

# Supporting Ontology Maintenance with Contextual Word Embeddings and Maximum Mean Discrepancy

Natasha Shroff<sup>1</sup>, Pierre-Yves Vandenbussche<sup>2</sup>, Véronique Moore<sup>2</sup>, and Paul Groth<sup>1</sup>

<sup>1</sup> University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands

<sup>2</sup> Elsevier Amsterdam, Radarweg 29a, 1043 NX Amsterdam, Netherlands

**Abstract.** Ontologies contain an abundance of concepts, are frequently structured as hierarchies, and can cover different domains of knowledge. Polysemous concepts need to be disambiguated for annotation purposes, for example, a concept such as *depression* has a different meaning in the fields of psychology and economics. In this paper, we introduce the use of the maximum mean discrepancy to indicate whether sets of concepts sharing the same meaning should be merged. This method is a novel approach to ontology maintenance because it provides an objective metric that supports the decision-making of subject matter experts during the concept evaluation process. Our objective is thus to assist ontology maintenance, in particular the organization of concepts, through an analysis framework that gives insights into the polysemy of concepts.

**Keywords:** Ontology maintenance · Contextual embeddings · NLP

## 1 Introduction

Ontologies, as defined by Gruber, are an explicit "specification of a conceptualization" [10] and provide a formal representation of knowledge across domains [14]. They serve to create a shared understanding of concepts between humans and information systems and provide the required structure for highly-efficient information storage, representation, and retrieval [33, 30, 7]. To remain up to date with domain developments, ontologies need to be frequently maintained and updated. However, current ontology maintenance tools are unable to fully offer insights into the polysemy of a concept, e.g., what *depression* means in *economics* versus *psychology*. Furthermore, tools are not able to accurately indicate if two similar concepts should be merged, under which conditions, or whether a concept does not belong to a given domain altogether [19, 26, 2]. Consequently, ontology curators find it difficult to have the best possible, unambiguous, representation of their domains of interest.

An important observation is that ontology concepts are frequently associated with corresponding text. The meaning of this text can be captured through the use of distributional semantics [16]. Given the linkage between concept and

text, we seek to apply advances in NLP based on distributional semantics for addressing ontology maintenance problems, such as the accurate organization of polysemous concepts into corresponding domains [8]. We study the potential of this approach on a large-scale production ontology - OmniScience [18]. Elsevier’s OmniScience ontology is used to generate annotations for research documents and links the metadata of multiple products in different domains together.

The contributions of this paper are as follows:

- The introduction of the maximum mean discrepancy (MMD) scores in combination with contextual word embeddings as a tool for ontology maintenance;
- The analysis of this approach on a large-scale real world production ontology.

## 2 State of the Art

Recent research in distributional semantics [16] has highlighted improvements in word sense disambiguation on polysemous words [35, 34] using Transformer-based models [31] such as BERT [5]. BERT uses bidirectional representations that help to capture the context of a sentence more accurately than the unidirectional language models such as Word2Vec [20], GloVe [22], or ELMo [23]. BERT produces contextual embeddings that are based on various semantic contexts of a word. While this is effective for polysemy, others have improved on this approach by developing polyseme-aware vector representation models that disambiguate polysemous words more effectively [11].

Recent literature on the automation of ontology maintenance shows that the manual evaluation of concepts with semantic similarity is still required as even state-of-the-art maintenance tools are not able to accurately map ambiguous concepts [12, 6]. In the field of ontology merging and mapping, several recently published studies should be highlighted. While our approach concerns the merging of two concepts into one as part of one ontology, advances in merging two or more ontologies are of relevance to this topic. CoMerger, a tool to automatically merge multiple ontologies in an efficient, scalable, and customizable manner, fulfills user requirements such as the quality assessment of the merged ontology and offers a check for compatibility of the ontologies to be merged [2]. The method is based on partitioning and first divides all ontologies to be merged into multiple blocks, followed by refinements per block according to the previously set merge requirements. Other merging tools have different approaches such as functional merging of ontologies through rules [32], letting users broadly define how to merge the ontology [21] or using granular computing techniques [24] that produce multiple levels of descriptions to merge two ontologies. Few ontology merging tools are made available open-source for easy use and, to the best of our knowledge, are limited to the matching of taxonomic structures rather than capturing the semantic meaning of concepts inside the ontology.

Recently published works such as Alin [4] in the field of ontology mapping and matching, mention an *interactive ontology matching process* in which an ontology matching tool is used by a domain expert to improve the accuracy of the matching results. The domain expert makes use of the tool to receive a

preliminary mapping of the merged ontologies and then gives feedback on the mappings to either accept or reject them. Common methods to conduct ontology matching are the use of classifiers [28], using a threshold to reject or accept mappings [4] and using domain experts to calculate such a threshold [15]. Since most accurate ontology matching tools still rely on the expertise and feedback of human domain experts, the challenge to accurately identify polysemous concepts for ontology maintenance using embedding models remains unsolved [13].

### 3 Method

Our approach consists of the selection of polysemous and synonymous ontology concepts and their sentences from the corpus of scientific text data. This was followed by the creation of corresponding contextual word embeddings, experimentation, and evaluation of the MMD scores for ontology maintenance.

**Data extraction.** The OmniScience ontology used for this study includes more than 769.000 concepts and covers 20 scientific domains. Each concept in the ontology is mentioned in a range of scientific articles that belong to a separate database. We first created a dataset of polysemous and synonymous concepts by extracting all sentences in which selected concepts were matched exactly from a corpus of 18 million scientific full-text articles. Since this paper focuses on the automatic recognition of concept sets that should be merged, we did not extract their taxonomic relations from the ontology and instead only took the concepts and the sentences in which they occurred into account. The sentences were matched to the concept using the Aho-Corasick algorithm on lowercase pattern matching [1]. Because the algorithm searches for exact matches, it did not return plural forms or other lemmas.

**Contextual embeddings.** Following the sentence extraction, corresponding contextual embeddings for each sentence were created with SciBERT[3], a language model based on BERT that was pretrained on a large corpus of 1.14 million scientific documents. We used the SciBERT-uncased-large model as recommended by Beltagy et al. [3]. The SciBERT embeddings were used for the experimentation with the MMD scores as described in the next paragraph.

**Ontology maintenance.** We experimented with the MMD to investigate the extent to which it can support ontology maintenance. The MMD was mainly used to calculate the distance between distributions of sentences containing synonymous concepts that should be merged. It measures the distance between distributions through the distance of mean embeddings from two distributions [29]. The MMD score calculation is based on probability measures in the Reproducing Kernel Hilbert Space (RKHS) [9]. Equation 1 shows the general MMD in which the kernel is  $k$  and  $\mathcal{H}$  stands for the RKHS.  $P$  is the probability measure, that is the mean element of the kernel mean. They result in the following formula, which is equal to the distance between the two distributional means [29].

$$MMD_k P, Q = \| \mu_P - \mu_Q \|_{\mathcal{H}}. \quad (1)$$

An MMD score of 0 means that the distributions are equal, a score of 1 or higher means that the distributions are separated.

## 4 Experiments

To test whether the MMD score could serve as a measure to indicate a merge for ontology concepts, a set of 117 synonymous concept pairs was extracted and analyzed. In addition, MMD scores for five polysemous concepts were computed to understand how the metric would perform for concepts with different meanings across domains. The sample size for the score computations was always 1000. This means that many datasets were upsampled to 1000, even if they initially only contained 100 sentences. We calculated a ratio of the sample size per MMD score to provide additional information around the score reliability and increase internal validity. A ratio of 1 is a perfect case scenario in which the datasets both contain at least 1000 sentences. The lower the ratio, the higher the difference in the number of original sentences, and the more up-sampling was performed to receive the required sample size. While up-sampling lead to overall less reliable scores, it was necessary for the creation of this larger evaluation dataset.

During the analysis of MMD scores, three categories were specified:

1. Red scores: MMD bigger than 0.2
2. Yellow scores: MMD between 0.15 and 0.2
3. Green scores: MMD smaller than 0.15

The color-coding helped distinguish high from low scores and supported the manual analysis of example cases to understand why some scores were higher than others. The category bounds were chosen based on the results of initial MMD score computations.

## 5 Results and Discussion

In this section, the MMD scores and their application for ontology maintenance are presented.

### 5.1 Organizing concepts

We calculated the MMD scores for sentences that contained a polysemous concept first, to get an indication of how far apart the distributions are in different contexts. A selection of MMD scores calculated with randomly selected distributions can be found in Table 1. We provide the open-source code to the research community for the MMD Score calculation on GitHub [27].

We found that MMD scores below 0.2 indicate that sentences with a polysemous concept such as *depression* in medicine and psychology are used in a similar context in both domains. Scores for two polysemous concepts in different domains well above 0.2 such as *coding* in biochemistry and computer science or

Table 1: MMD Scores for polysemous words.

Concept	Selected Domains	MMD Score	Samples	Ratio
Coding	Biochemistry, Comp. Science	0.6272	1000, 1000	1
Coding	Biochemistry, Medicine	0.3614	1000, 1000	1
Depression	Medicine, Psychology	0.1670	1000, 1000	1
Depression	Medicine, Earth Sciences	0.8137	1000, 1000	1

*depression* in medicine and earth sciences indicate a weak semantic relationship and indicate that concepts should be organized under different domains. We can conclude that MMD scores can indicate the ambiguity of domains, clarify the semantic distance between polysemous concepts, and can provide guidance for manual ontology curation of concepts across domains.

## 5.2 MMD score evaluation

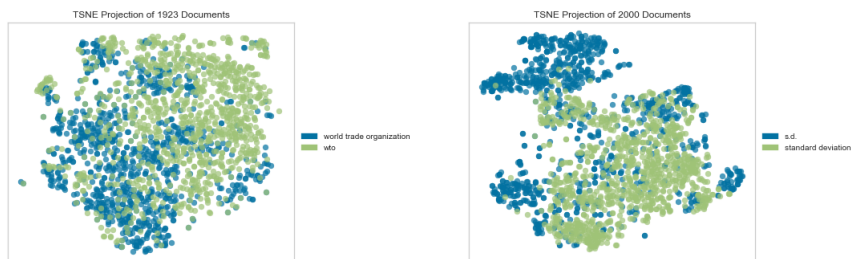
To investigate whether the scores can indicate if overlapping concepts should be merged, an evaluation dataset containing 117 pairs of previously merged economic synonyms was created as described in Section 3. A selection of scores is shown in Table 2.

Table 2: MMD Scores for sets of two synonyms from the same domain.

Synonyms	Domain	MMD Score	Samples	Ratio
Cost Benefit Analysis, Cost Benefit	Economics	0.0141	743, 1553	0.743
Risk Modeling, Risk Modelling	Economics	0.1314	238, 200	0.840
Gross National Income, GNI	Economics	0.1678	322, 1255	0.322
Standard Deviation, s.d.	Economics	0.3697	66953, 2611	1

We found that 67% of distributions with green MMD scores below 0.15 have a sample size ratio above 0.4. Half of the distributions with yellow scores between 0.15 and 0.2 on the other hand, contained sample size ratios below 0.3 and are less reliable than those with a higher sample size ratio. Finally, distributions with red scores above 0.2 contained significantly lower sample size ratios: 70% of scores had a ratio lower than 0.3. Lower sample size ratios often occurred in sets made up of a concept and its acronym, e.g. *world trade organization* and *wto* and due to the more frequent use of the acronym.

To validate MMD scores, we have computed t-SNE [17] corpus visualization maps. Figure 1 visualizes the corpus for two concept sets in two dimensions. We can infer that two synonyms’ embeddings are less likely to overlap as the MMD score increases. This finding confirms that the MMD score provides a valid measure of the distance between distributions and helps experts decide whether concepts should be merged. With the current combination of manual and automated ontology maintenance, our automated analysis framework for



(a) MMD score: 0.1953,  
green: wto,  
blue: world trade organization

(b) MMD score: 0.3697,  
green: standard deviation,  
blue: s.d.

Fig. 1: t-SNE corpus visualization of SciBERT embeddings per concept set

semantically similar concepts in the ontology can help curators decide on the merging of concepts. The accuracy and efficiency of the decision-making process to assess the semantic cohesion of concepts are hereby enhanced through the use of the MMD analysis. For instance, experts can use MMD scores, given a sample size ratio of 0.4 or higher, as a measure to confirm a merge of concepts and save time during the manual evaluation process.

### 5.3 Limitations

Regarding the usage of the MMD score to merge pairs of synonyms, there are a few limitations to take into consideration. First, the MMD score evaluation showed that not all scores were as low as expected for pairs of synonyms. Therefore, the need for a balanced and saturated dataset of 1000 sentences per synonym is considered a limitation that needs to be solved to arrive at a reliable MMD score. Furthermore, we assume that the extracted sentences are associated to the concepts that are mentioned. Since the MMD score gives an indication of how closely related two ontology concepts are, we create the dataset based on occurring instances in the corpus of data. Finally, the current MMD pipeline is resource and memory intensive, which might lead to computational infrastructure issues when applying the code on larger sample sizes.

## 6 Conclusion

The experiments in this study served to develop a method for ontology maintenance through the usage of contextual embeddings and the distance between their distributions. With the MMD scores, we can determine whether two separate concepts should be merged. Moreover, we found that the MMD score can

indicate the ambiguity of a concept within domains. This can help subject matter experts with their manual evaluation and give more insights into a concept's polysemy.

We see two main avenues for future work. First, the current research scope is limited to a case study with five polysemous concepts and 117 previously merged synonym pairs. A list of ambiguous OmniScience concepts in sets of two could be run through the analysis pipeline to carry out a systematic evaluation over a large sample. Concepts with MMD scores below 0.15 could be examined by ontology curators to confirm or reject a merge. Second, recent improvements to contextual embedding models such as Sentence-BERT [25] could provide a better semantic capture of ontology concepts within a sentence and is therefore suggested for future works.

## References

1. Aho, A.V., Corasick, M.J.: Efficient string matching: an aid to bibliographic search. *Communications of the ACM* **18**(6), 333–340 (1975)
2. Babalou, S., König-Ries, B.: Towards Building Knowledge by Merging Multiple Ontologies with CoMerger: A Partitioning-based Approach (may 2020)
3. Beltagy, I., Cohan, A., Lo, K.: SciBERT: Pretrained Contextualized Embeddings for Scientific Text (mar 2019)
4. Da Silva, J., Revoredo, K., Baião, F.A., Euzenat, J.: Alin: improving interactive ontology matching by interactively revising mapping suggestions. *Knowledge Engineering Review* **35**, e1 (2020)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (oct 2018)
6. Dragisic, Z., Ivanova, V., Lambrix, P., Faria, D., Jiménez-Ruiz, E., Pesquita, C.: User validation in ontology alignment. In: *International Semantic Web Conference*. pp. 200–217. Springer (2016)
7. Gao, W., Zhu, L., Guo, Y., Wang, K.: Ontology learning algorithm for similarity measuring and ontology mapping using linear programming. *Journal of Intelligent & Fuzzy Systems* **33**, 3153–3163 (oct 2017). <https://doi.org/10.3233/JIFS-169367>
8. Gargouri, Y., Lefebvre, B., Meunier, J.g.: Ontology maintenance using textual analysis. In: *Proc. 7TH World Multi Conference on Systemics, Cybernetics and Informatics, USA. List of Figures Figure*. vol. 1 (2003)
9. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. *Journal of Machine Learning Research* **13**(25), 723–773 (2012), <http://jmlr.org/papers/v13/gretton12a.html>
10. Gruber, T.R.: A translation approach to portable ontology specifications. *Knowledge Acquisition* **5**(2), 199–220 (1993), <http://www.sciencedirect.com/science/article/pii/S1042814383710083>
11. Guo, S., Yao, N.: Polyseme-aware vector representation for text classification. *IEEE Access* **8**, 135686–135699 (2020). <https://doi.org/10.1109/ACCESS.2020.3010981>
12. Harrow, I., Balakrishnan, R., Jimenez-Ruiz, E., Jupp, S., Lomax, J., Reed, J., Romacker, M., Senger, C., Splendiani, A., Wilson, J., Woollard, P.: Ontology mapping for semantically enabled applications. *Drug Discovery Today* **24**(10), 2068–2075 (2019), <http://www.sciencedirect.com/science/article/pii/S1359644618304215>

13. Khadir, A.C., Aliane, H., Guessoum, A.: Ontology learning: Grand tour and challenges. *Computer Science Review* **39**, 100339 (2021). <https://doi.org/https://doi.org/10.1016/j.cosrev.2020.100339>, <https://www.sciencedirect.com/science/article/pii/S1574013720304391>
14. Konys, A.: Knowledge systematization for ontology learning methods. *Procedia Computer Science* **126**, 2194–2207 (2018), <http://www.sciencedirect.com/science/article/pii/S1877050918312043>
15. Li, C., Cui, Z., Zhao, P., Wu, J., Xin, J., He, T.: Improving ontology matching with propagation strategy and user feedback. In: *Seventh International Conference on Digital Image Processing (ICDIP 2015)*. vol. 9631, p. 963127. International Society for Optics and Photonics (2015)
16. Lüddecke, T., Agostini, A., Fauth, M., Tamosiunaite, M., Wörgötter, F.: Distributional semantics of objects in visual scenes in comparison to text. *Artificial Intelligence* **274**, 44–65 (sep 2019), <https://www.sciencedirect.com/science/article/pii/S0004370219300177>
17. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *Journal of machine learning research* **9**(Nov), 2579–2605 (2008)
18. Malaisé, V., Otten, A., Coupet, P.: *OmniScience and Extensions – Lessons Learned from Designing a Multi-domain, Multi-use Case Knowledge Representation System BT - Knowledge Engineering and Knowledge Management*. pp. 228–242. Springer International Publishing, Cham (2018)
19. McDaniel, M., Storey, V.C., Sugumaran, V.: Assessing the quality of domain ontologies: Metrics and an automated ranking system. *Data Knowl. Eng.* **115**, 32–47 (2018)
20. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013)
21. Noy, N.F., Musen, M.A.: The prompt suite: interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies* **59**(6), 983–1024 (2003). <https://doi.org/https://doi.org/10.1016/j.ijhcs.2003.08.002>, <https://www.sciencedirect.com/science/article/pii/S107158190300137X>
22. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pp. 1532–1543 (2014)
23. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. *arXiv preprint arXiv:1802.05365* (2018)
24. Priya, M., Cherukuri, A.K.: An approach to merge domain ontologies using granular computing. *Granular Computing* **6** (01 2021). <https://doi.org/10.1007/s41066-019-00193-3>
25. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics (11 2019), <http://arxiv.org/abs/1908.10084>
26. Saeeda, L., Med, M., Ledvinka, M., Blaško, M., Křemen, P.: *Entity Linking and Lexico-Semantic Patterns for Ontology Learning BT - The Semantic Web*. pp. 138–153. Springer International Publishing, Cham (2020)
27. Shroff, N., Vandenbussche, P.Y., Moore, V., Groth, P.: *Ontology-maintenance-with-mmd*. <https://github.com/curiousseikatsu/Ontology-Maintenance-with-MMD> (2021)



28. Silva, J., Baiao, F.A., Revoredo, K.: Interactive ontology alignment using alignment antipatterns: A first experiment. In: Proceedings of the XII Brazilian Symposium on Information Systems on Brazilian Symposium on Information Systems: Information Systems in the Cloud Computing Era - Volume 1. p. 208–215. SBSI 2016, Brazilian Computer Society, Porto Alegre, BRA (2016)
29. Tolstikhin, I., Sriperumbudur, B.K., Schölkopf, B.: Minimax Estimation of Maximum Mean Discrepancy with Radial Kernels. Tech. rep.
30. Vallet, D., Fernández, M., Castells, P.: An ontology-based information retrieval model. In: European Semantic Web Conference. pp. 455–470. Springer (2005)
31. Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. (2017)
32. Verhodubs, O.: Merging of ontologies through merging of their rules (2020)
33. Wong, W., Liu, W., Bennamoun, M.: Ontology Learning from Text: A Look Back and into the Future. *ACM Computing Surveys - CSUR* **44**, 1–36 (jan 2011). <https://doi.org/10.1145/2333112.2333115>
34. Xypolopoulos, C., Tixier, A.J.P., Vazirgiannis, M.: Unsupervised word polysemy quantification with multiresolution grids of contextual embeddings (2021)
35. Yenicelik, D., Schmidt, F., Kilcher, Y.: How does BERT capture semantics? a closer look at polysemous words. In: Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP. pp. 156–162. Association for Computational Linguistics, Online (Nov 2020). <https://doi.org/10.18653/v1/2020.blackboxnlp-1.15>, <https://www.aclweb.org/anthology/2020.blackboxnlp-1.15>