Is it Useful to Support Users with Lexical Resources? A User Study.

Ernesto William De Luca

University of Applied Science Potsdam
Friedrich-Ebert-Str. 4, 14467 Potsdam, Germany
Phone: +49 331 580 1520
Fax: +49 331 580 1599
deluca@fh-potsdam.de

Abstract

Current search engines are used for retrieving relevant documents from the huge amount of data available and have become an essential tool for the majority of Web users. Standard search engines do not consider semantic information that can help in recognizing the relevance of a document with respect to the meaning of a query. In this paper, we present our system architecture and a first user study, where we show that the use of semantics can help users in finding relevant information, filtering it ad facilitating quicker access to data.

Keywords: Information Retrieval, Lexical Resources, User Studies, Semantics, Clustering, Wordnet, Word Sense Disambiguation.

1. Introduction

Search engines are an essential tool for the majority of Web users for finding information in the huge amount of documents available online. For most ad-hoc search tasks they perform satisfyingly, however certain fundamental properties still leave room for improvement. For example, if users perform general questions, they frequently get lost in the huge amount of documents returned. More recent approaches, such as Yippi[1] (formerly Vivísimo (Koshman et al., 2006)) organize search results into categories (hierarchical clusters), based on textual similarity. However, these methods only consider the word distribution in documents, without taking into account linguistic criteria derived from the underlying query, such as different meanings of a term. Therefore, the assigned categories usually do not represent the categories a user is expecting for the issued query. Linguistic information (e.g. semantics) can provide valuable support for the user's search process. Retrieved documents could for instance be grouped by the meanings of the query, allowing the user to choose one of these meanings and navigate only the documents related to it.

Wordnet-based Categorization. Currently there are no efficient systems that support the user in searching information with semantic knowledge. In this paper, we want to evaluate our approach that aims to help users in searching and browsing information using lexical resources, based on the idea of supporting them with meanings in an interactive search process.

Lexical resources contain the different meanings of the words and the related linguistic relations providing semantic information. In order to access information by their meanings, different problems have to be solved: disambiguating words contained in the query, and categorizing and presenting the results appropriately. Ideally, with a semantic-based approach the system should understand the users without having them reformulate their query.

Different work has been already done on semantic WordNet-based categorization methods ((Gonzalo et al., 1998; Tokunaga et al., 1998; Varelas et al., 2005)). These categorization techniques have been criticized from the performance point of view, where linguistic information seems not to be useful and too poor for text categorization (Moschitti and Basili, 2004; Bloehdorn et al., 2006). Hovewer, different other approaches have shown that it can be beneficial if linguistic information is combined with clustering techniques (De Luca and Nürnberger, 2006b; Elberrichi et al., 2008; Li et al., 2009; Buscaldi and Rosso, 2010).

2. Sense Folder System (SFS)

In the following we describe our Sense Folder System (SFS) architecture and how we build the semantic concepts provided for supporting the user during the interaction process. The SFS combines linguistic information with knowledge-based word sense disambiguation and clustering methods (De Luca, 2008) like described also in (Gliozzo et al., 2004; Mihalcea, 2006).

2.1. SFS Architecture

The search process starts after the user submits her query by means of a user interface (see Figure 1). The query is simultaneously sent to the search engine and to the *Sense Folder Engine* (SFE). While documents are retrieved and indexed, the SFE retrieves the different meanings $[q_1, ..., q_n]$ (SynSets) of the query from a lexical resource R (in our case Wordnet), thus forming different *Sense Folders*. For every meaning q_i a Sense Folder $\vec{S_i}$ is created as described in Section 2.2. Then, every document (retrieved from the search engine) is first assigned to its nearest prototype vector derived from the lexical resource and afterwards this classification is revised by the clustering process. This additional clustering step has been introduced in order to enhance the semantic-based classification (only based on lexical resources) by considering also similarities in-between documents. This approach has shown to strongly improve the classification (or disambiguation) performance (Gonzalo et al., 1998; De Luca and Nürnberger,

[1] http://search.yippy.com/

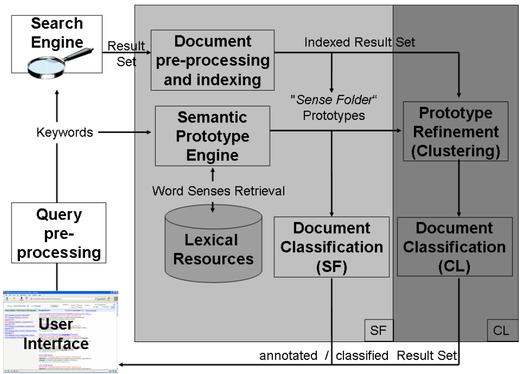


Figure 1: Sense Folder System Architecture

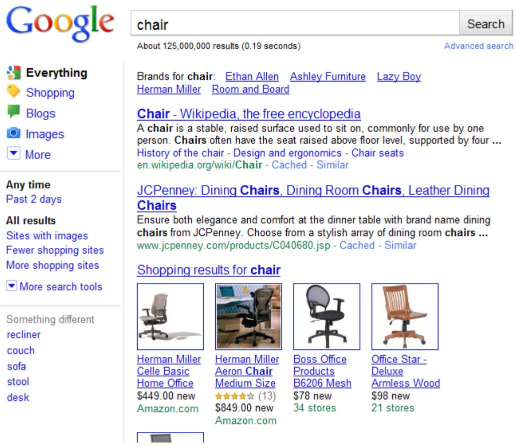
2006b; Elberrichi et al., 2008; Li et al., 2009). The semantic information assigned is appended to the document as additional information in order to help the user in finding the relevant documents, without the need of browsing all documents.

2.2. Sense Folder Definition

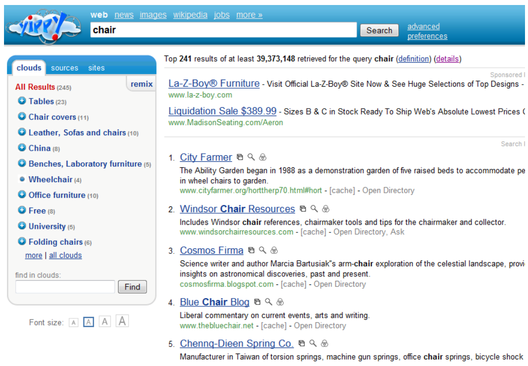
Users interacting with the SFS are asked to choose a "main query word" Q describing the concept they are searching for (in our case e.g. *chair*) and have the possibility to add other "optional query words" O that better specify the meant concept for disambiguation purposes (e.g. *furniture, wood, etc.*). Given the main query word $Q = \{q_1, ..., q_n\}$, we create different Sense Folders $S = \{\vec{S_1}, ..., \vec{S_n}\}$ that define the different meanings $[q_1, ..., q_n]$ of the query Q. We also define the linguistic context $C = \{c_1, ..., c_n\}$, that contains the different linguistic relations $[c_1, ..., c_n]$ of every meaning $[q_1, ..., q_n]$ retrieved from a lexical resource R. Different "optional query words" $O = \{o_1, ..., o_n\}$ can be given by the user and are included in all Sense Folders, because they could help in better describing the user context. Thus, the system builds, on the basis of the "main word", the Sense Folders (representing all meanings of this "main word"). Every Sense Folder $\vec{S_i}$ is a container that includes only one meaning of the main word q_i, all its related linguistic information (linguistic context) c_i retrieved from a lexical resource R (in our case Wordnet) and the user typed "optional query words" o_i.

3. User study

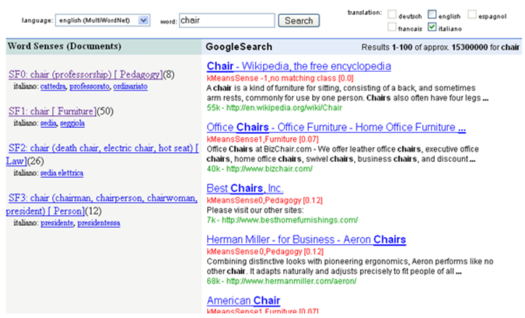
User studies (Dumais et al., 2001), (Labrou and Finin, 1999) have shown that categorized information can improve retrieval performance. Motivated by these evaluations, the SFS has been evaluated within an interactive user interface to verify the assumption that the semantic filtering support is beneficial for the users. We provided a user interface, where additional disambiguating information was added to the documents of a search result. This additional information is used in order to enable categorization, restructuring or filtering of the retrieved document set (see Figure 2(c)).



(a) Results for "chair" from Google. The results are a ranked list without considering the sense of the word.



(b) Results for "chair" from Yippi. Yippi clusters search results based in key terms, it does however not seem to base the clusters on the semantic meaning of the query.



(c) Results for "chair" in the Sense Folder System. There are 4 Sense Folders, one for each sense.

Word	Synonyms	Domain
chair	professorship	Pedagogy
chair		Furniture
chair	electric chair, hot seat	Law
chair	president, chairperson	Person

(d) WordNet noun collocation of the term "chair". The same presented by the Sense Folder System in Figure 2(c)

Figure 2: Examples of retrieved results for the query "chair" from three search engines.

Questionnaire
1. How many concepts can you find on the first page? (0-5 — 5-10 — More)
2. Would you agree with the concepts found by the SF engine? (Yes — No, why?)
3. How do you find the SF search engine concept? (Easy — Difficult — Abstract, why?)
4. Which are the differences you noticed to common search engines?
5. Did you notice the "Sense Folder" column on the left side? (Yes — No)
5.1 If yes, was it helpful to you? (Yes — No, why?)
5.2 Could you shortly describe its function?
6. Would you like to see similar features in search engines in the future? (Yes — No, why?)
7. What would you suggest to change/improve for the SF search engine? (Yes — No, why?)
8. Would you like to use this search engine instead of common search engines? (Yes — No, why?)
9. For which purpose would you use this search engine? (All searches — Some searches, give an example)

Table 1: Questionnaire used for the evaluation.

3.1. Evaluation Results.

In order to evaluate the performance of the SFS, a user study was conducted. A total of 9 questions covering different aspects of the Sense Folder search results presentation were asked. Sixteen participants were presented with results of the same queries from Google[2], Yippi[3] and a local deployment of the Sense Folder System (see Figure 2). Yippi was chosen because it groups similar results together into "clouds"[4]. Table 1 summarizes the questionnaire used in the user study including questions and answers given to the users. All questions had at least one negative and one positive answer, most allowed the participants to leave short comments and motivations to their answers.

1. Users were asked to recognize the different concepts of one given word related to the search results retrieved by the search systems. Most of them (94%) recognized between one and five concepts. 6% said that there are more than five concepts.

2. Users were asked whether or not they agreed with the concepts found by the Sense Folder System. 94% said that all expected concepts were presented. 6% said that other concepts were missing.

3. The participants had to estimate the difficulty of the interaction with the Sense Folder System. 75% found it easy, while 19% difficult and the remaining 6% abstract.

4. Users were asked to describe the differences between Google, Yippi and the Sense Folder System. 81% were positive to the added value of the Sense Folder Annotation. The system was intuitive and supported them in disambiguating the concepts related to the query. 50% of the users saw the list of semantic concepts as positive. 25% said that the Sense Folder System clustered documents similarly to Yippi. They also noticed that Google and Yippi only covered one dominant concept. 6% of the participants observed that the Sense Folder System did not allow search for different media types. 19% did not notice any difference.

5. We asked if the Sense Folder System had been helpful and why. 81% explained that the use of filtering by semantic concept for the query was very positive and they could access information quickly and categorized by concepts. They also saw an easy way of filtering results. 19% claimed the coverage of topics was incomplete and said that they prefered to use longer queries instead.

6. Participants were asked whether they would like to use features similar to the Sense Folder System in future search engines. 81% where positive. They also said that the feature reduces non-relevant information and gives quick access to good results. 19% said they preferred e.g. the Wikipedia disambiguation page due its clustering. However, Wikipedia only presents concepts, without clustering documents related to the found concepts or to a given query.

7. We asked if the participants had suggestions for improving the Sense Folder System. 31% liked it as-is. 25 % would like to choose concepts not to consider. This functionality is implicity available as documents are already filtered by concepts, by clicking on a concept, others are automatically excluded. 19% would however prefer a nicer graphical user interface. 13% wanted to search not only for text, but also for other media and 12% would like to have more explanations about the concepts retrieved.

8. Users were asked if they would use the Sense Folder System instead of common search engine. 81% would. 31% because of the better support for finding relevant of documents, as well as the filtering of search results (50%). 19% who would not, said that old habits die hard and questioned the usability.

9. We wanted to know from the 81% of the users (that would use the Sense Folder System instead of common search engines) for which kind of search they would use the Sense Folder System. 62% of them would use it for all searches, while 38% for specific searches.

4. Coarse-grained Evaluation

In addition to the user study, we decided to evaluate the performance of the Sense Folder approach considering two

[2]http://www.google.com
[3]http://search.yippy.com/
[4]http://search.yippy.com/about-yippy-search

word sense granularity levels. In some cases word senses derived in WordNet belong to the same meaning and could be merged in one sense (De Luca and Nürnberger, 2006a). This is what we call a *coarse-grained representation* of the sense inventory. It is useful for making results more meaningful for evaluation. On the other hand, the sense distinction might not be detailed enough, so that a *fine-grained representation* is needed in order to distinguish the word senses on a more detailed level. In the following, we wanted to discover, if a more accurate selection of linguistic parameters, in combination with document categorization techniques enhances the automatic classification performed by the Sense Folder approach.

The combination of different linguistic parameters with different clustering and stemming methods in a fine-grained setting has been already evaluated, in order to recognize the best performing setting for retrieval and categorization (De Luca, 2010).

The coarse-grained evaluation treats a more general distinction of word senses as semantically related word senses are merged. To achieve a coarse-grained word-senses representation an additional merging step is included. If the similarity between two reference vectors exceeds a given threshold, they are merged together. The merging methods are applied on the reference vectors created from the Sense Folders before the document classification starts. Four merging methods have been implemented for grouping word senses by context (considering all linguistic information available), domain, hyperonyms and hyponyms (considering respectively the overlapping domain information, hyperonyms and hyponyms available).

4.1. Parameter Settings

Different combinations of linguistic relations (synonyms, coordinate terms, hyperonyms, hyponyms, glosses and semantic domains, and the semantic domain hierarchy) were taken into account, in order to create the Sense Folders for classifying the documents semantically and evaluating them in a coarse-grained framework. The tf and $tf \times idf$ encoding, as well as pre-processing methods like *stemming* (for reducing - possibly only the potentially important terms - a document to a representative minimum, bringing all words to their base form (Croft, 1995)) are considered in order to find an optimal representation. The resulting parameter settings are: tf-based (`Tf`) and respective stemmed one (`Tf + Stem`), $tf \times idf$-based (`TfxIdf`) and respective stemmed one (`TfxIdf + Stem`).

Three Sense Folder clustering methods (k-Means, Density-Based and Expectation-Maximization algorithms) that have been implemented in previous work are used (De Luca, 2008).

The k-Means clustering algorithm (`KM`) does not require any parameters, because the number k of cluster centers, corresponds to the number of word senses available in WordNet used as initial prototypes for the clustering process. The Density-Based Clustering method (`DB`) uses similar to k-Means Sense Folders as initial prototypes. The parameter λ has been set to 0.9 and the n parameter that represents the number of neighbors to be considered, is set to 2. The Expectation-Maximization(`EM`)-λ algorithm adopt the Sense Folders in the role of "labeled data," whereas the vectors representing the documents supply the "unlabeled data". The weight parameter λ for the unclassified data points is set to 0.9. The parameter settings for the last two clustering algorithms (EM-λ and DB) have been set according to the evaluation results presented in (Honza, 2005).

4.2. Evaluation Results

The evaluation of the merging methods described above is combined with the evaluation of different thresholds that can be subdivided into four different steps:

- Combining the Sense Folder Approach with Single Merging Methods (SMM) and Tresholds (SMT)
  - Finding the Best Performing SMM
  - Finding the Best Performing SMT
- Combining SMM and SMT into Multi-Stage Merging Methods

In the following, we summarize the most important coarse-grained evaluation results.

Best Performing Single Merging Methods (SMM) While for the context and domain merging methods we use on the one side all linguistic relations available and on the other side the domain information, for merging by hyperonyms and hyponyms we added the respective linguistic parameter hyperonyms and hyponyms that are used for the creation of the semantic prototypes and influence the merging and classification process.

Basing on the best performance results reached in the fine-grained evaluation (see (De Luca, 2010) and Table 2) , if the $tf \times idf$ document vectors of the retrieved documents are used alone (`TfxIdf+Merge+NoStem`), while tf document vectors of the same documents perform best in combination with stemming methods (`Tf+Merge+Stem`). We decide to compare these two document representations, for analyzing which of them works better with coarse-grained concept classes.

Best Performing Single Merging Thresholds(SMT) In addition, we explore all thresholds in order to discover which are the most suitable in merging word senses. Because good results were already achieved by setting the thresholds to 0.5 (De Luca and Nürnberger, 2006a), we decide to hold this threshold for all merging methods and try a higher one (0.7), in order to see if the merging performance could be more enhanced (for merging by hyperonyms and hyponyms) and a lower one for all (0.1). Because the merging methods applying hyponyms did not merge any word senses in the previous evaluation (De Luca and Nürnberger, 2006a), we also use two lower thresholds to explore if these helped in merging similar word senses by hyponyms (0.1 and 0.3). We only use the 1.0 threshold for the merging by domains, where two word senses are only merged if they have exactly the same domain description. This decision is also based on the same experiments conducted in (De Luca and Nürnberger, 2006a).

SF [tf]	KM [tf]	EM [tf]	DB [tf]
0.37	0.43	0.28	0.33
SF [tf] [stem]	KM [tf] [stem]	EM [tf] [stem]	DB [tf] [stem]
0.41	0.45	0.28	0.43
SF [tfxidf]	KM [tfxidf]	EM [tfxidf]	DB [tfxidf]
0.42	0.41	0.30	0.41
SF [tfxidf] [stem]	KM [tfxidf] [stem]	EM [tfxidf] [stem]	DB [tfxidf] [stem]
0.31	0.39	0.32	0.32

Table 2: Overall Fine-grained Evaluation [Syn, dom, gloss]

	SF [tf] [stem] [merge]	KM [tf] [stem] [merge]	SF [tfxidf] [merge]	KM [tfxidf] [merge]
Dom 0.1/1.0	0.49	0.44	0.52	0.49
Cxt 0.7	0.45	0.46	0.45	0.43
Hyper 0.7	0.40	0.42	0.42	0.37
Hypo 0.3	0.40	0.41	0.43	0.38

Table 3: Overall Single Coarse-grained Evaluation [Syn, dom, gloss]

Single and Multi-Stage Coarse-Grained Evaluation Comparison Comparing the results presented in Table 3 and Table 4 it emerges that in both cases the single and multi-stage merging methods that consider the hyperonyms or hyponyms for merging word senses decrease the classification performance, because these terms are too general or too specific to be helpful for classification. When such linguistic information is combined with clustering methods, in some cases, the classification performance is strongly enhanced, because similar documents are recognized. Sometimes this semantic information already contained in lexical resources is sufficient to recognize the linguistic context of a document given a query, so that clustering methods are not needed or their use affects negatively the classification. An overall improvement of the "pure" Sense Folder classification performance applying the k-Means clustering algorithm has been shown. The few exceptions from that trend refer to cases where the only use of the Sense Folders was already sufficient to achieve good results.

Merging word senses using the the single "context" and "domain" merging methods (*Cxt0.7 and Dom0.1*) enhance good results, but their multi-stage coarse-grained combination results show a better performance (*Cxt0.7Dom0.1*), outperforming all baselines. We can face up to the facts that as in the fine-grained evaluation the multi-stage coarse-grained classification works better with the tf-based document representation and stemming methods than with the $tf \times idf$-based document representation. Comparing the results with the "Random" and "First Sense"baselines (see (McCarthy et al., 2004)) in more detail, we can see that all baselines are outperformed in the overall performance, except in single cases for "argomento" "lingua" or "regola", due to the known unbalanced class distribution problem that persists. Considering the "Most Frequent Sense" (MFS) baseline (Mihalcea, 2006), we can notice that the multi-stage merging methods behave similar to the respective single methods, apart from the multi-stage *Cxt0.7Dom0.1* merging method. It uses tf document vectors and outperforms the overall baseline, hence classifying better both for English and Italian. This baseline is outperformed of 0.10 for English with the k-Means algorithm and 0.03 for Italian and in the overall performance the "pure" Sense Folder classification overbids the baseline of 0.06 while k-Means of 0.02. The same multi-stage merging method combined with $tf \times idf$ document vectors outperformed the English MFS baseline, but not the Italian one. The lower quality of results achieved when applying the Italian stemmer are an indication that this stemmer does not reduce the terms to their correct base form. Its application is partially beneficial and partially not. For the English language the stemmer works better and could be used for this language, for achieving better results.

Nevertheless, the multi-stage coarse-grained classification works better with the tf-based document representation and stemming methods than with the $tf \times idf$-based document representation. This is most likely to be attributed to the idf measure that cannot be estimated very good and be meaningful adopted, because the document collection is still relative small. When the number of documents would be higher, a more reasonable weight of idf would be available.

5. Conclusions

In this work, lexical resources (WordNet) have been used for providing the possible meanings of query words. We presented a user study that proved that the use of semantics (and the related semantic classification methods) can help users in finding relevant information, filtering it and facilitating quicker access. It should be said that the user study showed that even this simple prototype received overall positive opinions from the participants. In addition, we discussed a coarse-grained evaluation of the semantic classification methods. Here, we have shown that an accurate selection of linguistic parameters, combined with clustering and merging methods, can enhance retrieval performance and the categorization of semantically similar documents. Best results have been achieved in most cases using the "pure" Sense Folder approach (SF) and the k-Means algorithm (KM) with synonyms, glosses and semantic domains as linguistic parameters. Using these settings, we outperformed the *Most Frequent Sense* heuristic (Mihalcea, 2006) by 10% for English and by 3% for Italian.

	SF [tf] [stem] [merge]	KM [tf] [stem] [merge]	SF [tfxidf] [merge]	KM [tfxidf] [merge]
Cxt 0.7 Dom 0.1	0.57	0.53	0.55	0.47
Cxt 0.7 Hyper 0.7	0.48	0.49	0.47	0.42
Hyper 0.7 Dom 1.0	0.53	0.52	0.46	0.46
Hypo 0.3 Dom 1.0	0.50	0.47	0.49	0.44

Table 4: Overall Multi-Stage Coarse-grained Evaluation [Syn, dom, gloss]

6. References

Stephan Bloehdorn, Roberto Basili, Marco Cammisa, and Alessandro Moschitti. 2006. Semantic kernels for text classification based on topological measures of feature similarity. In *Proceedings of the Sixth International Conference on Data Mining*, ICDM '06, pages 808–812, Washington, DC, USA. IEEE Computer Society.

D. Buscaldi and P. Rosso. 2010. Indexing with wordnet synonyms may improve retrieval results. In *Revised Selected Papers CLEF-2009*. Springer-Verlag, LNCS.

W. Bruce Croft. 1995. What do people want from information retrieval? (the top 10 research issues for companies that use and sell ir systems). *D-Lib Magazine*.

Ernesto William De Luca and Andreas Nürnberger. 2006a. Rebuilding Lexical Resources for Information Retrieval using Sense Folder Detection and Merging Methods. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genova, Italy.

Ernesto William De Luca and Andreas Nürnberger. 2006b. Using Clustering Methods to Improve Ontology-Based Query Term Disambiguation. *International Journal of Intelligent Systems*, 21:693–709.

Ernesto William De Luca. 2008. *Semantic Support in Multilingual Text Retrieval*. Shaker Verlag, Aachen, Germany.

Ernesto William De Luca. 2010. A corpus for evaluating semantic multilingual web retrieval systems: The sense folder corpus. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Susan T. Dumais, Edward Cutrell, and Hao Chen. 2001. Bringing order to the web: Optimizing search by showing results in context. In *CHI01*, pages 277–284.

Zakaria Elberrichi, Abdelattif Rahmoun, and Mohamed Amine Bentaalah. 2008. Using wordnet for text categorization. *The International Arab Journal of Information Technology*, 5.

Alfio Gliozzo, Carlo Strapparava, and Ido Dagan. 2004. Unsupervised and supervised exploitation of semantic domains in lexical disambiguation. *Computer Speech and Language*, 18(3):275–299.

Julio Gonzalo, Felisa Verdejo, Irina Chugur, and Juan Cigarrin. 1998. Indexing with wordnet synsets can improve text retrieval. In *Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems*, pages 38–44.

Frank Honza. 2005. Clustering mit a-priori annahmen über clusterspezifische attributwichtigkeiten. Master's thesis, University of Magdeburg, Germany.

Sherry Koshman, Amanda Spink, and Bernard J. Jansen. 2006. Web searching on the vivisimo search engine. *J. Am. Soc. Inf. Sci. Technol.*, 57(14):1875–1887.

Yannis Labrou and Timothy W. Finin. 1999. Yahoo! As an Ontology: Using Yahoo! Categories to Describe Documents. In *CIKM*, pages 180–187.

Jianqiang Li, Yu Zhao, and Bo Liu. 2009. Fully automatic text categorization by exploiting wordnet. In Gary Lee, Dawei Song, Chin-Yew Lin, Akiko Aizawa, Kazuko Kuriyama, Masaharu Yoshioka, and Tetsuya Sakai, editors, *Information Retrieval Technology*, volume 5839 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin / Heidelberg.

D. McCarthy, R. Koeling, J. Weeds, and J. Carroll. 2004. Finding predominant word senses in untagged text. In *42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain.

Rada Mihalcea, 2006. *Knowledge-Based Methods for WSD*, pages 107–131. Springer, Netherlands.

Alessandro Moschitti and Roberto Basili. 2004. Complex linguistic features for text classification: a comprehensive study. In *Proceedings of the 26th European Conference on Information Retrieval (ECIR*, pages 181–196. Springer Verlag.

Rila Mandala Tokunaga, Rila M, Tokunaga Takenobu, and Tanaka Hozumi. 1998. The use of wordnet in information retrieval. In *Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems*.

Giannis Varelas, Epimenidis Voutsakis, Paraskevi Raftopoulou, Euripides G.M. Petrakis, and Evangelos E. Milios. 2005. Semantic similarity methods in wordnet and their application to information retrieval on the web. In *Proceedings of the 7th annual ACM international workshop WIDM '05*, pages 10–16, New York, NY, USA. ACM.