Ontology-Based Word Sense Disambiguation for Scientific Literature

Roman Prokofyev[1], Gianluca Demartini[1], Alexey Boyarsky[2,3,4], Oleg Ruchayskiy[5], and Philippe Cudré-Mauroux[1]

[1] eXascale Infolab, University of Fribourg—Switzerland
{firstname.lastname}@unifr.ch

[2] Ecole Polytechnique Fédérale de Lausanne—Switzerland
{firstname.lastname}@epfl.ch

[3] Instituut-Lorentz for Theoretical Physics, U. Leiden—The Netherlands

[4] Bogolyubov Institute for Theoretical Physics, Kiev—Ukraine

[5] CERN TH-Division, PH-TH, Geneva—Switzerland
oleg.ruchayskiy@cern.ch

Abstract. Scientific documents often adopt a well-defined vocabulary and avoid the use of ambiguous terms. However, as soon as documents from different research sub-communities are considered in combination, many scientific terms become ambiguous as the same term can refer to different concepts from different sub-communities. The ability to correctly identify the right sense of a given term can considerably improve the effectiveness of retrieval models, and can also support additional features such as search diversification. This is even more critical when applied to explorative search systems within the scientific domain.
In this paper, we propose novel semi-supervised methods to term disambiguation leveraging the structure of a community-based ontology of scientific concepts. Our approach exploits the graph structure that connects different terms and their definitions to automatically identify the correct sense that was originally picked by the authors of a scientific publication. Experimental evidence over two different test collections from the physics and biomedical domains shows that the proposed method is effective and outperforms state-of-the-art approaches based on feature vectors constructed out of term co-occurrences as well as standard supervised approaches.

1 Introduction

The number of scientific papers getting published is rapidly increasing. To support the discovery of new scientific results as well as exploratory endeavors within a new field of interest, modern systems rely on annotated collections of scientific papers. One example of such systems is PubMed, which uses the MeSH taxonomy[6] to annotate the topic of a scientific paper and to enable search over

[6] http://www.nlm.nih.gov/mesh/

annotations. Annotations are usually created manually by the authors when creating or publishing a new document but can also, in some cases, be generated automatically, especially when performed at word-level. One example of such an automatic annotation system for scientific papers is ScienceWISE[7], which automatically annotates papers from the physics domain adopting an expert-curated ontology as background information. Another example is Utopia[8], a system integrating visualization and data-analysis features that has been used by the editors of the Biochemical Journal (BJ) in a successful pilot.

In automatic annotation systems, most annotations errors are originating from ambiguous terms, which may lead to the wrong concepts being identified. One example in the Physics domain is the term ‘cluster’ which may refer to a ‘cluster of galaxies’ or to a ‘cluster of stars’, which are two very different concepts. Usually, the right sense can be identified by the reader and by automated approaches using the context (e.g., the paper topic). In other cases, it might be necessary to be aware of some particular background knowledge related to the specific research topic addressed in the paper. While an expert in the field might be able to determine the right word sense in a scientific article thanks to his professional background, automatic approaches often fail to disambiguate the terms correctly without such knowledge.

For this reason, we propose in this paper a semi-supervised method for Word Sense Disambiguation (WSD) for the scientific literature domain. The task we address is the disambiguation of scientific terms and acronyms used in scientific abstracts. Our approach is based on the use of both contextual information from the document as well as a background knowledge-graph built and maintained by the scientific community. While no manually created annotated data is necessary to train our models, the proposed approach is semi-supervised in the sense that it exploits existing relations among concepts in a background ontology that can either be manually or automatically generated.

We experimentally evaluate our approach over two different test collections, one based on the ScienceWISE Web portal used to semantically annotate, bookmark, and share papers in the Physics domain, and one based on the MeSH index for the MEDLINE corpus in the biomedical domain [8].

The main contributions of this paper are:

- the definition of a WSD task for a collection of scientific abstracts that are semantically annotated via a background ontology;
- novel efficient and effective approaches to WSD that exploit both collection statistics as well as concept relations in the background ontology graph;
- a new test collection for WSD over a background ontology graph and its concept relations;
- an experimental comparison of the proposed approach against prior WSD approaches over two different test collections, showing that our ontology-based methods are both more effective and more efficient than state-of-the-

[7] `http://sciencewise.info`
[8] `http://getutopia.com/`

art approaches based on context vectors or automated classification when relying on a high-quality ontology.

The rest of this paper is organized as follows: Section 2 describes previous work in the area of WSD. We define the problem of WSD for semantically annotated scientific papers and propose a new approach leveraging collection statistics and relations among existing concepts in Section 3. Section 4 describes our experimental setting and presents the results of a series of experiments comparing our approach to existing WSD methods. Finally, we conclude in Section 5.

2 Related Work

In this paper, we target the scenario of WSD for scientific document collections. This is a compelling research topic, especially when considered in the context of online digital libraries offering metadata about scientific publications, like Bibsonomy [6] or ScienceWISE. The general problem of Word Sense Disambiguation has been widely studied in the past (see [10] for a survey). Both supervised and unsupervised approaches to WSD have been proposed.

Supervised approaches consider an initial set of training examples over which a model to disambiguate terms in documents is learned. A popular approach is Naïve Bayes [2], which is known to be effective but not particularly efficient. Other more efficient supervised methods based on Support Vector Machines have been proposed as well [9]. In this paper, we propose a semi-supervised method that does not require training evidence but that is based on existing relations among domain concepts within a manually curated ontology graph. We also compare our method against k-Nearest Neighbor (kNN), which is one of the most effective supervised approaches to WSD [3].

Knowledge-based methods are directly related to the approach we propose in this paper. Knowledge-based methods adopt background information to select the correct sense of a term in a document. The most popular resource used by such approaches is WordNet [4], a machine-readable lexicon of word senses and linguistic relations. While very useful, the disadvantage of such a general-purpose resource lie in its lack of domain-specific information. In a recent paper [11], Navigli *et al.* propose a similar approach to supervised WSD based on text classification that also exploits a WordNet graph as background information. Our approach is different in the sense that it is able to exploit *domain-specific* rather than general-purpose ontologies and does not require any training.

Another approach which, similarly to ours, proposes a method that leverages both a background knowledge-base as well as corpus statistics is [7]. In that piece of work, the authors propose the use of a machine-readable dictionary over which similarity values are computed and used for clustering terms. On the other hand, our work aims at analyzing semantic relations among terms in the ontology in order to understand the intended meaning of a term. Our experiments also show higher accuracy values as compared to [7].

Standard test collections exist to evaluate and compare WSD approaches. In this paper, we use an existing collection for WSD in the context of scientific documents which is based on the MeSH vocabulary [8]. Additionally, we create a novel test collection specifically targeting the scenario where a community-maintained background ontology as well as expert generated ground truth annotations are available.

3 Graph-Based Disambiguation Models

3.1 Ontology-based WSD: Task Definition

The task we are focusing on in this paper is Word Sense Disambiguation given a domain-specific ontology $O = \{C, R\}$ containing concepts C and relations R among them. In the context of this paper, we use "term" to denote a single word (separated by white spaces or any other punctuation symbol) and "concept" to denote the set of n-grams ($n \geq 1$) that define all possible forms of a concept in the ontology (e.g., "Milky Way Halo", "MW Halo"). We define the set $C_U \subset C$ as the set containing the n-grams that occur only once across the ontology (assuming that each concept in the ontology has at least one unique form, so-called *main form*) and a set $C_A \subset C$, which contains all the n-grams that occur more than once, such that $C_U \cup C_A = C$ and $C_U \cap C_A = \emptyset$.

The ontology is used to identify and extract concepts from textual documents: Given a document collection $D = \{d_1, .., d_n\}$, we extract from each document d_i a list of concepts $c_1, .., c_n$ based on normalized n-gram matching. In some ambiguous cases, an extracted n-gram may refer to different concepts in the ontology. In such cases, we define the WSD task for an ambiguous n-gram as the selection of the right concept in the ontology among a list of candidate-matching concepts. Details on the concept extraction process are provided in Section 4.4.

3.2 Concept Context Vectors for WSD

The first approach we adopt for WSD over a scientific document collection is based on *context vectors*, which is a commonly used unsupervised approach to WSD (see, for example, [1]). A context vector $\boldsymbol{cv}(c_i)$ for a concept $c_i \in C_A$ is defined as $\boldsymbol{cv(c_i)} = \{(t_j, score_j)|t_j \in T\}$, where T is the space of all terms from the document collection. Such vectors may either contain binary values indicating whether t_j co-occurs or not in the same documents as c_i, or more informative values such as the frequency score of such co-occurrences.

In this paper, we define and use an extension of context vectors which—instead of using all words in the document context—first identifies concepts in d based on the background ontology using entity linking methods (e.g., [5]). Thus, we define a *Concept Context Vector* (CCV) $\boldsymbol{ccv(c_i)}$ for a concept $c_i \in C_A$ as $\boldsymbol{ccv(c_i)} = \{(c_j, score_j)|c_j \in C_U\}$. The only difference with classic context vectors is that instead of considering all possible words in the textual context, we restrict our analysis on the co-occurrence between concepts described in the ontology. An example of CCVs from our test collections is shown in Table 1.

Similarly, a *Document Concept Context Vector* (DCCV) $\boldsymbol{dccv(d_i)}$ is a vector consisting of all the concepts identified in a document $d_i \in D$. Examples of

Table 1. Examples of CCVs (main form) from the ScienceWISE collection.

Concept	CCV(Concept)
Star formation efficiency	(Instability, 4), (Supernova, 2), (Milky Way, 3), ...
Support vector machine	(Bayesian, 1), (Neural network, 2), (Classification, 11), ...
Markov decision process	(Probability , 10), (Reinforcement learning, 4), ...

DCCVs are provided in Table 2. We define $\boldsymbol{dccv(d_i)} = \{(c_j, score_j)|c_j \in C_U\}$. Once CCVs and DCCVs have been constructed, it is possible to perform WSD by means of a similarity score between CCVs of the candidate matching concepts and the DCCV where the ambiguous concepts have been identified. In this paper, we rank candidate CCVs by cosine similarity scores with the target DCCV.

Table 2. Examples of DCCV from the ScienceWISE collection.

DocID	DCCV(DocID)
1	(Milky Way, 1), (Electron neutrino, 1), (Electron antineutrino, 1), ...
2	(Local analysis, 1), (Poynting-Robertson effect, 1), (White dwarf, 3), ...

3.3 Graph-based Approaches to WSD

Assuming that an ontology storing domain concepts and their relations is available, it is possible to define advanced WSD methods that exploit such relations as well. Let us define a graph $O = \{C, R\}$ where nodes $c \in C$ are concepts in the ontology and edges $r_l(c_i, c_j) \in R$ represent the labeled relations between different concepts.

A first possible WSD method (minDist) that exploits such an additional structure is based on the distance between concepts in the graph. Given an ambiguous n-gram and its candidate matching concepts $CC = \{cc_1 \ldots cc_n\}$ we select one sense based on the minimum distance with respect to all the other concepts $DC = \{dc_1 \ldots dc_n\}$ present in d, where the distance between two concepts $dist(cc_i, dc_j)$ is given by the shortest path connecting them in the O graph:

$$score(cc_i) = \min_{dc_j \in DC} dist(cc_i, dc_j) \tag{1}$$

A different approach (Ontology Shortest-Path, OSP) is also based on the ontology graph, but ranks candidate concepts based on the average distance to all concepts in d:

$$score(cc_i) = \frac{\sum_{dc_j \in DC} dist(cc_i, dc_j)}{|DC|} \tag{2}$$

Finally, the third approach (NN) we explore in this paper is based on the neighborhood of the candidate matching concepts given the ontology O. Thus, the confidence score to rank a candidate concept cc_i for a document d is given by the number of co-occurring neighbors c_j of cc_i in d:

$$score(cc_i) = |\{c_j | c_j \in DC \wedge dist(cc_i, c_j) = 1\}| \qquad (3)$$

Those three techniques to score and rank candidate matching concepts are experimentally compared over two different test collections in Section 4.

3.4 Combination of WSD approaches

The approaches to WSD described so far provide a score (e.g., similarity score among vectors) that indicates the confidence level of the disambiguation. Therefore, it is possible to combine different approaches together, for instance using a simple linear combination of their confidence scores and thus reach potentially better decisions based on multiple evidences. In this paper, we adopt a mixture model among pairs of approaches A and B to circumvent the problem of having several parameters to learn at once:

$$score(cc_i) = \alpha score_A(cc_i) + (1 - \alpha) score_B(cc_i), \alpha \in [0, 1] \qquad (4)$$

4 Experimental Evaluation

4.1 Experimental Setting

We evaluated the proposed models over two different test collections: one based on the MSH collection [8] and another one from the ScienceWISE system. Both collections contain a set of abstracts from scientific publications. In most cases, online digital libraries let guest users or crawlers only access the abstracts of the papers they store. Hence, we decided to restrict the document corpus to those abstracts only.

We consider four baseline approaches in the following. The first baseline approach we use for comparison is the random baseline (that is commonly used for comparison in WSD, see for instance [10]), which randomly assigns one among all the possible senses to the ambiguous n-gram. The more ambiguous the n-gram, the less effective this random baseline gets. Another simple baseline we consider is to always select the most frequent sense (as appearing in the document collection) among the candidate matching senses. We also compare the CCV-based approach against standard context vectors constructed over all the terms appearing in the document instead of only considering the extracted concepts. The fourth baseline we use for comparison is the state-of-art supervised method based on Naive Bayes (NB) classification. We train it over 7'641 and 2'952 manually disambiguated documents for the MSH and ScienceWISE collection respectively.

The evaluation measures commonly used for WSD are Precision and Coverage (i.e., Recall). As our approaches always retrieve a sense for each ambiguous n-gram extracted from the abstracts, we only report Precision values in the following. To compare different approaches and to validate potential improvements, we measure statistical significance by means of a paired t-test considering a difference significant when $p < 0.05$. We describe the two document collections we used for our experiments below.

4.2 MSH Collection

The first document collection we use for evaluating our approaches consists of abstracts from the biomedical domain [8]. Each element of the test collection represents one ambiguous n-gram, its corresponding abstract and the correct sense among all the available senses. The test collection also contains all possible senses for each n-gram in a separate file.

To build the appropriate Concept Context Vectors for the MSH collection, we used the RESTful text annotator service offered by bioontology.org [9]. As a backend ontology for the annotation process, we used the Medical Subject Headings (MeSH) ontology[10], which is used by MEDLINE indexers to annotate the textual contents of biomedical articles. To focus exclusively on important concepts, we also filtered out short one-word concepts (e.g.: *cell, administration*) (we manually experimented with different thresholds for filtering out one-word concepts and got the best results by filtering out ones that are shorter than 14 characters). Overall, 8'782 different concepts and 11'797 different n-grams were extracted. After this preprocessing step, 38'025 distinct relations among concepts were created.

4.3 ScienceWISE Collection

The second collection we consider is a testset for WSD we created based on public data obtained from the ScienceWISE system. The ScienceWISE system allows a community of scientists, working in a specific domain, to generate dynamically as part of their daily work a field-specific ontology with direct connections to research papers and scientific data management services. The two main functionalities of ScienceWISE are *annotations* (i.e., adding meta-data to scientific documents) and *semantic bookmarking* (i.e., creating virtual collections of research papers from arXiv).

The domain-specific ontology is central to the system and allows to integrate all heterogeneous pieces of data and content shared by the users. The initial version of the ontology was created by performing a semi-automated import from many science-oriented ontologies and online encyclopedias. After this step, ScienceWISE users (who are domain experts) were allowed to edit elements of the ontology (e.g., adding new definitions or new relations) in order to improve both its quality and coverage. Presently, the ScienceWISE ontology, which is publicly available in RDF[11], counts more than 60'000 unique entries, each with its own definitions, alternative forms, and semantic relations.

Using documents and human-created annotations over the ScienceWISE ontology, we created a testset for WSD. The generated test collection contains 1) a set of 4'691 abstracts from the Physics domain, 2) a set of 5'217 disambiguation decisions performed by experts in the Physics domain, and 3) the version

[9] `http://bioportal.bioontology.org/annotator`,
REST API description: `http://rest.bioontology.org/`

[10] The exact version of the ontology we used is *2012_2011_09_09*.

[11] `http://data.sciencewise.info/`

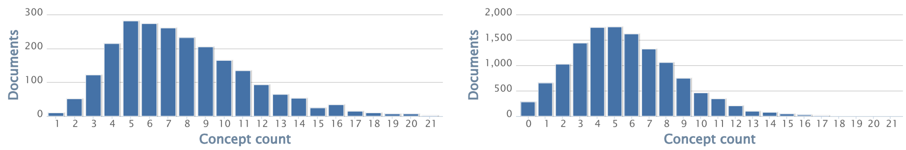

Fig. 1. Distribution of concepts per document in the ScienceWISE (left) and MSH (right) collections.

of the ScienceWISE ontology as of October 2012 which has been used for our domain-specific ontology-based WSD approach.

Formally, a test collection TC is represented as the following set of tuples: $TC = \{(d, c_a, c_u) | d \in D, c_a \in C_A, c_u \in C_A\}$, where c_a and c_u represent the ambiguous and unambiguous (main) forms of the same concept respectively. Both collections with detailed descriptions are available online at `http://exascale.info/papers/ecir2013disambig` for reproducibility purposes.

4.4 Concept Extraction and Distribution

Given the plain text abstracts and the corresponding ontology, we build the DCCVs for both collections as follows. First, we create an index from all the scientific concepts appearing in the collection ontology by considering stemming (Porter stemming algorithm) and stopword removal. Then, we process each abstract and match its textual contents to the concept index using an efficient and exact string matching method and using TF-IDF as scoring function. The final distributions of concepts for the MSH and the ScienceWISE documents are depicted in Figure 1. As we can observe, most paper abstracts contain 5-6 concepts in the ScienceWISE collection and 4-5 concepts in the MSH collection.

4.5 Experimental Results

In scientific articles, acronyms are often used to shorten commonly used concepts across the document. Usually, such acronyms are defined the first time they appear in the paper (e.g., Color Dipole Model (CMD)). Those occurrences make it easy to automatically detect the right sense of such ambiguous acronyms by simply using regular expressions to look for the definition given before or after the brackets. Using simple regular expressions, we discovered that we can directly solve 56% of the cases in the ScienceWISE collection and 67% in the MSH collection. Thus, we divide our test collection TC into 2 sub-collections $TC_R \cup TC_U = TC$ that represent the sub-collection containing the cases that can be simply resolved and the other cases respectively. For this reason, we report in the following the effectiveness of the proposed methods on the sub-collection TC_U. The supervised NB method is trained over the sub-collection TC_R.

Table 3 gives the effectiveness values for CCV approaches as compared to our baselines for the two test collections. Among the baselines, we observe that the supervised NB performs best on ScienceWISE.

Table 3. Precision of context vector based WSD over the subset of concepts that cannot be disambiguated using regular expressions. We indicate statistical significant improvements over NB with * and over the best unsupervised baseline with $^+$.

WSD Approach	Prec (ScienceWISE)	Prec (MSH)
Random	39.97	46.73
Most Frequent	74.46	43.60
Context Vectors	74.29	**95.29**
NB	**85.13**	67.31
TF-IDF CCV	80.72^+	90.46^*
Binary CCV	$\mathbf{93.34}^{*+}$	$\mathbf{90.77}^*$

Analyzing the results of the proposed ontology-based approaches, we see that CCVs outperforms basic unsupervised approaches and is comparable to supervised approaches (i.e, NB). Specifically, we note that CCVs outperforms standard Context Vectors in terms of effectiveness while also being more efficient in terms of indexing as the term space is considerably reduced since it only considers concepts in the ontology instead of all terms in the document collection.

On the MSH collection, Context Vectors perform best overall. This can be explained by the relatively low quality of its background ontology which has been automatically constructed. On the other hand, in ScienceWISE the ontology is manually built and curated by a community of domain experts which makes the approaches exploiting such information perform best. This hypothesis is supported when looking at the Precision/Coverage graph (Figure 2), where we observe that by lowering the coverage of matching concepts, Precision of CCV becomes bigger than Precision of CV also for the MSH collection.

Moreover, we note that for the ScienceWISE collection, the simpler Binary CCV approach (which considers only binary values in the vectors indicating co-occurrences) performs better than the TF-IDF CCV method, which instead uses TF-IDF scores for the concept context vectors. TF-IDF CCVs performs best however on the MSH collection, albeit by a small margin (less than 0.3%).

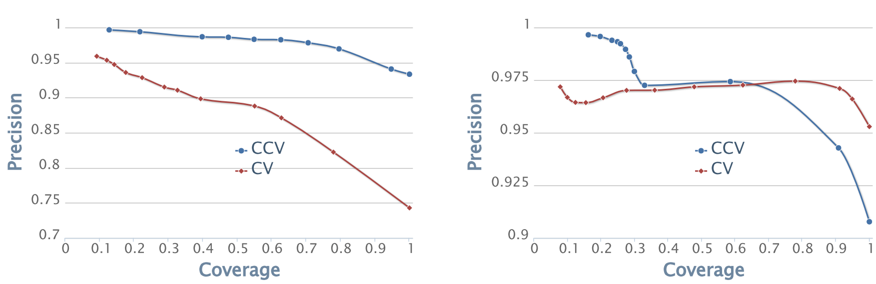

Fig. 2. Precision/Coverage graphs for CV and Binary CCV methods over the ScienceWISE (left) and MSH (right) collections.

Table 4. Precision of ontology graph based WSD.

WSD Approach	Prec (ScienceWISE)	Prec (MSH)
minDist (with Cat)	**88.82**	-
OSP (with Cat)	86.46	-
NN (with Cat)	73.93	-
minDist (without Cat)	**82.84**	67.28
OSP (without Cat)	77.42	56.77
NN (without Cat)	73.93	**72.37**

Table 4 presents the results for WSD approaches based on the ontology graph. While they also outperform unsupervised and, in some cases, supervised baselines, they are not better than CCV-based approaches. Among the graph-based approaches, the NN method yields the best effectiveness over the MSH collection while NN performs best on ScienceWISE. For the ScienceWISE dataset, we run our approaches over two different versions of the ontology graph: one that includes the edges about category information (similarly to Wikipedia articles and categories) and one containing exclusively edges that relate concepts to each other (note that for MSH the category information is not available). We observe that considering category links provides better WSD effectiveness. Thus, we only report results using the more complete ontology graph for ScienceWISE in the following.

Next, we evaluate the combination of ontology-based approaches to WSD. Specifically, Table 5 shows the combination of Binary CCV with methods based on the ontology graph. The methods are combined using the model from Equation 4 using equal weights for all the components. As we can see, the combination of CCVs and graph-based NN methods outperforms both individual approaches on the ScienceWISE collection. On the MSH collection, no significant improvement is observed.

Table 5. Precision of combined WSD semantic approaches. We indicate statistically significant improvement over Binary CCV with *.

WSD Approach	Prec (ScienceWISE)	Prec (MSH)
Binary CCV	93.34	**90.77**
+ minDist (with Cat)	92.68	77.56
+ OSP (with Cat)	94.44	80.77
+ NN (with Cat)	**94.53***	90.60

Parameter Sensitivity in the Mixture Model. As described in Section 3.4, we combine different approaches by considering a linear combination of their confidence scores. Figure 3 gives the results of a parameter sensitivity analysis we performed for such combinations. The figure shows precision values for the combination of the Binary CCV method with two different approaches exploiting the ontology graph, namely, NN and OSP. As we can see, optimal effectiveness values are obtained when more weight is put on matching evidence coming from the graph. Considering an equal weight is hence somewhat suboptimal, though it also results in high effectiveness values.

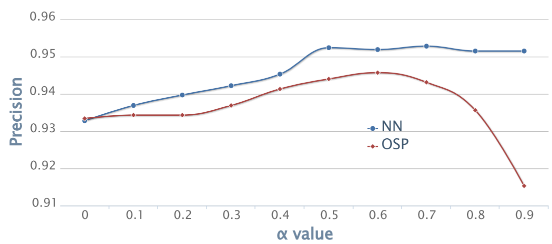

Fig. 3. Precision values varying the α parameter of the mixture model in Equation 4.

Efficiency Considerations. As very large collections of scientific documents are available in digital libraries, in addition to WSD effectiveness we are also interested in how efficient our methods are when deployed in large-scale, real settings. Table 6 reports the execution times of different WSD methods over the two test collections.

Table 6. Execution time of different WSD approaches over the two test collections.

WSD Approach	**Exec Time (ms) (ScienceWISE)**	**Exec Time (ms) (MSH)**
Context vectors	14'712 (+30 min indexing)	1'826 (+1 h indexing)
CCV	1'682 (+2 min indexing)	1'476 (+5 min indexing)
NN	35'363	41'947

We observe that the running times of graph-based approaches are higher than those only relying on vector similarities, mainly because of the costly access times to the database system. However, the preparation of both term vectors and concept vectors requires considerable time, which is not needed by the graph-based approaches.

5 Conclusions

Scientists originating from different sub-communities often use the same term to refer to different concepts, making it hard to automatically process their articles using simple NLP or indexing techniques. In this paper, we tackled the problem of correctly disambiguating terms appearing in the abstract of scientific publications using a series of techniques ranging from relatively simple approaches (e.g., most common sense) to several variants of context vectors and to a series of new ontology-based approaches we devised for this work.

While creating and maintaining a field-specific ontology represents a huge effort, more and more scientific portals rely on such ontologies to organize their contents (the two ontologies we used in the context of this paper are good examples of that trend). Following our experiments, we observe that such ontologies can represent crucial information when building word sense disambiguation systems, for two main reasons: i) ontologies typically regroup the most important terms of a scientific domain and can thus be used to build more efficient and effective context vectors based on ontologic concepts only and ii) the structure of the ontology can be leveraged to devise new techniques for WSD, for example using distance measures or nearest-neighbors on the ontology graph. Combining concept context vectors and graph-based approaches yields the best results according to our experiments, where our combined methods outperform both Bayes classifiers and conventional context vectors when leveraging on a high-quality and relatively complete ontology.

References

1. Khaled Abdalgader and Andrew Skabar. Unsupervised similarity-based word sense disambiguation using context vectors and sentential word importance. *ACM Trans. Speech Lang. Process.*, 9(1):2:1–2:21, May 2012.
2. Rebecca F. Bruce and Janyce M. Wiebe. Decomposable modeling in natural language processing. *Comput. Linguist.*, 25(2):195–207, June 1999.
3. Walter Daelemans, Antal Van Den Bosch, and Jakub Zavrel. Forgetting exceptions is harmful in language learning. *Mach. Learn.*, 34(1-3):11–41, February 1999.
4. C. Fellbaum. Wordnet. *Theory and Applications of Ontology: Computer Applications*, pages 231–243, 2010.
5. Xianpei Han, Le Sun, and Jun Zhao. Collective entity linking in web text: a graph-based method. In *SIGIR*, pages 765–774, New York, NY, USA, 2011. ACM.
6. A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Bibsonomy: A social bookmark and publication sharing system. In *Proceedings of the Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures*, pages 87–102, 2006.
7. Antonio Jimeno Yepes and Alan R. Aronson. Knowledge-based and knowledge-lean methods combined in unsupervised word sense disambiguation. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, pages 733–736, New York, NY, USA, 2012. ACM.
8. Antonio José Jimeno-Yepes, Bridget T. McInnes, and Alan R. Aronson. Exploiting mesh indexing in medline to generate a data set for word sense disambiguation. *BMC Bioinformatics*, 12:223, 2011.
9. Yoong Keok Lee and Hwee Tou Ng. An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation. In *ACL-02 - Volume 10*, EMNLP '02, pages 41–48, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.
10. Roberto Navigli. Word sense disambiguation: A survey. *ACM Comput. Surv.*, 41(2):10:1–10:69, February 2009.
11. Roberto Navigli, Stefano Faralli, Aitor Soroa, Oier de Lacalle, and Eneko Agirre. Two birds with one stone: learning semantic models for text categorization and word sense disambiguation. In *CIKM*, pages 2317–2320, New York, NY, USA, 2011. ACM.