are quite close, and comparable in error bar. For SBERT, however, the performance obtained with our method is systematically and considerably superior to that obtained by embedding the text without considering the semantic relations. SBERT outperforms GloVe in all test cases, confirming that it is a better model for sentence embedding even when the sentences are constructed as lists of words. In fact, SBERT-SIS achieves the best recall in all our experiments. The recall of SBERT-SIS is higher than that obtained by the alignment systems in the OAEI 2023 conference track, even in the case of k equal to one, which corresponds to only one attempt per entity in the first ontology. Clearly, these results cannot be directly compared without interpretation: high recall is obtained at the cost of significantly lower precision, which is exactly what we expect for a high recall filter. However, it is important to keep in mind that a recall of around ninety percent even for relatively low k values allows any alignment system you choose to apply later, and in particular LLM-based ones, to work with a decidedly lower computational cost and a higher chance of getting a good F1-score. As a final remark, we think that the errors in our experiments carry information as significant as the absolute performance. SBERT-SIS has the lowest percentage errors, that is the lowest ratio between the standard deviation and the mean recall. These percentage errors decrease as the value of k increases: for k greater than five, we have not only excellent results, but also low uncertainty on them when the alignment to be performed varies. This is particularly important, given that we want to have not only a method with good performance, but also reliability of the results when trying to align new ontologies.

Table 2. Comparison of Pearson Correlation Coefficients between recall outperformance ΔR and the product between the average number of predicates per entity in the two ontologies, N_p^{12}. Results for different values of k.

k	GloVe Embeddings	SBERT Embeddings
1	-0.13	0.20
2	0.15	0.42
3	0.06	0.31
5	0.21	0.39
8	0.40	0.59
10	0.40	0.55
12	0.44	0.61
15	0.41	0.57
18	0.25	0.51
25	0.19	0.54

4.2. Correlation of the performance improvement with the level of specification of the ontologies

For a better understanding of our method, we performed an additional analysis. We expect that our idea of separately considering different predicates related to a subject for ontology alignment is more relevant when the ontologies we need to align are specified in greater detail. Driven by this intuition, we expected that the improvement in recall is positively correlated with the level of specification of the two ontologies to align. However, it is not entirely clear how to express this characteristic of a pair of ontologies in quantitative terms. In our case, we have chosen to consider the product of the average number of predicates per entity as the variable capable of expressing the level of semantic information contained in the two ontologies. This quantity can be intuitively linked to the information that any pair of concepts in the two ontologies can share, understood as the probability of having the same predicate referring to both. Under some general assumptions, it can be shown that the overlap probability scales exactly as the product of the average number of predicates per entity. Starting from this idea, we calculated the correlation coefficient between the difference in recall in a SIS and RTX setting, $\Delta R = R_{SIS} - R_{RTX}$, and the product between the average number of predicates per entity in the two ontologies, $N_p^{12} = \langle N_p \rangle_{O_1} \langle N_p \rangle_{O_2}$ As can be clearly seen in Table 2, the correlation between these two variables is always positive, except for the single case of k equal to one with GloVe embeddings, which, however, is not statistically significant. In particular, although the correlation is relatively weak for GloVe, it is more robust for SBERT and becomes quite high for values of k greater than five.

This confirms the soundness of our idea: the method works better the higher the level of specification of the ontologies to be matched. The configuration with the highest correlation is also the one with the best performance in absolute terms, i.e., SBERT-based embeddings with mid-range to high k.

5. Conclusions and future work

In this paper we have introduced a strategy for using the text present in an ontology, in the form of literals or local names, to construct a high-recall filter for ontology alignment. Our idea is rooted in the practice of using textual information from ontologies to produce similarities between concepts: at the same time, to the best of our knowledge, none of the existing systems systematically uses the semantic information encoded by the predicates. The method proposed here, which we called Semantically-Informed Similarity (SIS), consists of separately embedding text chunks that are objects of the same predicate, and then computing a similarity that is the sum of the cosine similarities of this set of vectors. We evaluated it on the OAEI conference track, leveraging GloVe and SBERT embeddings. For both of them, and for any value of the parameters, we observed a strong and consistent improvement in performance over a baseline approach similar to those employed in existing systems.

This result suggests that our method is able to capture more meaningful information than standard ones and can help to improve alignment tools. In particular, with SBERT embeddings the SIS approach shows both high recall (exceeding 90% with relatively small k) and low variability across different ontology pairs. As additional analysis, we studied the correlation between our increase in recall and the level of specification of the ontologies; these two variables are positively correlated, confirming that SIS is especially beneficial when the ontologies to be matched are more detailed and rich in semantic relations. We believe that SIS can be used as a high-recall filter to reduce computational cost problems and hallucinations in LLM-based ontology alignment systems, such as [30]. Furthermore, the decomposition of text embeddings by predicate can potentially inform more refined matching algorithms that combine linguistic similarity with logical constraints and structural reasoning in a multi-layered fashion.

References

[1] Ehrig, M. Ontology alignment: bridging the semantic gap (Vol. 4). Springer Science & Business Media (2006).
[2] Shvaiko, P., & Euzenat, J. Ontology matching: state of the art and future challenges. *IEEE Transactions on knowledge and data engineering*, 25(1), 158–176 (2011).
[3] Hitzler, P., Janowicz, K., Haller, A., & Polleres, A. Are we better off with just one ontology on the Web?. *Semantic Web*, 11(1), 87–99 (2020).
[4] Janowicz, K., van Harmelen, F., Hendler, J. A., & Hitzler, P. Why the Data Train Needs Semantic Rails. *AI Magazine*, 36(1), 5–14 (2015).
[5] Faria, D., Pesquita, C., Mott, I., Martins, C., Couto, F. M., & Cruz, I. F. Tackling the challenges of matching biomedical ontologies. *Journal of Biomedical Semantics*, 9(9), 5–14 (2018).
[6] Rahm, E., & Bernstein, P. A. A survey of approaches to automatic schema matching. *The VLDB Journal*, 10, 334–350 (2001).
[7] Vaswani, A., et al. Attention is all you need. *Advances in Neural Information Processing Systems* 30 (2017).
[8] Zhang, S., et al. Instruction tuning for large language models: A survey. arXiv preprint arXiv:2308.10792 (2023).
[9] Liu, N. F., et al. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics* 12 (2024): 157–173.
[10] Macilenti, G., Stellato, A., Fiorelli, M. Prompting is not all you need: evaluating GPT-4 performance on a real world ontology alignment use case. *Procedia Computer Science* 246C (2024), 1289–1298.
[11] Pennington, J., Socher, R., Manning, C. D. GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP) (2014).
[12] Reimers, N., & Gurevych, I. Sentence-BERT: Sentence embeddings using siamese BERT networks. arXiv preprint arXiv:1908.10084 (2019).
[13] Zhang, Y., Wang, X., Lai, S., He, S., Liu, K., Zhao, J., & Lv, X. Ontology matching with word embeddings. *In Chinese computational linguistics and natural language processing based on naturally annotated big data*, Springer (2014).
[14] Pedersen, T., Patwardhan, S., & Michelizzi, J. Wordnet similarity: Measuring the relatedness of concepts. *Demonstration Papers at HLT-NAACL 2004*. Association for Computational Linguistics (2004), 38–41.
[15] Mikolov, T. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013).
[16] Dhouib, M. T., Zucker, C. F., & Tettamanzi, A. G. An ontology alignment approach combining word embedding and the radius measure. *In International Conference on Semantic Systems*, Springer, Cham (2019).
[17] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5 (2017): 135–146.
[18] Thieblin, E. Task-oriented complex alignments on conference organisation (2019). https://figshare.com/articles/dataset/Complexalignmentdatasetonconferenceorganisation/4986368/8.
[19] Kolyvakis, P., Kalousis, A., & Kiritsis, D. Deepalignment: Unsupervised ontology matching with refined word vectors. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, 787–798.
[20] Wang, L. L., Bhagavatula, C., Neumann, M., Lo, K., Wilhelm, C., & Ammar, W. Ontology alignment in the biomedical domain using entity definitions and context. arXiv preprint arXiv:1806.07976 (2018).
[21] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018).
[22] Neutel, S., & de Boer, M. H. T. Towards Automatic Ontology Alignment using BERT. *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering* (2021).
[23] He, Y., Chen, J., Antonyrajah, D., & Horrocks, I. BERTMap: a BERT-based ontology alignment system. *Proceedings of the AAAI Conference on Artificial Intelligence* 36(5): 5684–5691 (2022).
[24] Peng, Y., Alam, M., & Bonald, T. Ontology Matching using Textual Class Descriptions. *Proceedings of the 18th International Workshop on Ontology Matching co-located with the 22nd International Semantic Web Conference (ISWC 2023)*, Athens, Greece, November 7, 2023.
[25] Norouzi, S. S., Mahdavinejad, M. S., & Hitzler, P. Conversational Ontology Alignment with ChatGPT. CoRR abs/2308.09217 (2023).
[26] Amini, R., Norouzi, S. S., Hitzler, P., & Amini, R. Towards Complex Ontology Alignment using Large Language Models. arXiv preprint arXiv:2404.10329 (2024).
[27] He, Y., et al. Exploring large language models for ontology alignment. arXiv preprint arXiv:2309.07172 (2023).
[28] Wang, Q., Gao, Z., & Xu, R. Exploring the In-context Learning Ability of Large Language Model for Biomedical Concept Linking. arXiv preprint arXiv:2307.01137 (2023).
[29] He, Y., Chen, J., Dong, H., Jiménez-Ruiz, E., Hadian, A., & Horrocks, I. Machine Learning-Friendly Biomedical Datasets for Equivalence and Subsumption Ontology Matching. *The Semantic Web – ISWC 2022*, Springer, Cham (2022), 575–591.
[30] Hertling, S., & Paulheim, H. OLaLa: Ontology matching with large language models. *Proceedings of the 12th Knowledge Capture Conference 2023* (2023), 131–139.
[31] Hertling, S., & Paulheim, H. OLaLa results for OAEI 2023. *OM@ISWC* (2023).
[32] Zamazal, O., & Svátek, V. (2017). The ten-year ontofarm and its fertilization within the onto-sphere. *Journal of Web Semantics*, 43, 46-53.