Combining General Hand-Made and Automatically Constructed Thesauri for Information Retrieval

Rila Mandala, Takenobu Tokunaga and **Hozumi Tanaka**
Department of Computer Science
Tokyo Institute of Technology
Ôokayama Meguro Tokyo 152-8552, Japan
{rila,take,tanaka}@cs.titech.ac.jp

Abstract

One of the most intuitive ideas for enhancing the effectiveness of an information retrieval system is to include the use of a thesaurus. WordNet, as a hand-crafted and general-purpose thesaurus, intuitively should also work fine in information retrieval, but unfortunately, experimental results by many researchers have not been promising. Thereby in this paper we investigate why the use of WordNet in information retrieval has not been successful. Based on this analysis we propose a method to combine WordNet with predicate-argument-based and co-occurrence-based automatically constructed thesauri. Experiments using large test collection shows that our method results in a significant improvement of information retrieval performance.

1 Introduction

The task of the information retrieval system is to match the query against the document collection and return relevant documents to the user. Retrieval performance is usually express in terms of *recall*, the proportion of relevant document retrieved, and *precision*, the proportion of retrieved document that is relevant. Whereas a perfect retrieval run will have a value of 1.0 for both recall and precision, in practice precision and recall are inversely related.

A critical problem in information retrieval is the case of different words being used to describe the same thing, either in queries or in documents. This require some kind of knowledge base which can equate or relate the terms used in language, i.e., a thesaurus. A thesaurus is a data structure which groups synonymous terms and relates them as either broader or narrower. A thesaurus can be used to expand a query to include all synonymous or related terms. This method has been known as query expansion method [Schutze and Pederson, 1997].

WordNet is currently the most large, hand-crafted, general-purpose, machine-readable, and publically available thesaurus. It is the product of a research project at Princeton University which has attempted to model the lexical knowledge of English [Miller, 1990]. WordNet has been used in numerous natural language processing, such as semantic tagging [Segond *et al.*, 97], word sense disambiguation [Resnik, 1995a], text categorization [Gomez-Hidalgo and Rodriguez, 1997], information extraction [Chai and Biermann, 1997], and so on with considerable success. However, the use of WordNet in information retrieval have not been very successful.

Two sets of experiments using the TREC collection were performed to investigate the effectiveness of using WordNet for query expansion by Voorhees [1994]. The first set used handpicked synsets and the second set extends the expansion strategy to include automatically selecting the starting synsets. When the concepts were chosen manually, her method could improve the retrieval effectiveness for short queries, but failed to improve the retrieval effectiveness for long queries. When the concepts were chosen automatically, none of the expansion methods produced significant improvement as compared with an unexpanded run. She further tried to use WordNet as a tool for word sense disambiguation [Voorhees, 1993] and applied it to text retrieval, but the performance of retrieval was degraded.

Stairmand [1997] used WordNet to investigate the computational analysis of lexical cohesion in text using lexical chain method [Morris and Hirst, 1991]. Because lexical chains are associated with topics, he suggested that information retrieval, where the notion of topic is very pertinent, is a suitable application domain. He concluded that his method only succeed in small-scale evaluation, but a hybrid approach is required to scale-up to real-word information retrieval scenarios.

Smeaton and Berrut [1995] tried to expand the queries of the TREC-4 collection with various strategies of weighting expansion terms, along with manual and automatic word sense disambiguation techniques. Unfortunately all strategies degraded the retrieval performance.

Instead of matching terms in queries and documents, Richardson [1995] used WordNet to compute the semantic distance between concepts or words and then used this term distance to compute the similarity between a query and a document. Although he proposed two methods to compute semantic distances, neither of them increased the retrieval performance.

2 Limitations of WordNet

In this section we analyze why WordNet has failed to improve information retrieval performance. We ran exact-match retrieval against 9 small standard test collections [Fox, 1990] in order to observe this phenomenon. An information retrieval test collection consists of a collection of documents along with a set of test queries. The set of relevant documents for each test query is also given, so that the performance of the information retrieval system can be measured. We expand queries using a combination of synonyms, hypernyms, and hyponyms in WordNet. The results are shown in Table 1.

In Table 1 we show the name of the test collection (Collection), the total number of documents (#Doc) and queries (#Query), and all relevant documents for all queries (#Rel) in that collection. For each document collection, we indicate the total number of relevant documents retrieved (Rel-ret), the recall ($\frac{\text{Rel-ret}}{\text{\#Rel}}$), the total number of documents retrieved (Ret-docs), and the precision ($\frac{\text{Rel-ret}}{\text{Ret-docs}}$) for each of no expansion (Base), expansion with synonyms (Exp. I), expansion with synonyms and hypernyms (Exp. II), expansion with synonyms and hyponyms (Exp. III), and expansion with synonyms, hypernyms, and hyponyms (Exp. IV).

From the results in Table 1, we can conclude that query expansion can increase recall performance but unfortunately degrades precision performance. We thus turned to investigation of why all the relevant documents could not be retrieved with the query expansion method above. Some of the reasons are stated below :

- Two terms that seem to be interrelated have different parts of speech in WordNet. This is the case between *stochastic* (adjective) and *statistic* (noun). Since words in WordNet are grouped on the basis of part of speech in WordNet, it is not possible to find a relationship between terms with different parts of speech.
- Most of relationships between two terms are not found in WordNet. For example how do we know that Sumitomo Bank is a Japanese company ?
- Some terms are not included in WordNet (proper name, etc).

To overcome all the above problems, we propose a method to enrich WordNet with an automatically constructed thesaurus. The idea underlying this method is that an automatically constructed thesaurus could complement the drawbacks of WordNet. For example, as we stated earlier, proper names and their interrelations among them are not found in WordNet, but if proper names and other terms have some strong relationship, they often co-occur in the document, so that their relationship may be modeled by an automatically constructed thesaurus.

Polysemous words degrade the precision of information retrieval since all senses of the original query term are considered for expansion. To overcome the problem of polysemous words, we apply a restriction in that queries are expanded by adding those terms that are most similar to the entirety of query terms, rather than selecting terms that are similar to a single term in the query.

3 Method

In this section, we first describe the construction method for each type of thesaurus utilized in this research, and then describe a term weighting method using similarity measure based on these thesauri.

3.1 WordNet

In WordNet, words are organized into taxonomies where each node is a set of synonyms (a synset) representing a single sense. There are 4 different taxonomies based on parts of speech and also there are many relationships defined within it [Fellbaum, 1998]. In this experiment we use only noun taxonomy with hyponymy/hypernymy (or the is-a relation), which relates more general and more specific senses.

The similarity between word w_1 and w_2 is defined as a shortest path from each sense of w_1 to each sense of w_2, as below [Resnik, 1995b]:

$$sim(w_1, w_2) = max[-log(\frac{N_p}{2D})]$$

where N_p is the number of nodes in path p from w_1 to w_2 and D is the maximum depth of the taxonomy.

3.2 Co-occurrence-based Thesaurus

The general idea underlying the use of term co-occurrence data for thesaurus construction is that words that tend to occur together in documents are likely to have similar, or related, meanings [Qiu and Frei, 1993]. Co-occurrence data thus provides a statistical method for automatically identifying semantic relationships that are normally contained in a hand-made thesaurus. Suppose two words (A and B) occur f_a and f_b times, respectively, and co-occur f_c times, then the similarity between A and B can be calculated using a similarity coefficient such as the Tanimoto Coefficient :

$$\frac{f_c}{f_a + f_b - f_c}$$

3.3 Predicate-Argument-based Thesaurus

In contrast with the previous section, this method attempts to construct a thesaurus according to predicate-argument structures [Hindle, 1990; Grafenstette, 1994; Ruge, 1992]. The use of this method for thesaurus construction is based on the idea that there are restrictions on what words can appear in certain environments, and in particular, what words can be arguments of a certain predicate. For example, a *dog* may *walk*, *bite*, but can not *fly*. Each noun may therefore be characterized according to the verbs or adjectives that it occurs with.

Table 1: Term Expansion Experiment Results using WordNet

Collection	#Doc	#Query	#Rel		Base	Exp. I	Exp. II	Exp. III	Exp. IV
ADI	82	35	170	Rel-ret	157	159	166	169	169
				Recall	0.9235	0.9353	0.9765	0.9941	0.9941
				Ret-docs	2,063	2,295	2,542	2,737	2,782
				Precision	0.0761	0.0693	0.0653	0.0617	0.0607
CACM	3204	64	796	Rel-ret	738	756	766	773	773
				Recall	0.9271	0.9497	0.9623	0.9711	0.9711
				Ret-docs	67,950	86,552	101,154	109,391	116,001
				Precision	0.0109	0.0087	0.0076	0.0070	0.0067
CISI	1460	112	3114	Rel-ret	2,952	3015	3,076	3,104	3,106
				Recall	0.9479	0.9682	0.9878	0.9968	0.9974
				Ret-docs	87,895	98,844	106,275	108,970	109,674
				Precision	0.0336	0.0305	0.0289	0.0284	0.0283
CRAN	1398	225	1838	Rel-ret	1,769	1,801	1,823	1,815	1,827
				Recall	0.9625	0.9799	0.9918	0.9875	0.9940
				Ret-docs	199,469	247,212	284,026	287,028	301,314
				Precision	0.0089	0.0073	0.0064	0.0063	0.0060
INSPEC	12684	84	2543	Rel-ret	2,508	2,531	2,538	2,536	2,542
				Recall	0.9862	0.9953	0.9980	0.9972	0.9996
				Ret-docs	564,809	735,931	852,056	869,364	912,810
				Precision	0.0044	0.0034	0.0030	0.0029	0.0028
LISA	6004	35	339	Rel-ret	339	339	339	339	339
				Recall	1.0000	1.0000	1.0000	1.0000	1.0000
				Ret-docs	148,547	171,808	184,101	188,289	189,784
				Precision	0.0023	0.0020	0.0018	0.0018	0.0018
MED	1033	30	696	Rel-ret	639	662	670	671	673
				Recall	0.9181	0.9511	0.9626	0.9640	0.9670
				Ret-docs	12,021	16,758	22,316	22,866	25,250
				Precision	0.0532	0.0395	0.0300	0.0293	0.0267
NPL	11429	100	2083	Rel-ret	2,061	2,071	2,073	2,072	2,074
				Recall	0.9894	0.9942	0.9952	0.9942	0.9957
				Ret-docs	267,158	395,280	539,048	577,033	678,828
				Precision	0.0077	0.0052	0.0038	0.0036	0.0031
TIME	423	24	324	Rel-ret	324	324	324	324	324
				Recall	1.000	1.000	1.000	1.000	1.000
				Ret-docs	23,014	29,912	33,650	32,696	34,443
				Precision	0.0141	0.0108	0.0096	0.0095	0.0094

Nouns may then be grouped according to the extent to which they appear in similar constructions.

First, all the documents are parsed using the Apple Pie Parser, which is a probabilistic chart parser developed by Sekine and Grishman [1995]. Then the following syntactic structures are extracted :

- Subject-Verb
- Verb-Object
- Adjective-Noun

Each noun has a set of verbs and adjective that it occurs with, and for each such relationship, a Tanimoto coefficient value is calculated.

- $C_{sub}(v_i, n_j) = \frac{f_{sub}(v_i,n_j)}{f(v_i)+f_{sub}(n_j)-f_{sub}(v_i,n_j)}$, where $f_{sub}(v_i, n_j)$ is the frequency of noun n_j occurring as the subject of verb v_i, $f_{sub}(n_j)$ is the frequency of the noun n_j occurring as subject of any verb, and $f(v_i)$ is the frequency of the verb v_i
- $C_{obj}(v_i, n_j) = \frac{f_{obj}(v_i,n_j)}{f(v_i)+f_{obj}(n_j)-f_{obj}(v_i,n_j)}$, where $f_{obj}(v_i, n_j)$ is the frequency of noun n_j occurring as the object of verb v_i, $f_{obj}(n_j)$ is the frequency of the noun n_j occurring as object of any verb, and $f(v_i)$ is the frequency of the verb v_i
- $C_{adj}(a_i, n_j) = \frac{f_{adj}(a_i,n_j)}{f(a_i)+f_{adj}(n_j)-f_{adj}(a_i,n_j)}$, where $f(a_i, n_j)$ is the frequency of noun n_j occurring as argument of adjective a_i, $f_{adj}(n_j)$ is the frequency of the noun n_j occurring as argument of any adjective, and $f(a_i)$ is the frequency of the adjective a_i

We define the similarity of two nouns with respect to one predicate as the minimum of each Tanimoto coefficient with respect to that predicate, i.e.,
$SIM_{sub}(v_i, n_j, n_k)=min\{C_{sub}(v_i, n_j), C_{sub}(v_i, n_k)\}$
$SIM_{obj}(v_i, n_j, n_k)=min\{C_{obj}(v_i, n_j), C_{obj}(v_i, n_k)\}$
$SIM_{adj}(a_i, n_j, n_k)=min\{C_{adj}(a_i, n_j), C_{adj}(a_i, n_k)\}$

Finally the overall similarity between two nouns is defined as the average of all the similarities between those two nouns for all predicate-argument structures.

3.4 Expansion Term Weighting Method

A query q is represented by a vector $\vec{q} = (q_1, q_2, ..., q_n)$, where the q_i's are the weights of the search terms t_i contained in query q.

The similarity between a query q and a term t_j can be defined as belows [Qiu and Frei, 1993]:

$$simqt(q, t_j) = \sum_{t_i \in q} q_i * sim(t_i, t_j)$$

Where the value of $sim(t_i, t_j)$ can be defined as the average of the similarity values in the three types of thesaurus.

With respect to the query q, all the terms in the collection can now be ranked according to their $simqt$. Expansion terms are terms t_j with high $simqt(q, t_j)$.

The $weight(q, t_j)$ of an expansion term t_j is defined as a function of $simqt(q, t_j)$:

$$weight(q, t_j) = \frac{simqt(q, t_j)}{\sum_{t_i \in q} q_i}$$

where $0 \leq weight(q, t_j) \leq 1$.

An expansion term gets a weight of 1 if its similarity to all the terms in the query is 1. Expansion terms with similarity 0 to all the terms in the query get a weight of 0. The weight of an expansion term depends both on the entire retrieval query and on the similarity between the terms. The weight of an expansion term can be interpreted mathematically as the weighted mean of the similarities between the term t_j and all the query terms. The weight of the original query terms are the weighting factors of those similarities.

Therefore the query q is expanded by adding the following query :

$$\vec{q_e} = (a_1, a_2, ..., a_r)$$

where a_j is equal to $weight(q, t_j)$ if t_j belongs to the top r ranked terms. Otherwise a_j is equal to 0.

The resulting expanded query is :

$$\vec{q}_{expanded} = \vec{q} \circ \vec{q_e}$$

where the $\circ$ is defined as the concatenation operator.

The method above can accommodate the polysemous word problem, because an expansion term which is taken from a different sense to the original query term is given very low weight.

4 Experimental Results

In order to evaluate the effectiveness of the proposed method in the previous section we conducted experiments using the TREC-7 information retrieval test collection [Voorhees and Harman, to appear 1999]. TREC-7 documents consists of the Financial Times (FT), Federal Register (FR94), Foreign Broadcast Information Service (FBIS) and the LA Times. Table 2 gives its document statistics, Table 3 give topic statistics, and Table 4 is one example out of 50 topics. As a baseline we used SMART [Salton, 1971] without expansion. SMART is an information retrieval engine based on the vector space model in which term weights are calculated based on term frequency, inverse document frequency and document length normalization. The results are shown in Table 5. This table shows the average of non-interpolated recall-precision for each of baseline, expansion using only WordNet, expansion using only predicate-argument-based thesaurus, expansion using only co-occurrence-based thesaurus, and expansion using all of them. For each method we give the percentage of improvement over the baseline. It is shown that the performance using the combined thesauri for query expansion is better than both SMART and using just one type of thesaurus.

Table 2: TREC-7 Document statistics

Source	Size (Mb)	# Docs	Median # Words/Doc	Mean # Words/Doc
Disk 4				
FT	564	210,158	316	412.7
FR94	395	55,630	588	644.7
Disk 5				
FBIS	470	130,471	322	543.6
LA Times	475	131,896	351	526.5

Table 3: TREC-7 Topic length statistics

Topic Section	Min	Max	Mean
Title	1	3	2.5
Description	5	34	14.3
Narrative	14	92	40.8
All	31	114	57.6

Title :
journalist risks

Description :
Identify instances where a journalist has been put at risk (e.g., killed, arrested or taken hostage) in the performance of his work.

Narrative :
Any document identifying an instance where a journalist or correspondent has been killed, arrested or taken hostage in the performance of his work is relevant.

Figure 1: Topic Example

Table 4: Average non-interpolated precision for expansion using combined thesaurus and expansion using only one type of thesaurus.

Topic Type	Base	Expanded with			
		WordNet only	Pred-Arg only	Co-occur only	Combined
Title	0.117	0.121 (+3.6%)	0.135 (+15.2%)	0.142 (+21.2%)	0.201 (+71.7%))
Desc	0.142	0.145 (+2.5%)	0.162 (+13.1%)	0.167 (+17.3%)	0.249 (+75.3%)
All	0.197	0.201 (+1.7%)	0.212 (+7.5%)	0.217 (+10.2%)	0.265 (+34.5%)

5 Discussion

In this section we discuss why our method of using WordNet is able to improve the performance of information retrieval. The important points of our method are :

- the coverage of WordNet is broadened
- weighting method

The three types of thesaurus we used have different characteristics. Automatically constructed thesauri add not only new terms but also new relationships not found in WordNet. If two terms often co-occur together in a document then those two terms are likely bear some relationship. Why not only use the automatically constructed thesauri ? The answer to this is that some relationships may be missing in the automatically constructed thesauri [Grafenstette, 1994]. For example, consider the words *tumor* and *tumour*. These words certainly share the same context, but would never appear in the same document, at least not with a frequency recognized by a co-occurrence-based method. In general, different words used to describe similar concepts may never be used in the same document, and are thus missed by the co-occurrence methods. However their relationship may be found in the WordNet thesaurus.

The second point is our weighting method. As already mentioned before, most attempts at automatically expanding queries by means of WordNet have failed to improve retrieval effectiveness. The opposite has often been true: expanded queries were less effective than the original queries. Beside the "incomplete" nature of WordNet, we believe that a further problem, the weighting of expansion terms, has not been solved. All weighting methods described in the past researches of query expansion using WordNet have been based on "trial and error" or ad-hoc methods. That is, they have no underlying justification.

The advantages of our weighting method are:

- the weight of each expansion term considers the similarity of that term with all terms in the original query, rather than to just one or some query terms.
- the weight of the expansion term accommodates the polysemous word problem.

This method can accommodate the polysemous word problem, because an expansion term taken from a different sense to the original query term sense is given very low weight. The reason for this is that, the weighting method depends on all query terms and all of the thesauri. For example, the word *bank* has many senses in WordNet. Two such senses are the financial institution and the river edge senses. In a document collection relating to financial banks, the river sense of *bank* will generally not be found in the co-occurrence-based thesaurus because of a lack of articles talking about rivers. Even though (with small possibility) there may be some documents in the collection talking about rivers, if the query contained the finance sense of *bank* then the other terms in the query would also concerned with finance and not rivers. Thus rivers would only have a relationship with the *bank* term and there would be no relationships with other terms in the original query, resulting in a low weight. Since our weighting method depends on both query in its entirety and similarity in the three thesauri, the wrong sense expansion terms are given very low weight.

We also experimented this method using other similarity coefficient method and Roget thesaurus, and found significant improvement of retrieval performance although the contribution of Roget thesaurus is very limited [Mandala *et al.*, to appear 1999].

6 Conclusion

This paper analyzed why the use of WordNet, a large and hand-made publically available thesaurus was not so succesful in improving the retrieval effectiveness in information retrieval applications. We found that the main reason is that most relationships between terms are not found in WordNet, and some terms, such as proper names, are not included in WordNet. To overcome this problem we proposed a method to enrich the WordNet with automatically constructed thesauri.

Another problem in query expansion is that of polysemous words. Instead of using a word sense disambiguation method to select the appropriate sense of each word, we overcame this problem with a weighting method. Experiments proved that our method of using WordNet in query expansion could improve information retrieval effectiveness.

In the future, we will use anaphora resolution to accurately determine the nature of relationships involving proper names. We will also investigate the effect of using different similarity coefficient method to build thesauri to the retrieval performance.

7 Acknowledgements

The authors would like to thank Mr. Timothy Baldwin (TIT, Japan) and three anonymous reviewers for useful comments on the earlier version of this paper. We also thank to Dr. Chris Buckley (SabIR Research) for the SMART support, and Dr. Satoshi Sekine (New York University) for providing the Apple Pie Parser program. This research is partially supported by the JSPS project number JSPS-RFTF96P00502.

References

[Chai and Biermann, 1997] J.Y. Chai and A. Biermann. The use of lexical semantics in information extraction. In *Proceedings of the ACL-EACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources*, pages 61–70, 1997.

[Fellbaum, 1998] Christiane Fellbaum. *WordNet, An Electronic Lexical Database.* MIT Press, 1998.

[Fox, 1990] Edward A. Fox. *Virginia Disk One.* Blacksburg: Virginia Polytechnic Institute and State University, 1990.

[Gomez-Hidalgo and Rodriguez, 1997] J.M. Gomez-Hidalgo and M.B. Rodriguez. Integrating a lexical database and a training collection for text categorization. In *Proceedings of the ACL-EACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources*, pages 39–44, 1997.

[Grafenstette, 1994] Gregory Grafenstette. *Explorations in Automatic Thesaurus Discovery.* Kluwer Academic Publisher, 1994.

[Hindle, 1990] Donald Hindle. Noun classification from predicate-argument structures. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistic*, pages 268–275, 1990.

[Mandala *et al.*, to appear 1999] Rila Mandala, Takenobu Tokunaga, and Hozumi Tanaka. Complementing WordNet with Roget and corpus-based thesauri for information retrieval. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics*, to appear, 1999.

[Miller, 1990] George A. Miller. Special issue, WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4), 1990.

[Morris and Hirst, 1991] Jane Morris and Graeme. Hirst. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 21–45, 1991.

[Qiu and Frei, 1993] Yonggang Qiu and Hans-Peter Frei. Concept based query expansion. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 160–169, 1993.

[Resnik, 1995a] Philip Resnik. Disambiguating noun grouping with respect to wordnet senses. In *Proceedings of 3rd Workshop on Very Large Corpora*, 1995.

[Resnik, 1995b] Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 448–453, 1995.

[Richardson and Smeaton, 1995] R. Richardson and Alan F. Smeaton. Using WordNet in a knowledge-based approach to information retrieval. Technical Report CA-0395, School of Computer Applications, Dublin City University, 1995.

[Ruge, 1992] Gerda Ruge. Experiments on linguistically-based term associations. *Information Processing and Management*, 28(3):317–332, 1992.

[Salton, 1971] Gerard Salton. *The SMART Retrieval System: Experiments in Automatic Document Processing.* Prentice-Hall, 1971.

[Schutze and Pederson, 1997] Hinrich. Schutze and Jan O. Pederson. A cooccurrence-based thesaurus and two applications to information retrieval. *Information Processing and Management*, 33(3):307–318, 1997.

[Segond *et al.*, 97] F. Segond, A. Schiller, G. Grefenstette, and J. Chanod. An experiment in semantic tagging using hidden markov model tagging. In *Proceedings of the ACL-EACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources*, pages 78–81, 97.

[Sekine and Grishman, 1995] Satoshi Sekine and Ralph Grishman. A corpus-based probabilistic grammar with only two non-terminals. In *Proceedings of the International Workshop on Parsing Technologies*, 1995.

[Smeaton and Berrut, 1995] Alan F. Smeaton and C. Berrut. Running TREC-4 experiments: A chronological report of query expansion experiments carried out as part of TREC-4. In *Proceedings of The Fourth Text REtrieval Conference (TREC-4)*, 1995. NIST special publication.

[Stairmand, 1997] Mark A. Stairmand. Textual context analysis for information retrieval. In *Proceedings of the 20th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pages 140–147, 1997.

[Voorhees and Harman, to appear 1999] Ellen M. Voorhees and Donna Harman. Overview of the Seventh Text REtrieval Conference (TREC-7). In *Proceedings of the Seventh Text REtrieval Conference*, to appear, 1999. NIST Special Publication.

[Voorhees, 1993] Ellen M. Voorhees. Using wordnet to disambiguate word senses for text retrieval. In *Proceedings of the 16th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pages 171–180, 1993.

[Voorhees, 1994] Ellen M. Voorhees. Query expansion using lexical-semantic relations. In *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pages 61–69, 1994.