Author manuscript, published in "Lecture notes in computer science (2010) 24-35"

Sense-Based Biomedical Indexing and Retrieval

Duy Dinh and Lynda Tamine

University of Toulouse,
118 route de Narbonne, 31062 Toulouse, France
{dinh,lechani}@irit.fr

Abstract. This paper tackles the problem of term ambiguity, especially for biomedical literature. We propose and evaluate two methods of Word Sense Disambiguation (WSD) for biomedical terms and integrate them to a sense-based document indexing and retrieval framework. Ambiguous biomedical terms in documents and queries are disambiguated using the Medical Subject Headings (MeSH) thesaurus and semantically indexed with their associated correct sense. The experimental evaluation carried out on the TREC9-FT 2000 collection shows that our approach of WSD and sense-based indexing and retrieval outperforms the baseline.

Keywords: Word Sense Disambiguation, Semantic Indexing, Term Weighting, Biomedical Information Retrieval.

hal-00553710, version 1 - 10 Jan 2011

1 Introduction

Nowadays, the volume of biomedical literature is constantly growing at an ever increasing rate. The exploitation of information from heterogeneous and diverse medical knowledge sources becomes a difficult task for any automated methods of data mining and searching. In general, biomedical texts are expressed in documents using natural human language, which causes the common problem of *ambiguity*. Indeed, human natural language is ambiguous by its nature. For example, the term "gold" may be a noun or adjective depending on the context where it appears. The noun may refer to a yellow metallic element in chemistry, while the adjective may refer to the characteristic of something having the color of this metal. Ambiguity is a common problem in the general text as well as in the specific domain such as biomedicine. For instance, polysemous words that are written and pronounced the same way as another, but which have different senses, frequently indicate both a gene and encoded protein, a disease and associated proteins ... Such polysemous words are called ambiguous. Meanings or senses of a given word or term are usually defined in a dictionary, thesaurus, ontology and so on. Recognizing and assigning the correct sense to these words within a given context are referred to as Word Sense Disambiguation (WSD).

Many investigations on WSD in general text have been done during the last years. Current approaches to resolving WSD can be subdivided into four categories: *Knowledge-based* [1, 2, 3], *Supervised* [4, 5], *Unsupervised* [6] and *Bootstrapping* methods [7]. Knowledge-based approaches use external resources such

C.J. Hopfe et al. (Eds.): NLDB 2010, LNCS 6177, pp. 24–35, 2010.
© Springer-Verlag Berlin Heidelberg 2010

hal-00553710, version 1 - 10 Jan 2011

as Machine Readable Dictionaries (MRDs), thesauri, ontologies, etc. as lexical knowledge resources. Supervised machine learning (ML) methods (*Decision Trees, Naive Bayes, Vector Space Model, Support Vector Machines, Maximum Entropy, AdaBoost ...*) use manually annotated corpus for training classifiers. Unsupervised classifiers are trained on unannotated corpora to extract several groups (clusters) of similar texts. Bootstrapping (semi-supervised) approaches rely on a small set of seed data to train the initial classifiers and a large amount of unannotated data for further training. Most of the developed systems for WSD of biomedical text are based on the supervised ML approach [5, 8, 9], which depends totally on the amounts and quality of training data. This constitutes the principal drawback of those approaches: they require a lot of efforts in terms of cost and time for human annotators.

This paper proposes a sense-based approach for semantically indexing and retrieving biomedical information. Our approach of indexing and retrieval exploits the poly-hierarchical structure of the Medical Subject Headings (MeSH) thesaurus for disambiguating medical terms in documents and queries. The remainder of this paper is organized as follows. Section 2 presents related work on WSD in biomedical literature. In section 3, we detail our methods of WSD for biomedical terms in documents. Section 4 deals with the document relevance scoring. Experiments and results are presented in section 5. We conclude the paper in section 6 and outline some perspectives for our future work.

2 Word Sense Disambiguation in Biomedical Text

Term ambiguity resolution in the biomedical domain becomes a hot topic during the last years due to the amount of ambiguous terms and their various senses used in biomedical text. Indeed, the UMLS[1] contains over 7,400 ambiguous terms [10]. For instance, biomedical terms such as *adjustment, association, cold, implantation, resistance, blood_pressure, etc.* have different meanings in different contexts. In MeSH, a concept may be located in different hierarchies at various levels of specificity, which reflects its ambiguity. As an illustration, figure 1 depicts the concept "Pain", which belongs to four branches of three different hierarchies whose the most generic concepts are: *Nervous System Disease* (C10); *Pathological Conditions, Signs and Symptoms* (C23); *Psychological Phenomena and Processes* (F02); *Musculoskeletal and Neural Physiological Phenomena* (G11).

Biomedical WSD has been recently the focus of several works [11, 5, 12, 8, 9, 13, 14, 15]. Indeed, related works can be subdivided into two main categories: *Knowledge-based* and *Supervised-based* WSD. In the knowledge-based approach, ambiguous terms are processed using knowledge sources such as MRDs, thesauri, ontologies, list of words or abbreviations together with their meanings, etc. For example, UMLS contains information about biomedical and health related concepts, their various names, and the relationships among them, including the Metathesaurus, the Semantic Network, and the SPECIALIST Lexicon and Lexical Tools. Humphrey *et al.* [11] employed the Journal Descriptor Indexing

[1] Unified Medical Language System.

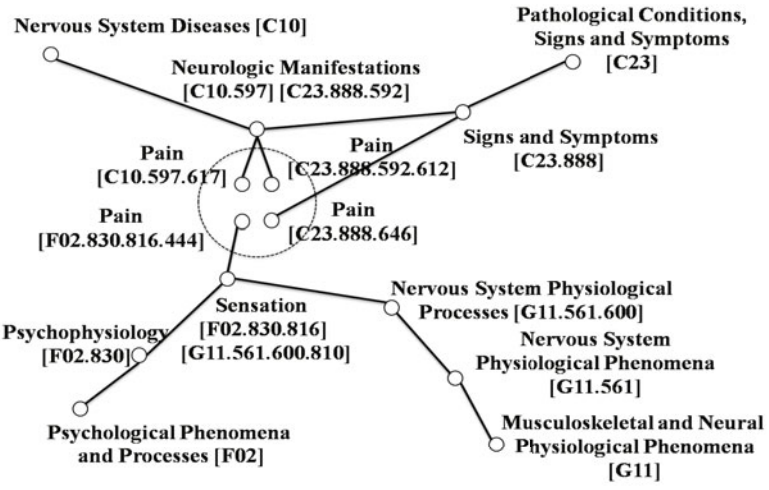


Fig. 1. Concept Pain in MeSH

(JDI) methodology to perform the disambiguation of terms in a large collection of MEDLINE citations when trying to map free text to the UMLS concepts. JDI-based WSD consists in selecting the best meaning that is correlated to UMLS semantic types assigned to ambiguous concepts in the Metathesaurus. More precisely, it is a statistical, corpus-based (training set of MEDLINE citations) method using pre-existing medical knowledge sources (a small set of journal descriptors in MeSH).

Most research works for WSD of biomedical text are based on the supervised ML approach [5, 12, 8, 9, 13, 14] and inspired by the Senseval[2] and BioCreative[3] challenges. The work in [13] used UMLS to disambiguate biomedical terms in documents using the information about term co-occurrence defined by using the naive Bayesian learning. Abbreviations in MEDLINE abstracts are resolved using SVM when trying to build a dictionary of abbreviations occurring with their long forms [12]. Briefly, works basically exploit features of general text then apply them to the biomedical text such as head word, part-of-speech, semantic relations and semantic types of words [8], unigrams, bigrams [9], surrounding words, distance, word collocations [5], lexical, syntactic features [14], etc. Most recently, the work in [15] proposed to disambiguate biomedical terms using the combination of linguistic features, which have been commonly used for WSD in general text, to the biomedical domain by augmenting it with additional domain-specific and domain-independent knowledge sources. The WSD method described in [15] combines the Concept Unique Identifiers (CUIs), which are automatically obtained from MetaMap [16] and MeSH terms, which are manually assigned to the abstracts to build feature vectors for training three WSD

[2] http://www.senseval.org/
[3] http://biocreative.sourceforge.net/

hal-00553710, version 1 - 10 Jan 2011

classifiers based on: Vector Space Model, Naive Bayes Network and Support Vector Machine. However, such methods become inflexible when always requiring a lot of efforts in terms of cost and time for human annotators.

Our contribution is outlined through the following key points:

1. We propose WSD methods that map free text to the MeSH concepts by assigning the most appropriate sense, indicated by its tree number, to each term or phrase in the local context of documents. Compared to other related WSD methods in the biomedical literature, our method has the following features: (1) do not require any training corpus, but only based on the local context of documents, and (2) exploit the MeSH semantic hierarchies to identify the correct sense for ambiguous concepts.
2. We then exploit our WSD algorithms as the basis of a sense-based indexing and retrieval model for biomedical text.

hal-00553710, version 1 - 10 Jan 2011

3 WSD Using MeSH Hierarchical Structure

Our objective here is to assign the appropriate *sense* related to a given *term* in the local context of the document mapped to the MeSH poly-hierarchy. Documents are at first tagged with Part-Of-Speech labels using a lexical tool such as TreeTagger [17]. A list of *concepts* is extracted in each document using the left to right maximum string matching algorithm. In what follows, we give some definitions and key notations and then detail our WSD methods.

3.1 Definitions and Notations

In MeSH, the preferred term, used for indexing, represents the name of the concept, also known as main heading. Otherwise, non-preferred ones, which are synonymous terms, are used for retrieval. In that poly-hierarchical structure, each concept is represented by a node belonging eventually to one (non-ambiguous) or multiple hierarchies (ambiguous), each of which corresponds to one of the sixteen MeSH domains: *A-Anatomy, B-Organisms, C-Diseases*, ... The following definitions are given based on the MeSH vocabulary.

- **Definition 1:** A **word** is an alphanumeric string delimited by spaces.
- **Definition 2:** A **term** is a group of one or more words comprising the basic unit of the vocabulary.
- **Definition 3:** A **concept** is the bearer of linguistic meaning consisting of synonymous term elements.
- **Definition 4:** The **sense** of a concept is represented by a tree node, indicated by the tree number in the poly-hierarchy. The set of **senses** of a concept c is denoted as $syn(c)$.
- **Definition 5:** The relationship **is-a** links concepts in the same hierarchy from various levels of specificity.

3.2 Left-To-Right Disambiguation

The first algorithm concerns the selection of the correct sense for each concept based on the following assumptions:

- The one-sense-per-discourse assumption [18], i.e., if a polysemous term appears two or more times in a discourse (sentence, paragraph, document), it is extremely likely that they will all share the same sense.
- The correlation between concepts in a local context expresses their semantic closeness.
- The priority of meanings is defined by the precedence of concepts: the leftmost concept impacts the overall meaning of the discourse, which inspires the semantic chain of the document from the beginning to the end.

Based on these hypotheses, we firstly compute the semantic similarity between the leftmost concept with its nearest neighbor. The third concept is disambiguated based on the meaning of the second and so on. Afterwards, their meanings will be propagated in all of their occurrences in the document. We visually illustrate such principle of WSD in Figure 2. Formally, given a sequence of n

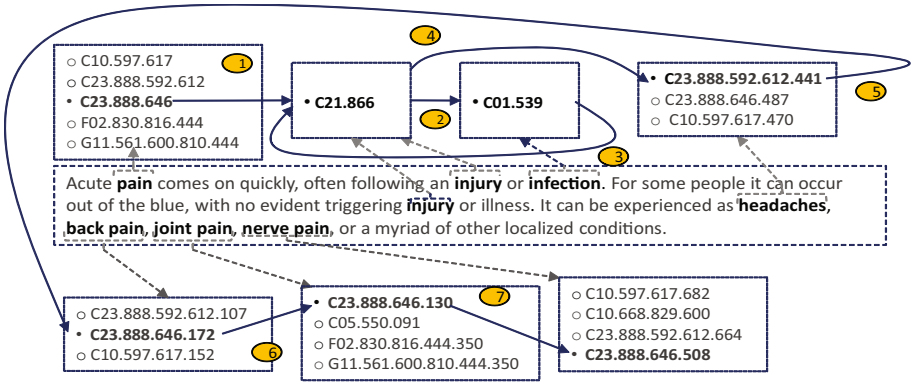


Fig. 2. Concept disambiguation

MeSH concepts, $L_n = \{c_1, c_2, ..., c_n\}$, we propose the following formula to identify the best sense for each concept c_k:

$$\begin{cases} (s_1, s_2) = \sum_{s_1 \in syn(c_1), s_2 \in syn(c_2)} sim(s_1, s_2) & \text{if } k \leq 2 \\ s_k = \arg\max(\sum_{s \in syn(c_k)} sim(s_{k-1}, s)) & \text{if } k > 2 \end{cases} \quad (1)$$

where s_k: the sense of the concept c_k,
$syn(c_k)$: the set of senses of c_k,
$sim(s_1, s_2)$: the similarity between s_1 and s_2.

hal-00553710, version 1 - 10 Jan 2011

The similarity between two senses referred to two concepts in a hierarchy is calculated using the graph-based similarity of related concepts [19]:

$$sim(s_1, s_2) = max \left\{ -\log \frac{length(s_1, s_2)}{2 * D} \right\} \quad (2)$$

where $length(s_1, s_2)$ is the shortest path between s_1 and s_2, and D is the depth of the hierarchy.

3.3 Cluster-Based Disambiguation

Concept clusters have been mentioned in [20], but they are not employed for WSD. Inspired by this work, we exploit the concept clusters for biomedical WSD. In this approach, the one-sense-per-discourse [18] is applied and furthermore we assume that the sense of a concept depends on the sense of other hierarchical related concepts, whatever their location is within a document. Based on these assumptions, our algorithm functions as follows: related concepts in the document are grouped into clusters, each of which belongs to the same hierarchy. A concept can appear in one or more clusters since it can have multiple senses as defined in the thesaurus. Senses that maximize the similarity between concepts within a cluster will be assigned to the related concepts.

In our case, clusters are named according to the sixteen MeSH domains (A, B, C, ...), formally $K = \{k_1, k_2, ..., k_{16}\}$. Given a list of concepts in a document, $L_n = \{c_1, c_2, ..., c_n\}$, we assign to each concept c_i its correct sense based on the cluster-based similarity, defined as following:

$$s_i = \arg\max(\sum_{c_i, c_j \in k_u, i \neq j} \; \sum_{s_a \in syn(c_i), s_b \in syn(c_j)} sim(s_a, s_b)) \quad (3)$$

The cluster-based WSD is similar to the Left-to-Right WSD while using the same similarity in formula 2. The difference between two WSD algorithms is that the considered context changes, i.e., the cluster of concepts for the former and the left-most concept and the previously disambiguated concept for the later.

4 Sense-Based Indexing and Retrieval

At this level, our objective is to compute the relevance score of documents with respect to each query. The principle of our proposed approach aims at representing both documents and queries using semantic descriptors and then matching them using a sense-based weighting scheme. Hence, each concept in the document (query) is tagged with the appropriate sense in the local context and indexed with its unique sense in the document. The retrieval makes use of the detected sense related to a given term in the query compared to documents where it appears. We formulate our sense-based indexing and retrieval process through the following steps:

hal-00553710, version 1 - 10 Jan 2011

Step 1: Build the document index. Let D_i be the initial document, then D_i contains both disambiguated concepts in the thesaurus and single words in the vocabulary. Formally:

$$\begin{array}{l} D_i^s = \{d_{1i}^s, d_{2i}^s, ..., d_{mi}^s\} \\ D_i^w = \{d_{1i}^w, d_{2i}^w, ..., d_{ni}^w\} \end{array} \tag{4}$$

where D_i^s, D_i^w are respectively the set of concepts and single words, m and n are respectively the number of concepts and words in D_i, d_{ji}^s is the *j-th* concept and d_{ji}^w is the *j-th* word in the document D_i.

Step 2: Build the query index. Queries are processed in the same way as documents. Thus, the original query Q can be formally represented as:

$$\begin{array}{l} Q^s = \{q_1^s, q_2^s, ..., q_u^s\} \\ Q^w = \{q_1^w, q_2^w, ..., q_v^w\} \end{array} \tag{5}$$

where Q^s, Q^w are respectively the set of concepts and single words, u and v are respectively the number of concepts and words in Q, q_k^s is the *k-th* concept and q_k^w is the *k-th* word in the query Q.

Step 3: Compute the document relevance score. The relevance score of the document D_i with respect to the query Q is given by:

$$RSV(Q, D_i) = RSV(Q^w, D_i^w) + RSV(Q^s, D_i^s) \tag{6}$$

where $RSV(Q^w, D_i^w)$ is the *TF-IDF* word-based relevance score and $RSV(Q^s, D_i^s)$ is the sense-based relevance score of the document w.r.t the query, computed as follows:

$$\begin{array}{l} RSV(Q^w, D_i^w) = \sum_{q_k^w \in Q^w} TF_i(q_k^w) * IDF(q_k^w) \\ RSV(Q^s, D_i^s) = \sum_{q_k^s \in Q^s} \alpha_k * (1 + h(q_k^s)) * TF_i(q_k^s) * IDF(q_k^s) \end{array} \tag{7}$$

where TF_i: the normalized term frequency of the word q_k^w or concept q_k^s in document D_i, IDF: the normalized inverse document frequency of q_k^w or q_k^s in the collection, α_k: the meaning rate of q_k^s between D_i and Q^s, $h(q_k^s)$: the specificity of q_k^s associated with its meaning in the query, calculated as follows:

$$h(q_k^s) = \frac{level(q_k^s)}{MaxDepth} \tag{8}$$

where $level(q_k^s)$: depth level of q_k^s, $MaxDepth$: maximum level of the hierarchy.

$$\alpha_k = \left\{ \begin{array}{ll} 1 & \text{if } sense(q_k^s, Q^s) = sense(q_k^s, D_i) \\ 1 - \beta & \text{otherwise} \end{array} \right\} \tag{9}$$

where $sense(q_k^s, Q^s)$ (resp. $sense(q_k^s, D_i)$) indicates the sense of q_k^s in the query (resp. the document D_i) (see definition 4); β is an experimental parameter obtained the value in the interval $[0, 1]$. Indeed, we assume that information at a

hal-00553710, version 1 - 10 Jan 2011

more fine-grained level of specificity is more relevant for search users. The specificity factor in formula 8 is integrated to favour documents containing concepts at a more fine-grained level of specificity. The meaning rate α_k is considered to alleviate the relevance score of any document D_i in which the sense of the concept q_j^s is different from the query.

5 Experimental Evaluation

In our experimental evaluation, we studied the effects of assigning the sense of concepts in documents during the process of biomedical IR. Hence, we performed a series of experiments to show the impact of sense tagging on the retrieval performance. We describe in what follows the experimental settings, then present and discuss the results.

5.1 Experimental Setup

– *Test collection*: We use the OHSUMED test collection, which consists of titles and/or abstracts from 270 medical journals published from 1987-1991 through the MEDLINE database [21]. A MEDLINE document contains six fields: *title (.T), abstract (.W), MeSH indexing terms (.M), author (.A), source (.S), and publication type (.P)*. To facilitate the evaluation, we converted the original OHSUMED collection into the TREC standard format. Some statistical characteristics of the collection are depicted in Table 1.
We have selected 48 TREC standard topics, each one is provided with a set of relevant documents judged by a group of physicians in a clinical setting. The *title* field indicates *patient description* and the *description* field announces *information request*.

Table 1. Test collection statistics

Number of documents	293.856
Average document length	100
Number of queries	48
Average query length	6 (TITLE) 12 (TITLE+DESC)
Average number of concepts/query	1.50 (TITLE) 3.33 (TITLE+DESC)
Average number of relevant docs/query	50

– *Evaluation measures*: P@5, P@10 represent respectively the mean precision values at the top 5, 10 returned documents and MAP (Mean Average Precision) over the total of 48 testing queries. For each query, the first 1000 documents are returned by the search engine and average precisions (P@5, P@10, MAP) are computed for measuring the IR performance.

hal-00553710, version 1 - 10 Jan 2011

– *Medical Subject Headings:* MeSH is a medical domain knowledge resource that has been developed at the US National Library of Medicine (NLM) since 1960. The latest version of MeSH released in 2010 consists of 25,588 entries, each one represents a preferred concept for the indexing of publications included in the MEDLINE database.

5.2 Experimental Results

For evaluating the effectiveness of our WSD methods and the performance of our sense-based indexing approach, we carried out two sets of experiments: the first one is based on the classical index of titles and/or abstracts using Terrier standard configuration based on the state of the art weighting scheme OKAPI BM25 [22], used as the baseline, denoted *BM25*. The second set of experiments concerns our semantic indexing method and consists of three scenarios:

1. The first one is based on the naive selection of the first sense found in the hierarchy for each concept, denoted *WSD-0*,
2. The second one is based on the *Left-To-Right WSD*, denoted *WSD-1*,
3. The third one is based on the *Cluster-based WSD*, denoted *WSD-2*.

We use both terms representing MeSH entries and single words that do not match this thesaurus. In the classical approach, the documents were first indexed using the Terrier IR platform (`http://ir.dcs.gla.ac.uk/terrier/`). It consists in processing single terms occurring in the documents through a pipeline: removing stop words, identifying concepts in documents and stemming[4] of English words.

In our sense-based approaches that employ semantic information from WSD, documents and queries are firstly disambiguated and indexed with appropriate senses of concepts defined in the MeSH vocabulary. The semantic weighting scheme is then applied for each term in the query using Formula 7. In our experiments, the β parameter in formula 9 is set to 0.15 for the best results.

Table 2 depicts the IR performance based on the baseline and the proposed WSD methods both for querying with only *title* and with *title and description* together. Figure 3 shows the improvement rates of our methods over the baseline. We obtained the following results: both *WSD-1* and *WSD-2* methods outperform the *baseline* both for short queries (title) and long queries (title and description). The improving rates obtained are 5.61% and 4.77% for *WSD-1* and *WSD-2* respectively. This proves the interest of taking into account the document and query's semantics along with the specificity of the documents as well as of the queries in the IR process. Furthermore, the results show that randomly selecting the concept's sense (as in *WSD-0*) does not help, but correctly assigning the appropriate sense for each concept improves the IR performance. We see that *WSD-1* and *WSD-2* always give a better precision than *WSD-0*.

We have also tested the query evaluation with only *title* and *title and description* together in order to show the impact of the query length on the IR performance. We see that the semantic chain inspired the *WSD-1* gets a higher

[4] `http://snowball.tartarus.org/`

hal-00553710, version 1 - 10 Jan 2011

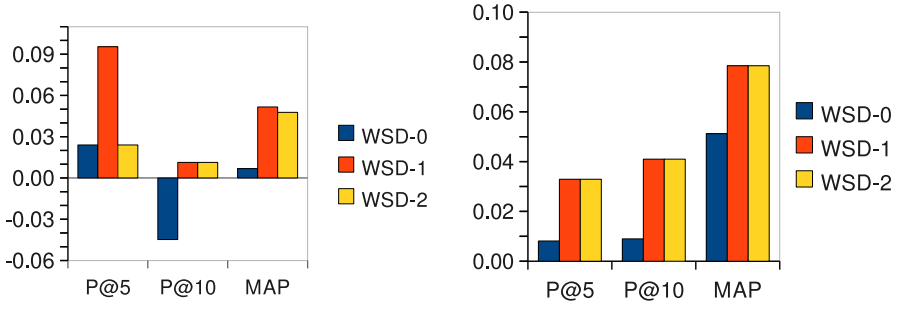
Table 2. Official results on the OHSUMED collection

(a) title

Measure	BM25	WSD-0	WSD-1	WSD-2
P@5	0.17500	0.1792	0.19170	0.17920
P@10	0.18540	0.1771	0.18750	0.18750
MAP	0.10270	0.1034	0.10800	0.10760

(b) title + description

Measure	BM25	WSD-0	WSD-1	WSD-2
P@5	0.50420	0.50830	0.52080	0.52080
P@10	0.45630	0.46040	0.47500	0.47500
MAP	0.24210	0.25450	0.26110	0.26110



(a) title (b) title + description

Fig. 3. Improving rate over the baseline

improving rate over the baseline (5.61%) compared to *WSD-2* (4.77%) for short queries. In addition, both of them get the same improving rate over the baseline (7.48%) for long queries. Indeed, for longer queries, both of our two methods identify better the sense of each concept and then induce its appropriate specificity level in the document. The only difference between the two methods is the selection of the context where a concept appears: the former from the left-side concepts and the later from clusters of concepts in the same hierarchy.

In a finer-grained analysis, we have reviewed the IR performance for each long query in the testing set to verify the impact of the concept's specificity with respect to the query length and the number of concepts in the query on the IR performance. For each query, we computed the average specificity of the query and we obtained a range from 2 to 6. For each group of queries having the specificity from 2 to 6, we compute the average query length, the average number of concepts and the average improving rates. Figure 4 shows the analysis of the results according to the query specificity. We notice that the more the query specificity, i.e., the average concept's specificity level in the query is, the less the number of concepts used in the query is. This could be explained by the fact that a few of more specific concepts are enough to express the user needs while many of generic ones are required to express better the user information need. In most of cases, our approach favours documents containing concepts at a higher level of specificity and shows a consistent improvement over the baseline. However, if the query is long but the number of concepts having a high specificity level in the query is less, our system tends to return first documents containing those

hal-00553710, version 1 - 10 Jan 2011

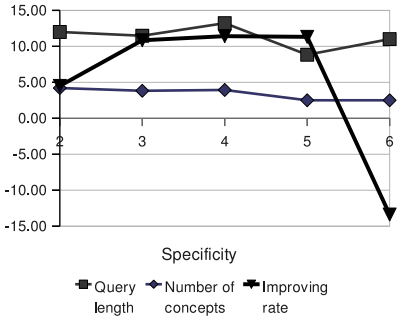


Fig. 4. Analysis of results according to query specificity

concepts. This could be the reason of the decrease of the performance when a few most specific concepts impact other terms in the query.

6 Conclusion

In this work, we have proposed and evaluated a sense-based approach of indexing and retrieving biomedical documents. Our approach relies on two WSD methods for identifying ambiguous MeSH concepts: *Left-To-Right WSD* and *Cluster-based WSD*. The evaluation of the indexing method on the standard OHSUMED corpus proves the evidence of integrating the sense of concepts in the IR process. Indeed, most of ongoing IR approaches match documents using term distribution without the sense of query terms along with the ones of the documents. The more the meaning of the query term matches the meaning of the document term, the more the retrieval performance is improved. The more the query terms are specific, the more the specificity of the returned documents is fine-grained. Future works will focus on automatically expanding the query using concepts extracted from the hierarchy indicated by the correct sense of concepts occurred in documents.

References

1. Lesk, M.: Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In: SIGDOC 1986, pp. 24–26 (1986)
2. Gale, W., Church, K., Yarowsky, D.: A method for disambiguating word senses in a large corpus. Computers and the Humanities, 415–439 (1993)
3. Mihalcea, R.: Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In: HLT 2005, pp. 411–418 (2005)
4. Lee, Y.K., Ng, H.T., Chia, T.K.: Supervised word sense disambiguation with support vector machines and multiple knowledge sources. In: Senseval-3: Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp. 137–140 (2004)

hal-00553710, version 1 - 10 Jan 2011

5. Liu, H., Teller, V., Friedman, C.: A multi-aspect comparison study of supervised word sense disambiguation. J Am. Med. Inform. Assoc. 11(4), 320–331 (2004)
6. Yarowsky, D.: Unsupervised word sense disambiguation rivaling supervised methods. In: ACL 1995, pp. 189–196 (1995)
7. Abney, S.P.: Bootstrapping. In: ACL, pp. 360–367 (2002)
8. Leroy, G., et al.: Effects of information and machine learning algorithms on word sense disambiguation with small datasets. Medical Informatics, 573–585 (2005)
9. Joshi, M., Pedersen, T., Maclin, R.: A comparative study of support vector machines applied to the word sense disambiguation problem for the medical domain. In: IICAI 2005, pp. 3449–3468 (2005)
10. Weeber, M., Mork, J., Aronson, A.: Developing a test collection for biomedical word sense disambiguation. In: Proc. AMIA Symp., pp. 746–750 (2001)
11. Humphrey, S.M., Rogers, W.J., et al.: Word sense disambiguation by selecting the best semantic type based on journal descriptor indexing: Preliminary experiment. J. Am. Soc. Inf. Sci. Technol. 57(1), 96–113 (2006)
12. Gaudan, S., Kirsch, H., Rebholz-Schuhmann, D.: Resolving abbreviations to their senses in medline. Bioinformatics 21(18), 3658–3664 (2005)
13. Andreopoulos, B., Alexopoulou, D., Schroeder, M.: Word sense disambiguation in biomedical ontologies with term co-occurrence analysis and document clustering. IJDMB 2(3), 193–215 (2008)
14. Mohammad, S., Pedersen, T.: Combining lexical and syntactic features for supervised word sense disambiguation. In: CoNLL 2004, pp. 25–32 (2004)
15. Stevenson, M., Guo, Y., Gaizauskas, R., Martinez, D.: Knowledge sources for word sense disambiguation of biomedical text. In: BioNLP 2008, pp. 80–87 (2008)
16. Aronson, A.R.: Effective mapping of biomedical text to the UMLS Metathesaurus: the metamap program. In: Proceedings AMIA Symposium, pp. 17–21 (2001)
17. Schmid, H.: Part-of-speech tagging with neural networks. In: Proceedings of the 15th conference on Computational linguistics, pp. 172–176 (1994)
18. Gale, W.A., Church, K.W., Yarowsky, D.: One sense per discourse. In: HLT 1991: Proceedings of the workshop on Speech and natural Language, pp. 233–237 (1992)
19. Leacock, C., Chodorow, M.: Combining local context and wordnet similarity for word sense identification. An Electronic Lexical Database, 265–283 (1998)
20. Kang, B.Y., Kim, D.W., Lee, S.J.: Exploiting concept clusters for content-based information retrieval. Information Sciences - Informatics and Computer Science 170(2-4), 443–462 (2005)
21. Hersh, W., Buckley, C., Leone, T.J., Hickam, D.: Ohsumed: an interactive retrieval evaluation and new large test collection for research. In: SIGIR 1994, pp. 192–201 (1994)
22. Robertson, S.E., Walker, S., Hancock-Beaulieu, M.: Okapi at trec-7: Automatic ad hoc, filtering, vlc and interactive. In: TREC, pp. 199–210 (1998)

hal-00553710, version 1 - 10 Jan 2011