Experiments in Automatic Word Class and Word Sense Identification for Information Retrieval

Published in: Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval.

Susan Gauch*
Robert P. Futrelle
*Electrical Engineering and Computer Science
1013 Learned Hall
University of Kansas
Lawrence, KS 66045
sgauch@eecs.ukans.edu

†Biological Knowledge Laboratory
College of Computer Science
Northeastern University
futrelle@ccs.neu.edu

Abstract

Automatic identification of related words and automatic detection of word senses are two long-standing goals of researchers in natural language processing. Word class information and word sense identification may enhance the performance of information retrieval systems. Large online corpora and increased computational capabilities make new techniques based on corpus linguistics feasible. Corpus-based analysis is especially needed for corpora from specialized fields for which no electronic dictionaries or thesauri exist. The methods described here use a combination of mutual information and word context to establish word similarities. Then, unsupervised classification is done using clustering in the word space, identifying word classes without pretagging. We also describe an extension of the method to handle the difficult problems of disambiguation and of determining part-of-speech and semantic information for low-frequency words. The method is powerful enough to produce high-quality results on a small corpus of 200,000 words from abstracts in a field of molecular biology.

1. Introduction

Information retrieval is an inexact science. While valuable information can be found, typically many irrelevant documents are also retrieved and many relevant ones are missed. One problem is terminology mismatches between the user's query and document contents. These mismatches occur for three main reasons:

Ambiguity: one word with multiple senses

Synonymy: two words with the same sense

Hyponymy: two words where one is an example of the other

Ambiguous query words can lead to "false drops" - the retrieval of documents which share some of the same words as the query, but do not share the same word sense as the query. Synonymy and hyponymy between words in the query and words in the documents may lead to relevant documents being missed. Of these problems, recall failures due to overlooking related terms (synonyms or hyponyms) seems to be the most serious, but ambiguity does contribute to lower precision [Krovetz & Croft, 1989].

Expanding a user's query with related terms can improve search performance. Relevance feedback systems where the related terms come from the contents of user-identified relevant documents have been shown to be quite effective [Harman, 1992]. An expert system which automatically reformulated queries by including terms from an online thesaurus was also able to improve search results [Gauch & Smith, 1993; Gauch & Smith 1991] without requiring relevance judgments from the user. Some systems [Anick et al, 1990] present related terms to the user and allow them to selectively augment the query. However, the latter two approaches require the presence of an online thesaurus whose words closely match the contents of the database.

Where can such a thesaurus come from? In some cases, it is hand-built [Gauch & Smith, 1991], a time-consuming and ad hoc process. In other cases, the thesaurus is an online version of a published thesaurus or semantically coded dictionary [Liddy & Myaeng, 1993]. However, an online published thesaurus or dictionary will have serious coverage gaps if used for technical domains which have their own distinct sublanguages. Because of ambiguity, this type of thesaurus may also be difficult to use with a database of general English documents because they show all possible classifications for a word when only one or a few senses may be actually present in the database.

Our goal is to automatically discover word classes directly from the contents of the textual database and to incorporate that information in an vector-space information retrieval system. We will modify and apply techniques from the field of corpus linguistics which seem particularly well-suited for this task. These techniques may also prove useful for summarizing the contents of retrieval sets [Futrelle & Gauch, 1993a] and for syntactic as well as semantic classification [Futrelle & Gauch, 1993b].

2. Corpus Linguistics Approach

Excellent methods have been developed for part-of-speech (POS) tagging using Markov models trained on partially tagged corpora [Church, 1988; Cutting, et al, 1992; Kupiec, 1992]. Semantic issues have been addressed particularly for sense disambiguation by using large contexts, e.g., 50 nearby words [Gale et al, 1992] or by reference to online dictionaries [Krovetz, 1991; Lesk, 1986; Liddy & Paik, 1992; Zernik, 1991]. More recently, methods to work with entirely untagged corpora have been developed which show great promise[Brill & Marcus, 1992; Finch & Chater, 1992; Myaeng & Li, 1992; Schütze, 1992]. They are particularly useful for specialized text with a specialized vocabularies and word-use. These methods of unsupervised classification typically have clustering algorithms at their heart [Jain & Dubes, 1988]. They use similarity of contexts as a measure of distance in the space of words and then cluster similar words into classes. This paper demonstrates a particular approach to this classification technology.

In our approach, we take into account both the relative positions of the nearby context words as well as the mutual information [Church & Hanks, 1990] associated with the occurrence of a particular context word. The similarities computed from these measures of the context contain information about both syntactic and semantic relations. For example, high similarity values are obtained for the two semantically similar nouns, "diameter" and "length" , as well as the two adjectives "nonmotile" and "nonchemotactic".

We demonstrate the technique on three problems, all using a 200,000 word corpus composed of 1700 abstracts from a specialized field of biology: #1: Generating the full classification tree for the 1,000 most frequent words (covering 80% of all word occurrences). #2: The classification of 138 occurrences of the *-ed* forms, "cloned" and "deduced" into four syntactic categories. #3: The classification of 100 words that only occur once in the entire corpus (*hapax legomena*), using expanded context information derived from #1. The results are discussed in terms of the semantic fields they delineate, the accuracy of the classifications and the nature of the errors that occur.

3. Corpus

The Biological Knowledge Laboratory is pursuing a number of projects to analyze, store and retrieve biological research papers, including working with full text and graphics. The work is focused on the biological field of bacterial chemotaxis. A biologist has selected approximately 1,700 documents representing all the work done in this field since its inception in 1965. This study uses the titles for all these documents plus all the abstracts available for them. The resulting corpus contains 227,408 words with 13,309 distinct word forms, including 5,833 words of frequency 1 and 1,889 words of frequency 2. There are 1,686 titles plus 8,530 sentences in the corpus. The sentence tagging algorithm requires two factors — contiguous punctuation (".", "!", or "?") and capitalization of the following token. To eliminate abbreviations (of which there are many in the corpus), the token prior to the punctuation must not be a single capital letter and the capitalized token after the punctuation may not itself be followed by a contiguous ".".

An example of a sentence from the corpus is,

> "$pre2$ $pre1$ one of the open reading frames was translated into a protein with pct amino acid identity to S. typhimurium FliI and pct identity to the beta subunit of E. coli ATP synthase $pos1$ $pos2$"

The positional items $pre... and $pos... have been added to furnish explicit context for sentence initial and sentence final constituents. Numbers have been converted to three forms corresponding to integers, reals and percentages.

The terminology we will use for describing words is as follows:

- Word form (or simply "word"): An orthographically distinguished item,
  "Cat" ≠ "cat"
- Word instance: An instance of a word form at some point in the text
- Labeled word instances: "the <u>cat1</u>"; "a <u>cat2</u>", to distinguish instances
- Target word: A word to be classified.
- Context words: Appearing within some distance of a target word,
  "The <u>big</u> <u>brown</u> <u>cat</u> <u>on</u> <u>the</u> mat...".
- Word class: Any defined set of word forms or labeled instances.

4. Similarity and Clustering

Word Contexts

Similar to [Finch & Chater, 1992], the context of a word (the target word) is described by the preceding two context words and the following two context words, Figure 1. Each position is represented by a vector corresponding to the occurrence of the 150 highest frequency words in the corpus, giving a 600-dimensional vector describing the context. Initially, the counts from all instances of a word form w_i are summed so that the entry in the corresponding context word position in the vector is the sum of the occurrences of that context word in that position for the corresponding target word form; it is the joint frequency of the context word.

Subsequently, a 600-dimensional vector of mutual information values, *MI*, is computed from the frequencies as follows,

$$MI(cw) = \log_2\left(\frac{Nf_{cw}}{f_c f_w} + 1\right)$$

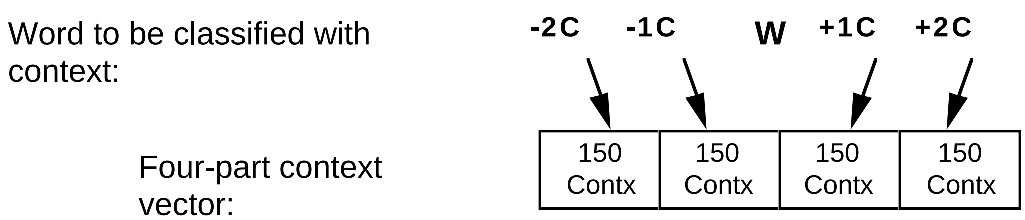

Figure 1. The 600-dimensional context vector around a target word W. Each subvector describes the frequency and mutual information of the occurrences of the 150 highest frequency words in the corpus.

This expresses the mutual information value for the context word *c* appearing with the target word *w*. The mutual information is large whenever a context word appears at a much higher frequency, f_{cw}, in the neighborhood of a target word than would be predicted from the overall frequencies in the corpus, f_c and f_w. The formula adds 1 to the frequency ratio, so that a 0 (zero) occurrence corresponds to 0 mutual information. Another smoothing strategy, used by some researchers [Church, et al., 1991], is capable of generating negative mutual information for the non-occurrence or low-frequency occurrence of a very high-frequency word and has the form,

$$MI(cw) = \log_2\left(\frac{N(f_{cw}+1)}{f_c f_w}\right)$$

In either case, some smoothing is necessary to prevent the mutual information from diverging when $f_{cw} = 0$.

Clustering and classification

When the mutual information vectors are computed for a number of words, they can be compared to see which words have similar contexts. The comparison we chose is the inner product, or cosine measure, which can vary between -1.0 and +1.0 [Myaeng & Li, 1992]. Once this similarity is computed for all word pairs in a set, various techniques can be used to identify classes of similar words. The method we chose is hierarchical agglomerative clustering [Jain & Dubes, 1988]. In this approach, the two words with the highest similarity are first joined into a two-word cluster. A new description for the cluster is generated and the cluster and remaining words are again compared, choosing the most similar to join. In this way, a binary tree is constructed with words at the leaves leading to a single root covering all words.

The updating was done by element-wise addition of the context frequencies in the vectors for the two words or clusters joined at each step. Each cluster, starting with individual target words has a frequency which is the sum of the frequencies of its two child clusters.

5. Experimental Results

Experiment #1: Classification of the 1,000 highest frequency words

The first experiment was to classify the 1,000 highest frequency words in the corpus by clustering, producing 999 clusters (0-998) during the process. $pre... and $pos... words were included in the context set, but not in the target set. Near the leaves, words clustered by syntactic class (part of speech) *and* by semantics (synonymy). Later, larger nodes tended to contain words of the same syntactic class, but with less semantic homogeneity. In each example below, the words listed are the entire contents of the node mentioned. The most striking property of the clusters produced was the classification of words into coherent semantic fields. Grefenstette has pointed out [Grefenstette, 1992] that the *Deese antonyms*, such as "large" and "small" or "hot" and "cold" show up commonly in these analyses. Our methods discovered entire graded domains, rather than just pairs of opposites. The following sample node shows a set of seventeen adjectives describing comparative magnitudes:

> decreased, effective, few, greater, high, higher, increased, large, less, low, lower, more, much, no, normal, reduced, short

Note that pairs such as "high" and "higher" and "low" and "lower" appear. "no", corresponding in meaning to "none" in this collection, is located at one extreme .

For obvious reasons, the entire tree of 1000 terms cannot be included in this paper. However, here are more sample nodes representative of the quality of the results. One node contains six concepts in the semantic domain of causal factors and the discovery of essential elements:

required, necessary, involved, responsible, essential, important

Another contains five compounds that are attractants used in chemotaxis studies:

aspartate, maltose, galactose, ribose, serine

Yet another consists of six physical units of measurement:

degrees, min, s, mM, microM, nm

At the lower levels of the tree, most nodes contain words that seem to belong together. However, given the limited context, the classification algorithm is bound to make mistakes, though a study of the text concordance will tell us why the algorithm failed in any specific case. For example, as the similarity drops, we see the adverb triple "greatly", "rapidly", "almost", which is still acceptable, then still later we see the triple, "them", "ring", "rings". This pronoun and the two nouns appear to have been grouped together because they are often sentence-final or followed by commas. At the end of the clustering, there is only a single cluster which includes all words and it forms stubbornly with a negative similarity of -0.51.

Experiment #2: Disambiguation of *-ed* forms

In addition to identifying related words, this technique can be used for disambiguation. The corpus uses the passive voice almost exclusively, with some use of the editorial "We". This results in a profusion of passive participles such as "detected", "sequenced" and "identified". But these *-ed* forms can also be simple past tense forms or adjectives. In addition, we identified their use in a postmodifying participle clause such as, "... the value <u>deduced</u> from this measurement." Each one of the 88 instances of "cloned" and the 50 instances of "deduced" was given a unique label that identified its part of speech and sentence number and clustering was applied to the resulting collection, giving the result shown in Figure 2A.

Using the classes identified in experiment #1, the results shown in Figure 2A were improved. Because we are dealing with single occurrences, only one element, at most, in each of the four context word vectors is filled, with frequency 1, the other 149 elements have frequency (and mutual information) 0.0. These sparse vectors will therefore have little or no overlap with vectors from other occurrences. In order to try to improve the classification, we expanded the context values in an effort to produce more overlap, using the following strategy: We proceed as if the corpus is far larger and in addition to the actual context words already seen, there would be many occurrences of highly similar words in the same positions. So for each non-zero context in each set of 150, we expand it to a set of similar words in the 150, picking words above a fixed similarity threshold (0.3 for the experiments reported here). Such a set is called a *word-rooted class*.

The apparent frequency of each additional (expansion) word was based on its corpus frequency relative to the corpus frequency of the word being expanded. For the expansion of a single context word instance c_i appearing at total frequency f_{ik} in the context of 1 or more occurrences of center word w_k, choose all c_j such that $c_j \in$ {set of high-frequency context words} and the similarity $S(c_i,c_j) \geq S_t$, a threshold value. These are the *expansion words*. Set the apparent frequency of each expansion word c_j to $f_{jk} = S(c_i,c_j) * f_{ik} * f_j / f_i$, where f_i and f_j are the corpus frequencies of c_i and c_j. Normalize the total frequency of the context word plus the apparent frequencies of the expansion words to f_{ik}. For the example being discussed here, $f_{ik} = 1$ and the average number of expansion words was 6.

Recomputing the classification of the *-ed* forms with the expanded context words results in the improved classification shown in Figure 2B. The number of classification errors is halved.

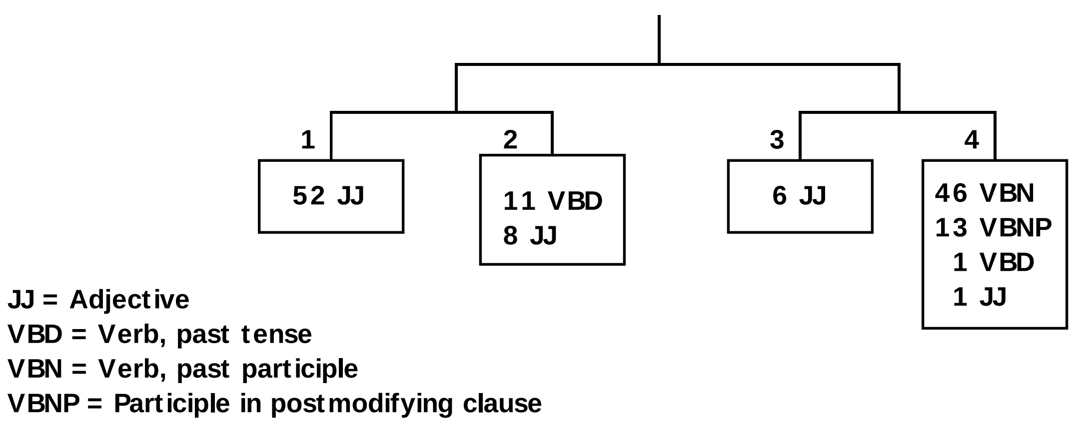

Figure 2A. Clustering of 88 occurrence of "cloned" and 50 occurrences of "deduced" into four syntactic categories. The abbreviations are based on [Francis & Kucera, 1982]. There is a strong admixture of adjectives in node 2 and all the postmodifiers are confounded with the past participles in node 4. The total number of errors (minority classes in a node) is 23 for an error rate of 23/138 = 17%.

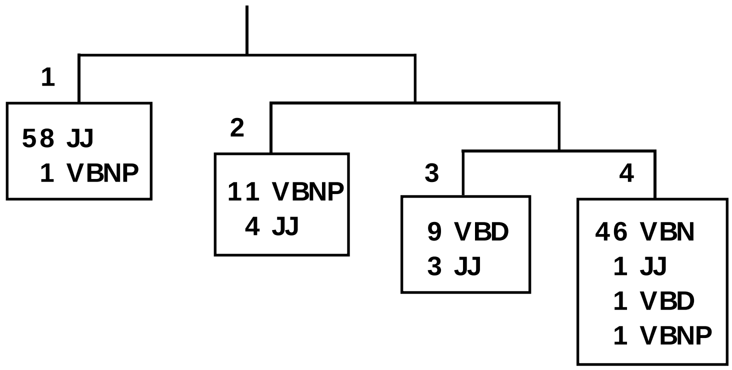

Figure 2B. Clustering of "cloned" and "deduced" after expansion of the context words. The postmodifying form, not isolated before, is fairly well isolated in its own subclass. The total number of errors is reduced from 23 to 11, to an error rate of 8%.

Experiment #3: Classification of single word occurrences

When classifying multiple instances of a single word form as we did in the previous example, there are numerous collocations that aid the classification. For example, there are 16 occurrences in the corpus of the phrase, "of the deduced amino acid sequence". But with words of frequency 1, we cannot rely on such similarities. Nevertheless, we experimented with classifying 100 words of corpus frequency 1 with and without expanding the context words. Though hand scoring the results is difficult, we estimate that there were 8 reasonable pairs found without using expansion and 26 pairs with expansion.

6. Discussion of Experimental Results

Experiment #1. Classifying 1000 words

An open issue is how to devise decision procedures that will tell us which classes are semantically or syntactically homogeneous, i.e., where to cut the tree into classes. The nice examples shown earlier broke down later on, when words or clusters began to be added that in our judgment were weakly related. Another fundamental question is - how good is good enough? Empirically, the lower nodes in the tree seem to contain semantically similar words, but are the words which are classified together good enough to be used as input to the query expansion process of an information retrieval system? There is some hope that this is the case because when the word-rooted classes were used to expand the context vectors used to classify single word instances in experiments 2 and 3.

Discussion of Experiment #2. Classifying *-ed* words

This experiment has major ramifications for the future. The initial classifications we did merged all identical word forms together, both as targets and contexts. But disambiguation techniques such as in this experiment can be used to differentially tag subsets of word occurrences, at least to some degree of accuracy. These newly classified items can in turn be used as new target and context items (if their frequencies are adequate) and the analysis can be repeated. Iterating the method in this way should be able to refine the classes until a fixed point is reached in which no further changes in the classification occur. The major difficulty with this approach will be to keep it computationally tractable.

Discussion of Experiment #3. Classifying frequency 1 words

The amount of information available about frequency 1 word can vary from a lot to zero, and most frequently tends to the latter, viz., "John and Mary looked at the blork." Nevertheless, such words are prominent, 44% of our corpus' vocabulary. From an information retrieval standpoint, these words are usually ignored. However, the improvement in their classification due to expanding the context vectors using the results of experiment #1 is encouraging.

7. Implications for Information Retrieval

We are developing intelligent interactive systems to help users navigate through online documents, focusing on techniques which are practical for large corpora and applicable to different domains. Current practice in information retrieval produces systems which fall short of desired performance, particularly for full-text databases [Blair & Maron, 1985]. Hopefully, the conceptual classification work presented here can be used to increase the number of relevant documents retrieved (i.e., improve recall) [Salton & McGill, 1983].

Currently, our retrieval system, the *Abstract Browser*, represents queries and abstracts as vectors which contain every unique word in the corpus, its weight in the abstract/query (using the tf*idf weighting scheme). These vectors are compared using the cosine similarity measure, and the top ranking abstracts are returned in response to a query. In a manner similar to that used to expand the context vectors for classification of singly occurring words, the query and/or abstract vectors could be expanded. For each word occurring in a query/abstract, the nearest neighbors can be located (either from the similarity matrix or the hierarchical cluster tree) and its weight in the vector increased. Thus, the vectors will contain non-zero entries not just for the words actually occurring in the text, but related words as well. The vectors will be broader in scope and, hopefully, matches will be found between text pieces which discuss similar concepts, not necessarily in the same words. A prototype using this approach is currently under development.

If the classification information does indeed improve the search results, a more efficient implementation is planned. In the planned second prototype, each classified word would have its expansion vector precalculated and stored in a lexicon. Then, the vector-space model would be extended to deal with entries that were not merely weights but rather vectors themselves. This approach is similar to the *MatchPlus* system [Gallant et al, 1993]. However, the context vectors in our system are vectors of related words identified by corpus linguistics techniques described earlier. The *MatchPlus* system uses vectors of words and other features learned by a neural net. Extensive experimentation will be necessary to determine which words are the most important to classify for retrieval purposes - low frequency? high-idf? some other group?

Although ambiguity does not appear to be a large problem in our narrowly focused technical corpus, for broader corpora it can be an issue. The disambiguation techniques discussed here could be applied to index and search on word senses rather than words. This should decrease the number of irrelevant documents retrieved (i.e., improve precision). Finally, indexing the corpora using phrases rather than single words may improve search results, although that remains to be demonstrated [Lewis, 1992]. Words which occur frequently near another word, with high mutual information, can be identified using the context vectors. These could be assumed to indicate a phrase and documents containing the words within a small context could be indexed with that phrase.

8. Conclusions

Corpus linguistics techniques based on word position and frequency yield promising results. We are able to automatically cluster words into classes which appear to contain related words. These classes can be used to improve word sense disambiguation, giving us hope that they will be of sufficient quality to improve the effectiveness of an information retrieval system.

References

Anick, P.G., Brennan, J.D., Flynn, R.A., Hanssen, D.R., Alvey, B., and Robbins, J.M. (1990). A Direct Manipulation Interface for Boolean Information Retrieval via Natural Language Query. Proc. 13th Ann. International ACM SIGIR Conf., (pp. 135-150). Brussels, Belgium: ACM Press.

Blair, D. C., & Maron, M. E. (1985). An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System. Comm. ACM, 28(3), 289-299.

Brill, E., & Marcus, M. (1992). Tagging an Unfamiliar Text with Minimal Human Supervision. In AAAI Fall Symposium Series: Probabilistic Approaches to Natural Language (Working Notes), (pp. 10-16). Cambridge, MA.

Church, K. W. (1988). A Stochastic Parts Program and Noun Parser for Unrestricted Text. In Proc. 2nd Conf. on Applied Nat. Lang. Processing, (pp. 136-143). Austin, TX.

Church, K.W., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. In U. Zernik (Eds.), Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon (pp. 115-164). Hillsdale, NJ: Lawrence Erlbaum.

Church, K. W., & Hanks, P. (1990). Word Association Norms, Mutual Information and Lexicography. Computational Linguistics, 16(1), 22-29.

Cutting, D., Kupiec, J., Pedersen, J., & Sibun, P. (1992). A Practical Part-of-Speech Tagger. In Proc. 3rd Conf. on Applied Natural Language Processing, (pp. 133-140).

Finch, S., & Chater, N. (1992). Bootstrapping Syntactic Categories Using Statistical Methods. In W. Daelemans & D. Powers (Ed.), Proc. 1st SHOE Workshop, (pp. 229-235). Tilburg U., The Netherlands.

Francis, W. N., & Kucera, H. (1982). Frequency Analysis of English Usage. Boston, MA: Houghton Mifflin.

Futrelle, R.P. & Gauch, S. (1993a). The Role of Automated Word Classification in the Summarization of the Contents of Sets of Documents. 2nd International Conf. on Information and Knowledge Management, (to appear). Washington, D.C.

Futrelle, R.P. & Gauch, S. (1993b). Experiments in Syntactic and Semantic Classification and Disambiguation Using Bootstrapping. Assoc. for Computational Linguistics Workshop on Acquisition of Lexical Knowledge from Text, (pp. 117-127), Cambridge, MA.

Gale, W. A., Church, K. W., & Yarowsky, D. (1992). Work on Statistical Methods for Word Sense Disambiguation. In AAAI Fall Symposium Series: Probabilistic Approaches to Natural Language (Working Notes), (pp. 54-60). Cambridge, MA.

Gallant, S.I., Hecht-Nielsen, R., Caid, W., Qing, K., Carleton, J., and Sudbeck, C.D. (1993). HNC's MatchPlus System, 1st Text Retrieval Conf. (TREC-1), (pp. 107-111). NIST #500-207.

Gauch, S., & Smith, J.B. (1993). An Expert System for Automatic Query Reformulation. J. of the Amer. Society of Inf. Sci., 44 (3), 124-136.

Gauch, S., & Smith, J.B. (1991). Search Improvement via Automatic Query Reformulation. ACM Trans. on Information Systems, 9 (3), 249-280.

Grefenstette, G. (1992). Finding Semantic Similarity in Raw Text: the Deese Antonyms. In AAAI Fall Symposium Series: Probabilistic Approaches to Natural Language (Working Notes), (pp. 61-66). Cambridge, MA.

Harman, D. (1992). Relevance Feedback Revisited. Proc. 15th Ann. International ACM SIGIR Conf., (pp. 1-10). Copenhagen, Denmark: ACM Press.

Jain, A. K., & Dubes, R. C. (1988). Algorithms for Clustering Data. Englewood Cliffs, NJ: Prentice Hall.

Krovetz, R. (1991). Lexical Acquisition and Information Retrieval. In U. Zernik (Eds.), Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon (pp. 45-64). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Krovetz, R., & Croft, W.B. (1989). Word Sense Disambiguation Using Machine-Readable Dictionaries. Proc. 12th Ann. International ACM SIGIR Conf., (pp. 127-136). Cambridge, MA: ACM Press.

Kupiec, J. (1992). Robust part-of-speech tagging using a hidden Markov model. Computer Speech and Language, 6, 225-242.

Lesk, M. (1986). Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from and Ice Cream Cone. In ACM SIGDOC Conf., (pp. 24-26). Toronto, Ont.: ACM Press.

Lewis, D.D. (1992). An Evaluation of Phrasal and Clustered Representations on a Text Categorization Task. Proc. 15th Ann. International ACM SIGIR Conf., (pp. 37-50). Copenhagen, Denmark: ACM Press.

Liddy, E.D., & Myaeng, S.H. (1993). DR-LINK's Linguistic-Conceptual Approach to Document Detection, 1st Text Retrieval Conf. (TREC-1), (pp. 113-129). NIST #500-207.

Liddy, E. D., & Paik, W. (1992). Statistically-Guided Word Sense Disambiguation. In AAAI Fall Symposium Series: Probabilistic Approaches to Natural Language (Working Notes), (pp. 98-107). Cambridge, MA:

Myaeng, S. H., & Li, M. (1992). Building Term Clusters by Acquiring Lexical Semantics from a Corpus. In Y. Yesha (Ed.), CIKM-92, (pp. 130-137). Baltimore, MD: ISMM.

Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81-106.

Salton, G., & McGill, M. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.

Schütze, H. (1992). Context Space. In AAAI Fall Symposium Series: Probabilistic Approaches to Natural Language (Working Notes), (pp. 113-120). Cambridge, MA:

Zernik, U. (1991). Train1 vs. Train2: Tagging Word Senses in Corpus. In U. Zernik (Eds.), Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon (pp. 97-112). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.