TREC-2003 Novelty and Web Track at ICT

Jian Sun, Zhe Yang, Wenfeng Pan, Huaping Zhang, Bin Wang, Xueqi Cheng
Institute of Computing Technology (ICT), Chinese Academy of Sciences, P.R. China
sunjian@ict.ac.cn

1 Overview

In this paper, we will present our approaches and experiments on the following two tracks of TREC-2003: Novelty track and Web track.

The novelty track can be treated as a binary classification problem: relevant vs. irrelevant sentences, or new vs. non-new. In this way, we applied variants of techniques that have been employed for text categorization. To retrieve the relevant sentences, we compute the similarity between the topic and sentences using vector space model (VSM). In addition, we tried several techniques in an attempt to improve the performance: using narrative field and adopting dynamic threshold for different documents. We also have implemented the KNN algorithm and Winnow algorithm for classifying the sentences into relevant and irrelevant in the novelty task 3. To detect the new sentences, we used Maximum Marginal Relevance (MMR) measure, Winnow algorithm and so on. In addition, we attempted to detect novelty by computing semantic distance between sentences using WordNet.

For the Web track, we improved the basic SMART system, and the *Lnu-Ltu* weighting method was introduced into the system. The improved system has been proved to be effective in last year's task. In addition, we implemented a simple retrieval system using the probability model that is adopted by Okapi.

The structure of the paper is as follows: The section 2 reports the approaches and experiments in novelty track. The section 3 describes the experiments in web track. Finally, in section 4, we conclude by summarizing our experiments and presenting the future work.

2 Novelty Track

In the novelty track, there are four tasks which vary the kinds of data available to the systems and the kinds of results that need to be returned. In what follows, we will describe our approach to each task together with results of TREC experiments.

2.1 Task1 of Novelty Track

2.1.1 Relevant Sentences Retrieval

We retrieved the relevant sentences by comparing the topic to them using VSM and also applied several techniques in an attempt to improve the performance.

2.1.1.1 Vector Space Model (VSM)

In the VSM, the feature selection is required to decrease the dimensionality and improve the efficiency of classification. We used a χ^2 statistic, which measures the lack of independence between term *t* and topic *c*, to select features [Yang and Pedersen, 1997] for each topic in our novelty track.

We denote the feature set, which is obtained using χ^2 statistic, as F_1. Finally, the feature set F of one topic can be obtained using the following formula:

$$F = F_1 \cup (TW \cap \overline{S}) \tag{2.1}$$

where TW is the set of title words of the topic, $\overline{S}$ is the complement of stop words. The weighting of each term we used is also χ^2 statistic.

After the feature selection and weighting, both the topic and the sentence are represented as weighted vectors. The sentence is scored based on its similarity (in our experiments, we use cosine function) to the topic vector. If the similarity score is greater than a specific threshold δ, the sentence is regarded as relevant.

2.1.1.2 Dynamic threshold (DTH)

We observe that, in general, there is a specific time period during which the topic occurs. For example in topic N5 titled "*World Cup soccer*", the docs are mainly from June 03, 1998 to July 12, 1998. We call the specific time period "*focus*". Before and after the *focus*, there are fewer reports on the topic. As a result, we think the number of relevant sentences in docs not in *focus* is fewer than that in docs in *focus*. The observation motivates us to dynamically adjust the threshold according to the time of the document. Before and after the *focus* of the topic, the threshold is increased. Our strategy on the dynamic threshold for each topic is to: (1) obtain the starting and ending time of the all documents; (2) divide the time period into 4 equal time periods; (3) map each document to one of the 4 periods and obtain the number of docs in that period, *n;* (4) compute the document density, *density=n/N*, for each period. N is the number of docs.

The threshold of each period is as follows:

$$\delta = \begin{cases} \delta_0 - 0.05, & density > 0.75 \\ \delta_0, & density > 0.5 \\ \delta_0 + 0.05, & density > 0.25 \\ \delta_0 + 0.1, & density > 0 \end{cases} \tag{2.2}$$

where δ_0 is the basic threshold, which is obtained using the TREC-2002 Novelty data.

2.1.1.3 Enlarging the data (XTD)

There are at most 25 relevant documents for each topic. We want to know whether the data is enough or not. Therefore we investigated whether the performance can be improved by enlarging the text data. We retrieved more documents (75 docs for each topic) from AQUAINT collection using the SMART system. The query for each topic was the content words in the title field. As a result, there will be 100 documents for each topic. We then used the same feature selection method to determine the terms and their weighting.

2.1.1.4 Using Narrative Field (NAR)

We know that the narrative field describes the information requirement in detail: (1) what is we need (e.g. *References to Dolly's children are relevant if Dolly's name is included.*) and (2) what is irrelevant (e.g. *Mention of Polly and Molly are not relevant*). Traditional methods usually did not use the information (we call it *negative information*) in (2). In order to utilize the negative information, we first obtained the negative features by selecting the content words in the narrative field that tell us what information is not needed, and then built a negative vector for each topic.

We determine the relevance of the sentence by computing the following similarity as follows:

$$sim(topic, s_i) = sim(V_{si}, V_{tp}) - sim(V_{si}, V_{tn}) \tag{2.3}$$

where V_{si} denotes the vector of sentence *i*, V_{tp} the (positive) vector of topic, V_{tn} the negative vector of topic.

2.1.1.5 Features based on local co-occurrence (CUR)

We also tried another method of extracting the features. For example, we adopted the approach, *local co-occurrence*, proposed by [Zhang. etc, 2002]. The fixed window we used was -2 ~ 2. The weighting was also χ^2 statistic. One official run using the method was submitted.

2.1.2 Novelty Detection

As for the novelty detection, we applied two methods: word overlapping between two sentences and maximum marginal relevance (MMR).

2.1.2.1 Word Overlapping between two sentences

The word overlapping (OLP) between two sentences, s_i and s_j, is similar to [Zhang. etc, 2002]. The method did not use the similarity between topic and sentence. Since the method is simple, we did not use it in task1. In task2, we submitted one run that used the method.

2.1.2.2 Maximum Marginal Relevance (MMR)

Another approach to select novelty sentences from relevant sentences is Maximum Marginal Relevance (MMR) [Carbonell and Goldstein, 1998] measure, which is given by:

$$MMR(s_i) = \lambda Sim_1(V_{si}, V_{tp}) - (1-\lambda)\max_{S_j \in R} Sim_2(V_{si}, V_{sj})] \tag{2.4}$$

where R is the set of selected relevant sentences, and Sim_1 and Sim_2 are similarity metrics.

To obtain the novelty sentences from the relevant set, we first ranked the relevant sentences according to one of above measures and then selected a specific percentage from them.

2.1.3 Submitted Results

We submitted five runs for task 1. The five runs were:

1) ICT03NOV1BSL: local co-occurrence, threshold=0.48; MMR
2) ICT03NOV1SQR: χ^2 statistic, threshold=0.40; MMR
3) ICT03NOV1NAR: NAR, local co-occurrence, threshold=0.20; MMR
4) ICT03NOV1XTD: XTD, χ^2 statistic, threshold=0.20; MMR
5) ICT03NOV1DTH: DTH, local co-occurrence, threshold=0.46; MMR

The results of our official runs at TREC-2003 Novelty Task 1 were shown in Table 2.1. We observed that the run that adopted the dynamic threshold achieved the best performance of five runs.

Table 2.1 Performance of Official Run of Novelty Task 1

Run#	Relevant Part			Novelty Part		
	P	R	F	P	R	F
ICT03NOV1BSL	0.62	0.51	0.486	0.41	0.48	0.379
ICT03NOV1SQR	0.63	0.54	0.489	0.40	0.49	0.368
ICT03NOV1NAR	0.58	0.46	0.434	0.37	0.44	0.334
ICT03NOV1XTD	0.61	0.39	0.408	0.39	0.36	0.310
ICT03NOV1DTH	0.63	0.50	0.489	0.42	0.44	0.370

2.2 Task2 of Novelty Track

For the novelty detection of task 2, we select a specific percentage of relevant sentences as new sentences. We believe the percentage of new sentences in relevant sentences decreases as the ID of document increases for each topic. To model the intuition, we applied a simple method, *dynamic percentage*, as follows:

1) set the average percentage p_v, which is obtained using the TREC-2002 data, of new sentences in relevant sentences
2) set the percentage p_1 of new sentences in relevant sentences of 1th document: $p_1 = p_v + 12.5\%$
3) set the percentage p_i of new sentences in relevant sentences of ith document: $p_i = p_1 - i*1\%$

2.2.1 Submitted Results

The five runs we submitted were:

1) ICT03NOV2SQR: χ^2 statistic, MMR
2) ICT03NOV2CUR: local co-occurrence, MMR
3) ICT03NOV2PNK: χ^2 statistic, NAR described in Section 2.1.1.4, MMR
4) ICT03NOV2LPP: OLP described in Section 2.1.2.1, dynamic percentage
5) ICT03NOV2LPA: OLP described in Section 2.1.2.1, fixed percentage

Table 2.2 showed our official runs at TREC-2003 Novelty Task 2. Comparing the results of the five runs, we noticed that the performance of run ICT03NOV2LPA was better than that of other runs. Actually, the run only applied the simplest techniques, i.e. counting the words that occur in both sentences.

2.3 Task3 of Novelty Track

For task 3, we concentrated on the retrieval of relevant sentences. We implemented KNN algorithm and Winnow algorithm for selecting the relevant sentences. The method of novelty detection was similar to that of task 1.

2.3.1 KNN algorithm for Retrieval of Relevant Sentences

In the task 3, we examined the KNN algorithm at the sentence level. And two strategies were taken to predict the class (relevant or irrelevant) of a sentence. One was that the prediction will be the class that has the largest number of members in the *k* nearest neighbors. The other was that the class with maximal average similarity will be the winner. These two strategies were denoted as KNN1 and KNN2, respectively.

2.3.2 Winnow algorithm for Retrieval of Relevant Sentences

The Winnow [Dagan, 19997] algorithm has been shown that it functions well in text domain. In the experiments presented here, we used it at the sentence level.

In this experiment, the strength of the feature is taken to indicate only the presence or absence of it in the sentence, that is, it is either 1 or 0.

2.3.3 Submitted Results

The five runs we submitted were:

1) ICT03NOV3KNN: all content words as features, KNN1; MMR
2) ICT03NOV3IKK: all content words as features, KNN2; MMR
3) ICT03NOV3KNS: χ^2 statistic, KNN2; MMR
4) ICT03NOV3WND: Winnow; MMR, dynamic percentage
5) ICT03NOV3WN3: Winnow; MMR, fixed percentage

The Table 2.3 showed our official runs at TREC-2003 Novelty Task 3. Comparing the first three runs, the run ICT03NOV3KNS that applied feature selection achieved better results. Comparing the last two runs with the first three runs, we observed that the precision increased using the Winnow algorithm while the recall decreased. As for the novelty detection, we made a mistake in the official run and the first relevant sentence in the 6^{th} doc was taken to be the first relevant sentence in the topic by us. The mistake resulted in the bad performance of the novelty detection. We believe that we can improve the performance of novelty track further in future work.

Table 2.2 Performance of Official Run of Novelty Task 2

Run#	Novelty Part		
	P	R	F
ICT03NOV2SQR	0.65	0.74	0.677
ICT03NOV2CUR	0.65	0.73	0.677
ICT03NOV2PNK	0.65	0.73	0.676
ICT03NOV2LPP	0.65	0.74	0.679
ICT03NOV2LPA	0.73	0.87	0.783

Table 2.3 Performance of Official Run of Novelty Task 3

Run#	Relevant Part			Novelty Part		
	P	R	F	P	R	F
ICT03NOV3KNN	0.57	0.56	0.547	0.35	0.37	0.346
ICT03NOV3IKK	0.57	0.58	0.548	0.36	0.39	0.348
ICT03NOV3KNS	0.60	0.58	0.572	0.37	0.39	0.362
ICT03NOV3WND	0.65	0.53	0.557	0.39	0.35	0.346
ICT03NOV3WN3	0.68	0.49	0.537	0.43	0.41	0.381

2.4 Task4 of Novelty Track

To detect the new sentences from the relevant sentences for the task 4, we applied several methods, such as MMR measure and Winnow algorithm. In addition, we attempted to detect novelty by computing semantic distance between two sentences using WordNet.

2.4.1 Winnow Algorithm for Novelty Detection

Since we know the new sentences and non-new sentences in the relevance for the task 4, we can train a classifier using Winnow algorithm. The classifier represents a sentence as a set of features F = $\{f_1, f_2 \ldots f_m\}$. The number of active features in the sentence we used was 5. We compute the strength of each feature as follows:

(1) compute the similarity between the sentence and the topic, f_1';

(2) compute the similarities between the sentence and all of those thatoccurred before it using VSM and obtain the two biggest similarities, f_2' and f_3';

(3) compute the word overlapping between the sentence and all of those that occurred before it and obtain the two biggest overlapping, f_4' and f_5';

(4) compute the strength of each feature as $f_1 = f_1$', $f_2 = 1- f_2$', $f_3 = 1- f_3$', $f_4 = 1- f_4$' and $f_5 = 1- f_5$', respectively.

The weight vector was estimated on training data using Winnow algorithm. After the weight vector was obtained, the Winnow algorithm was used to predict the novel sentence.

2.4.2 Computing Semantic Distance using WordNet for Novelty Detection

2.4.2.1 Motivation

Compared with the traditional IR, the Novelty track returned ranked only new and relevant sentences rather than a large amount of relevant document. The information content within a sentence is very small. Traditional methods can be used in sentence level; however, we think it is not the best choice if we only focus on word form, as almost all words occur once within a sentence. For example, there are two sentences in the N12 topic:

1) *Daily we read news stories about dissatisfaction with managed care, Medicare fraud and overbilling*.
2) *Eighty percent agreed with this*.

Both sentences described opinions on universal healthcare. However, it's difficult to detect relevance or novelty between them, since the words "dissatisfaction" and "agreed" seem irrelevant in terms of word form.

2.4.2.2 WordNet-based Semantic Distance between Words

We assume that the distance between same words is zero. If the two words have different word form, we compute the distance $Dist(w_1,w_2)$ as described in [Jiang and Conrath, 1997].

We have found that in many cases, it is still far from requirement if we only cover hypernym between words. Based on the above work, we introduce more word relations. They mainly include similarity and derivation between words. For example, “friendly’ is derived from “friend” and “friendly” is similar to “amicably”. We assigned the distance between such words with 0.5.

Apart from above relations, distance between any other words is set to be a large value.

2.4.2.3 Computation on Semantic Distance between Sentences

Before semantic distance between sentences, we define word-sentence semantic distance (WSSD) to be the minimum distance between the word w and words within the sentence S. Therefore, we estimate $WSSD(w, S)$ with the following formula:

$$WSSD(w, S) = \min \{Dist(w,w_i) | w_i \in S\} \quad (2.5)$$

where w is a word, S is a sentence and w_i is a word in sentence S.

Based on $WSSD(w, S)$, we define sentence distance SSSD(Sa,Sb) as follows:

$$\text{SSSD(Sa, Sb)} = \frac{\sum_{w_i \in Sa} WSSD(w_i, Sb) + \sum_{w_j \in Sa} WSSD(w_j, Sa)}{|Sa| + |Sb|} \quad (2.6)$$

where Sa and Sb are sentences; |Sa| and |Sb| are word numbers in sentence, respectively.

2.4.2.4 Novelty Detection

For novelty detection, we consider the following features:

1) *f1*: Semantic distance between a relevant sentence S and topic T.Let it be $SSSD(S, T)$. Naturally, S is more likely to be new if $SSSD(S, T)$ is less.
2) *f2*: minimal semantic distance from S to previous valid context P. Let it be $SSSD(S, P)$.
3) *f3*: Word overlapping from sentence S to topic T. Let it be $Overlapping(S, T)$. For two sentence Si and Sj, we define word overlapping with:

   $$Overlapping(\text{Si,Sj}) = \frac{|\text{Si} \cap \text{Sj}|}{|\text{Si}|} \quad (2.7)$$

   Similarly, larger $Overlapping(S, T)$ indicates that the sentence is more relevant to the topic.
4) *f4*: Word overlapping with previous valid context P. Let it be $Overlapping(S, T)$. Less $Overlapping(S, T)$ tends to be new sentence.
5) *f5*: Is S a head sentence? Let it be $IsHead(S)$. $IsHead(S)$ is equal to 1 if S is a head sentence in the paragraph. Otherwise, it is 0.

Then we can define a 5-tuple feature vector as $F=(f_1, f_2, f_3, f_4, f_5)=(1- SSSD(S,T), SSSD(S,P), Overlapping(S,T), 1-Overlapping(S,P), IsHead(S))$. After defining the features, we apply Winnow algorithm to estimate weight.

For the task 4, the novel sentences in the first 5 documents were treated as positive training set, while other relevant sentences were treated as negative training set. After the weight vector was obtained, the Winnow algorithm predicted the novel sentence as described in Section 2.4.1.

2.4.3 Submitted Results

We submitted two runs (ICT03NOV4SQR and ICT03NOV4WNW) using the methods described in Section 2.4.1. The MMR was adopted in run ICT03NOV4SQR and the Winnow algorithm was adopted in run ICT03NOV4WNW. The three runs (ICT03NOV4ALL, ICT03NOV4LFF and ICT03NOV4OTP) adopted the methods described in Section 2.4.2. The detailed feature vectors of the three runs were shown in Table 2.4:

Table 2.4 The three runs and their features

Run#	Feature vector
ICT03NOV4OTP	(f_3, f_4)
ICT03NOV4LFF	(f_1, f_2, f_3, f_4)
ICT03NOV4ALL	$(f_1, f_2, f_3, f_4, f_5)$

Table 2.5 Performance of Official Run of Novelty Task 4

Run#	P	R	F
ICT03NOV4ALL	0.60	0.68	0.598
ICT03NOV4LFF	0.59	0.64	0.568
ICT03NOV4OTP	0.59	0.70	0.610
ICT03NOV4SQR	0.61	0.66	0.623
ICT03NOV4WNW	0.65	0.72	0.636

The Table 2.5 showed our official runs at TREC-2003 Novelty Task 4. From the results, the run ICT03NOV4WNW achieved better performance than other runs. Although the computation of semantic distance using WordNet did not show any improvement in our official runs, we think that there are still potential improvements if we can make good use of prior knowledge and other information imbedded in the sentence.

3 Web Track

3.1 Introduction

This year, Web track consists of two subtasks: the Named/Home Page Finding task, and the Topic Distillation task. The former task is introduced to investigate methods for effective navigational search, with a mixture of home page and named page queries: finding a particular page desired by the user. This task involves a mixture of tasks from two previous years: home page finding and named page finding. In both cases, there is only one target page and user's queries are often the name of the page. For the Topic Distillation task, it is introduced to investigate methods for finding a set of the best home pages given a broad query. For this task, the key is to find as many different websites (represented by their entry pages) as possible within the first ten results. The test collection of this year's Web track is *.Gov* data set as the last year.

As the last year, our retrieval system was based on SMART. We modified the basic SMART system, and the *Lnu-Ltu* weighting method was introduced into the system. This system has been proved to be effective in last year's task. In addition, we implemented a simple retrieval system using the probability model that is adopted by Okapi.

3.2 Named/Home Page Finding

As introduced in above section, the goal of named page finding task is to find the page that named as user's query. For the home page finding task, the difference is that home page finding queries are restricted to home pages. Since in named/home page finding task the user explains his goal explicitly, every word in the query is more important than that be in ad-hoc task, which is a tough reason to request nearly all the words in the query appearing in the relevant document and content should be emphasized. The run ICTWebKI12A is an original result retrieved by content, which determine other runs' performance.

As the task's name, the target page always has a name being similar with the query, and the title of web page is also a very important component. To some extent, the anchor text is also the page's name: it is the index by which users can visit the page from other pages. This useful structure information is the key of the improvement of performance. In the run ICTWebKI12B, we combined the scores of content, title and the anchor text into a unified measure using the linear interpolation, and obtained a better result than the original result.

In this year's task, home page finding topics are mixed with named page finding topics. If we can divide these topics, the URL can be used in the home page finding task. For this purpose, we used a simple strategy: the topic described as entity, such as a special person, a special location or a special organization, was judged to be a home page finding topic. We used two different combining methods to the divided topics: for the named page finding topics, content, anchor text and title was used, for the home page finding topics, URL length was added to be an important factor which computes the probability of a page to be a home page. The run ICTWebKI12C divided the topics into NP topics and HP topics as described, and obtained a better result than ICTWebKI12B.That is to say: our simple method to identify home page finding task is useful. The Table 2.6below showed the results of three runs for this year's task.

Table 2.6 Results of Named/home Page Finding Task in TREC-2003

RunId	MRR	Founded Answers	Not Found
ICTWebKI12A	0.308	207/300	93/300
ICTWebKI12B	0.449	247/300	53/300
ICTWebKI12C	0.568	265/300	35/300

3.3 Topic Distillation

As described in the TREC-2003 Web Track Guideline, to be judged a "key resource", the page returned should be a good entry point to a website which: 1) is principally devoted to the topic, 2) provides credible information on the topic and 3) is not part of a larger site also principally devoted to the topic.

In TREC-2002, almost all the participants reached the same conclusion: The structure information will hurt the effect of retrieve. Our further experiments proved it again: Only anchor text or title can only get 0.03 for P@10, and the performance of retrieve almost can not benefit from such a result. In TREC-2002, our original retrieval result get 0.2360 at P@10, that is to say that there are more than 2 relevant documents in the first ten results. Such a good result motivated us to adopt the pseudo relevance feedback. Assuming the first five or ten results as relevance documents, we used Rocchio method to expend the original queries, and then retrieved with the new queries and obtained the final results. In our experiments, a small number of relevant documents assumed (such as 5) and moderate expended query terms (such as 15) can get a considerable improvement. These experiments suppose us to use pseudo relevance feedback in this year's task. But in this year's task, no query was offered like before: only narratives were given. We must create queries manually and it leads to disastrous original results.

Pseudo relevance feedback is influenced by the original results, and bad original results lead to worse final results. The table 2.7 below showed the different results of pseudo relevance feedback based on different original results.

Table 2.7 Feedback results and Original results in TREC-2002 and TREC-2003

Run	Average precision	P@10
Orig_2002	0.1620	0.2306
Fdb_2002	0.1748	0.2510
Orig_2003	0.0728	0.0520
Fdb_2003	0.0639	0.0380

Table 2.8 Multiple retrieval systems in TREC-2003

Run	Average precision	P@10
1.VSM	0.0728	0.0520
2.VSM with stemming	0.0571	0.0480
3.Okapi	0.0441	0.0320
4.Okapi with stemming	0.0440	0.0480
5.1+2+3+4	0.1036	0.0536

In addition, we used multiple retrieval systems vote mechanism. We combined the results of four different retrieval systems: VSM with stemming, VSM without stemming, probability model with stemming and probability model without stemming. As the results shown in Table 2.8, VSM system without stemming represented better than the others, and multiple retrieval systems vote mechanism showed its contribution for improving the performance.

4. Summary and Future Work

This paper presented our work in the Novelty track and Web track evaluated at TREC-2003. In the Novelty track, we applied the KNN algorithm and Winnow algorithm to retrieve relevant sentences. To detect the new sentences, we tried several methods, including semantic distance between sentences using WordNet, MMR measure, Winnow algorithm and so on. We also conducted our experiments on: (1) the use of χ^2 statistic for feature selection, (2) dynamic threshold for different document to retrieve relevant sentence, (3) the use of narrative field in topic, and (4) dynamic percentage of relevant sentences as novelty sentences within a document.

In the experiments, we observed that the performance of novelty detection greatly depends on the system's ability to retrieve the relevant sentences. The χ^2 statistic for feature selection and dynamic threshold for different document to retrieve relevant sentence were shown to be effective in the track.

The experiments on the Web track showed that the vote mechanism can improve the performance and the VSM retrieval systems without stemming always work a little better than those systems with stemming.

Our future work includes: (1) Further studying the problem of how to expand the information of sentence level using WordNet or other resources for similarity computation; (2) Exploiting the use of type information of topic; (3) Investigating how to determine the number of features for each topic.

References

J. Allan, C. Wade and A. Bolivar. Retrieval and Novelty Detection at the Sentence Level, Proceedings of the 26th annual international ACM SIGIR, pp314-321, 2003

T. Ault and Yiming Yang. kNN, Rocchio and Metrics for Information Filtering at TREC 10. Proceedings of the Tenth Text Retrieval Conference, pp 84-93. 2001

J. Carbonell and J. Goldstein, The use of MMR, Diversity-based Reranking for reordering document and producing summaries. Proceeding of SIGIR 98, pp335-336, 1998

Dagan, Ido, Yael Karov and Dan Roth. Mistake-driven learning in text categorization, in Proceedings of Second Conference on Empirical Methods in Natural Language Processing (EMNLP-2), 1997

Rocchio, J. J. Relevance Feedback in Information Retrieval. In The SMART Retrieval system, Prentice-Hall, Englewood NJ. 232-241,1971

Jay J. Jiang and David W. Conrath, Semantic similarity based on corpus statistics and lexical taxonomy. In proceedings of International Conference Research on Computational Linguistics, 1997.

H. Kazawa, T. Hirao, H. Isosaki and E. Maeda. A machine learning approach for QA and Novelty Tracks: NTT system description. To appear in the Proceedings of the Eleventh Text Retrieval Conference, 2002

G.Salton,Chris Buckley. 1990. Improving Retrieval Performance By Relevance Feedback. JASIS 41.288-297.

Thijs Westerveld, Wessel Kraaij, and Djoerd Hiemstra. 2001. RetrievingWeb Pages using Content, Links, URLs and Anchors, In The Tenth Text REtrieval Conference (TREC 10), pp663, 2001

Yang, Y., Pedersen J.P. A Comparative Study on Feature Selection in Text Categorization Proceedings of the Fourteenth International Conference on Machine Learning (ICML'97), pp412-420, 1997

M. Zhang, R. Song, C. Lin, S. Ma, Z. Jiang, Y. Jin, Y. Liu, L. Zhao and S. Ma. Expansion-Based Technologies in Finding Relevant and New Information: THU TREC2002 novelty track experiments. To appear in the Proceedings of the 11th TREC Conference, 2002