Using Co-occurrence Data for Query Expansion: Wrong Paradigm or Wrong Formulas?

Reginald Ferber, GMD - IPSI, Dolivostr. 15, D - 64293 Darmstadt, Germany, ferber@darmstadt.gmd.de

The paper discusses possible reasons for the failure of studies using co-occurrence data for query expansion. It suggests that the choice of similarity measures, the way expansion is done and the size of the corpus used to extract the co-occurrence data may be the reasons for this failure and not the co-occurrence paradigm per se. This view is substantiated by results of a study, that simulated the selection of search terms by professional searchers of a bibliographic data base.

1. Indexing, Retrieval, and Co-Occurrence Data

The last thirty years have seen several approaches to use co-occurrence data to enhance indexing and query construction ranging from hard wired machines like ACRON (Associative Content Retrieval Network) with 40 “documents” and 40 “terms” (Giuliano & Jones 1963) over studies on probabilistic retrieval and term selection (Van Rijsbergen 1977, Willett 1985) to approaches using coarse syntactic analysis to extract co-occurrences (Grefenstette 1992, Ruge 1992, Jing & Croft 1994) or an approach based on a logical model (Crestani & Van Rijsbergen 1995). Many of these approaches have been judged as not successful (Salton & Buckley 1988). Peat and Willett (1991) for example state: "The weight of the experimental evidence to date hence suggests that query expansion based on term cooccurrence data is unlikely to bring about substantial improvements in the performance of document retrieval systems.” (page 379)

However, there are several possible reasons for the failure of such studies that are not due to the use of co-occurrence data per se, but can be found in the way these data are used:

1. some studies used similarity measures that favored frequent terms
2. the expansion was often done for each single query term in isolation, and not for the query as a whole
3. the size of the test collections from which the co-occurrence data were extracted was rather limited resulting in weak estimations of probabilities of co-occurrence

This paper discusses some of these points and then presents data from a study which tried to avoid these problems.

The Vector Space Model

Most studies use the vector space model (see for example Jones & Furnas (1987), Deerwester, Dumais, Furnas, Landauer & Harshman (1990), or Peat & Willett (1991)): For a set of *objects* or *documents* $O = \{o_1, ..., o_m\}$ and a set of *(index-)terms* $T = \{t_1, ..., t_n\}$ there is a $m \times n$–matrix $W = \{w_{ij}\}_{\substack{i=1..n \\ j=1..m}}$ which describes the relevance of the terms for the documents. The entry $w_{ij} \in \mathbb{R}$ is the *relevance* or *weight* of term i for document j. Hence the columns of the matrix represent the documents as vectors in an n-dimensional space.

To construct a document vector, in general the terms that occur in the document are taken as index terms. Their weight can depend on their overall statistical properties and on the place and frequency of their occurrence in the document. To retrieve documents a query is transformed into a (query-) vector and the document vectors are identified, that are close to that query vector.

Most similarity measures used for this purpose are based on the presence of terms in both the query and the document.

Within this process three steps are crucial:

1. The indexing transformation forming a document vector out of a document,
2. the query transformation constructing a query vector out of a query,
3. the measure of similarity that compares these vectors

Of course these three operations are not independent. What is done in one could often as well be done in the other. In this paper we concentrate on the interaction of similarity measures and query construction: How to expand the query by new terms, to enhance the results of a search.

The meaning of the expression "query expansion" is not uniform within Information Retrieval research. Whereas in the theory of feedback methods it is used in the sense of adding more terms to an existing and previously used (query language) query, Peat and Willet (1991) seem to use it in the sense of adding more terms to those extracted from a natural language query given by a user. At least the queries coming with the Cranfield collection they used are natural language queries. This is the way it will be used in this paper.

Two fundamental problems of information retrieval are synonymy and polysemy (see Deerwester, Dumais, Furnas, Landauer, & Harshman 1990). Synonymy means that people use different terms to describe the same object; polysemy means that people describe different objects using the same term (in different contexts). In the vector space model synonyms are treated as two different terms. If one is used in a document and the other in a query they will not contribute to the similarity of the corresponding vectors.

Ambiguous words have only one index term for all of their meanings. If one meaning is mentioned in a query, documents in which the term appears with the other meanings are also estimated as closer to the query.

These problems can be attacked using associations between terms. Whereas human searchers often use terms that "come into their mind" for a specific problem, automated systems can use associations calculated on the basis of co-occurrence of terms in huge corpora of texts of the respective domain. The basic idea is in both cases the same: In the construction of a query vector the weights of the terms associated to the terms occurring in the query are increased. In the case of two terms describing the same object it is likely that the terms associated to the first term are mainly the same as those associated to the second term. If one term is used in a document and the other is used in a query the increase in the weights of the associated terms in the query vector will move the query vector and the document vector closer together.

In case of polysemy there will be terms associated to both meanings of an ambiguous term, but the query will contain more terms that are related to its intended meaning. These terms and their associated terms will form a cluster, which is associated to the intended meaning and which outweighs the unintended meaning. Hence the vector will be moved away from the unintended meaning and closer to the intended one.

These mechanisms are not restricted to real synonyms and ambiguous words. If several terms are specific to a topic that is searched for, but not all of them are mentioned in the query, the weights of those not mentioned will be increased through the associations. If a term is used in several topics the weights of other terms of the topic searched for will be increased by other terms of the query.

Whereas the above description is given in terms of vector similarity the same mechanism can also be used for query expansion. All terms can be ranked according to their weights in the

query vector, such that the terms with the highest weights will get the lowest ranks. These terms can be used to expand the query.

Similarity Measures

Several formulas have been used to calculate the associations between terms from co-occurrence data. Peat and Willet (1991) analyzed three different similarity measures

$$COSINE(X,Y) = \frac{H(X,Y)}{\sqrt{H(X)H(Y)}}$$

$$DICE(X,Y) = \frac{2 \cdot H(X,Y)}{H(X) + H(Y)}$$

and

$$TANIMOTO(X,Y) = \frac{H(X,Y)}{H(X) + H(Y) - H(X,Y)}$$

where $H(X)$ denotes the number of documents in which the term X occurs and $H(X,Y)$ the number of documents in which the terms X and Y occur together.

Peat and Willet (1991) summarize their results:

> This article demonstrates that the similar terms identified by cooccurrence data in a query expansion system tend to occur very frequently in the data base that is being searched. Unfortunately, frequent terms tend do discriminate poorly between relevant and nonrelevant documents, and the general effect of query expansion is thus to add terms that do little or nothing to improve the discriminatory power of the original query (p. 378)

This statement would be more precise if it would say that the terms identified by the co-occurrence data *with the formulas used* tend to occur very frequently. The selection of a specific formula is not implied by the use of co-occurrence data, but a decision made by the designer of the model. With a different formula one will get different results from the same co-occurrence data.

To reduce the number of frequent terms selected by a similarity measure, one can analyze the measure and change it in such a way that less frequent terms are selected. To do so, one can interpret the formulas as probabilities of co-occurrence of terms instead of interpreting them as similarity measures:

Let $p(X)$ denote the probability that term X occurs in a document and $p(X\&Y)$ the probability that term X and term Y occur together in one document. If the occurrence of the two terms is statistically independent then

$$p(X\&Y) = p(X) \cdot p(Y)$$

holds by the definition of statistical independence. This means that the quotient

$$\frac{p(X\&Y)}{p(X) \cdot p(Y)}$$

is less than 1 iff the two terms occur less often together than expected in case of independence, it is equal to 1 iff they co-occur by chance, and it is larger than 1 if they occur more often together than expected in case of independence.

Replacing probabilities by relative frequencies one gets a similarity measure or “association”

$$U(X,Y) = A \cdot \frac{H(X\&Y)}{H(X) \cdot H(Y)} \tag{6}$$

where A is the number of documents in the data base i. e. a constant factor that can be ignored in the following considerations. A comparison of this measure and the measures used by Peat & Willett (1991) reveals that their measures favor frequent terms. In case of the $COSINE(X,Y)$ there is a factor $F_C = \sqrt{H(X)H(Y)}$:

$$COSINE(X,Y) = \frac{H(X,Y)}{H(X)H(Y)} \cdot \sqrt{H(X)H(Y)} = F_C \cdot U(X,Y)$$

This means that the more frequent the terms are, the bigger is the factor by which the measure differs from U and hence more frequent terms are more likely to be added to a query. This is the effect described by Peat and Willett (1991).

To calculate the factor F_D for the Dice measure we start with

$$f \cdot DICE(X,Y) = U(X,Y)$$

$f \in \mathbb{R}$ and get

$$f = \frac{H(X,Y)}{H(X) \cdot H(Y)} \cdot \frac{2(H(X) + H(Y))}{H(X,Y)} = 2\frac{H(X) + H(Y)}{H(X,Y)}$$

and

$$f = 2\left(\frac{1}{H(Y)} + \frac{1}{H(X)}\right)$$

thus with $F_D = \frac{1}{f}$ we get altogether

$$DICE(X,Y) = F_D \cdot U(X,Y)$$

Again the Factor F_D increases with increasing frequency of the terms, producing a similar effect as with the cosine measure.

For the Tanimoto measure

$$f \cdot TANIMOTO(X,Y) = U(X,Y)$$

leads to

$$f = \frac{H(X,Y)}{H(X) \cdot H(Y)} \cdot \frac{H(X) + H(Y) - H(X,Y)}{H(X,Y)}$$

and

$$f = \frac{H(X) + H(Y) - H(X,Y)}{H(X) \cdot H(Y)} = \frac{1}{H(Y)} + \frac{1}{H(X)} - \frac{H(X,Y)}{H(X) \cdot H(Y)}$$

From $H(X) \geq H(X,Y)$ and $H(Y) \geq H(X,Y)$ it follows that

$$\frac{1}{H(Y)} + \frac{1}{H(X)} \geq \frac{H(X,Y)}{H(X) \cdot H(Y)}$$

and further that

$$f \leq \frac{1}{H(Y)} + \frac{1}{H(X)} =: g$$

Hence finally with $F_T = \frac{1}{g}$ it holds that

$$TANIMOTO(X,Y) = \frac{1}{f} \cdot U(X,Y) \geq F_T \cdot U(X,Y)$$

Again the factor F_T increases with increasing frequencies of terms and hence the measure favors frequent terms.

In the study described in this paper the associations were computed using formula (6) from which the value in case of independence was subtracted. This leads to formula

$$V(X,Y) = A \cdot \frac{H(X\&Y)}{H(X) \cdot H(Y)} - 1 \tag{18}$$

as association between the terms X and Y.

Expansion of Single Terms or Expansion of a Query

Most studies do query expansion in such a way that they add new terms to every single term in a query (Rijsbergen, Harper & Porter 1981, Peat & Willett 1991, Grefenstette 1992).

Many terms used in human communication are ambiguous or have several meanings. But in most cases these ambiguities are resolved without any problem, or even without noticing the ambiguity. The way this is done by humans in still an open problem of psychological research, but it is almost certain, that the context in which a term occurs plays a central role.

The expansion of single terms of a query ignores the context in which a term is used. There is no way that influences coming from several terms of the query can accumulate in a term or that negative associations coming from a query term can lower the probability of another term to be selected for expansion. A very simple way to allow such influences is a linear model, that results in a superposition of the influences coming from all query terms.

A related aspect is the size of the query to be expanded. If it consists of only a few terms, this superposition of influences is rather limited. If the query is longer, there are more terms to influence the selection of new terms.

Size of the Data Base

Most studies used document test collections ranging in size from less than 100 (Giuliano & Jones 1963) up to approximately 27,000 documents (Peat & Willett 1991). These are rather small sizes to estimate statistical properties of terms. Studies simulating association experiments using co-occurrence data have shown that the results rely heavily on the size of the corpora they are based on (Rapp 1991, Rapp 1993, Wettler, Rapp & Ferber 1993). This is not astonishing if one takes into consideration that the relative frequencies of words in language are rather small, and that the co-occurrence of words is of course even smaller.

Evaluation

Experiments in information retrieval are in general evaluated using document test collections and precision and recall curves. This method relies on many influences like the composition of the collection and the relevance judgements of raters. Harman (1992) states: “Performance improvements for query expansion using the probabilistic model seem to be heavily dependent on the test collection being used.” (page 2)

On the other hand a lot of research was done in the last years concerning the behavior of searchers (Belkin & Vickery (1985), Fidel (1984, 1991a, 1991b, 1991c), Glöckner-Rist (1993)). In 1988 Saracevic and Kantor (1988) formulated the following program:

> A main component of the basic research agenda for information science for time to come should be (i) (...) (ii) test of models of information seeking and retrieving involving the human elements, be they users or intermediaries. (...) To build a machine (including an intelligent interface with a machine) that does some information searching tasks at least as well as humans do, we must first study the patterns of human behavior, as well as the patterns that relate relevant texts (in whatever form or image) to questions and problems at hand. (page 215)

The study presented in this paper restricts the evaluation to the simulation of the behavior of professional searchers in a bibliographic data base. The aim is to predict which terms a searcher would use given a written end user request.

2. The Simulation

In this section we describe a simulation model for the selection of search terms by professional data base searchers. More details can be found in Ferber (1992) and Ferber, Wettler, and Rapp (1995).

The Model

The simulation was based on 94 search records and the co-occurrence of terms in the documents of the data base PsycLIT (1989). The basic ideas of the model were the following:

1. For every pair of terms used in the records an association was calculated based on their co-occurrence in a sample of documents of the data base and formula (18). These associations were stored in a weight matrix.
2. Then for every single user request a query vector was constructed that contained the number of occurrences of the terms in the request.
3. This vector was multiplied by the matrix containing the associations to include the influence of associated terms.
4. Finally the terms were ordered according to their entry in the vector; within this ranking the mean ranks

   a. of those terms that the searcher selected from the request,
   b. of those terms that the searcher added, and
   c. of those query terms that the searcher did not use

   were compared

A good simulation would yield low ranks for the terms used in the query and high ranks for those not used. In the following sections the model is described in more detail.

Material, Terms, and Frequencies

The searches were made independently of the study by professional searchers at a German psychological information agency. Each record consisted of an end-user's written request and the corresponding searches in the data bases PSYCINFO and PSYNDEX. An example of a record is given in Table 1

Figure 1: An End - User Request and the Search of the Professional Searcher

Inhaltliche Beschreibung der Fragestellung in Form eines Arbeitstitels in deutscher, möglichst auch in englischer Sprache: Einfluß von Geschlechtsstereotypen auf sexuelles Verhalten. Influence of sex role stereotypes on sexual behavior
Suchstichworte möglichst aus der anglo-amerikanischen Fachsprache: 1.) Sex role stereotype 2.) Androgyny

```
PSYNDEX
1 177 Find CT All (Sex Role Att$;Feminism;Feminity;Masculinity)
2 11 Find ALL Androgyn$/PQ
3 734 Find CT D Psychosexual Behavior
4 428 Find 3 Not CT=Sex Roles
5 17 Find (1;2) And 4
```

The problem descriptions in the records as well as the searches were partly in English and partly in German. Altogether the records contained 2108 different German and English words. From these words a set of terms was constructed in such a way that German words and their English translations as well as different morphological variants and shortened forms with identical root were combined into word groups. In what follows these groups of words will be called *terms*. A term's occurrence in a text was defined as the occurrence of any of its member words in the text. This procedure corresponded to stemming and translation and resulted in 1061 terms. From these terms only those 872 that occurred in 40 or more documents were used for the simulation. The co-occurrence data for these terms were taken from the free-text fields (i.e. title, abstract, or key phrase) of documents of the data base PsycLIT.

Calculation of the Associations

With the co-occurrence data and formula (18) raw values for the associations were calculated. For terms with low frequencies the estimation of their probability of co-occurrence by the relative frequency is unreliable (Gale & Church, 1990). To smooth the values their range was reduced to $[-t, 1]$, $0 \leq t \leq 1$, using a monotonic nonlinear sigmoidal transformation with two parameters.[1] The associations were organized in a 872×872–matrix. The values in the diagonal of the matrix were calculated with the same formula which can be written in this case as

$$V(X,X) = A \cdot \frac{H(X\&X)}{H(X) \cdot H(X)} - 1 = A \cdot \frac{1}{H(X)} - 1 \tag{19}$$

and submitted to the same transformation as the other values. They describe the fraction of the weight of a term that is kept during the multiplication. These elements are small for frequent terms and larger for less frequent terms. A factor on the diagonal elements was included as parameter.

[1] The transformation was composed of two functions $f(x) = \frac{a}{x-b} + c$ for $x \leq 0$ and $g(x) = \frac{d}{x-e} + h$ for $x \geq 0$ with the same value $f(0) = g(0) = 0$ and the same slope $f'(0) = g'(0) = m$ at the point 0 . With the additional assumptions $f(-1) = -t$ and $g(max) = 1$ for the maximal value max the values of $a, b, c, d, e, h \in \mathbb{R}$ can be determined. m, the slope in 0 and t the minimal value are the two parameters.

Simulation of Query Expansion

To simulate the searchers selection of terms for a query the terms of the end user request were read in automatically and a query vector was constructed, that contained for each term the number of its occurrences in the request. This vector was multiplied by the matrix with the associations and the terms were ranked according to their weights.

Figure 2: Results of the Simulation of Example 1

Rank	Class	Weight	Term
1	P&¬Q	0.009727	STEREOTYPEN STEREOTYPES STEREOTYPE
2	P&Q	0.009691	ANDROGYN ANDROGYNY
3	¬P&Q	0.008722	MASCULINITY
4	P&¬Q	0.007178	GENDER GESCHLECHT
5	P&Q	0.007116	SEX SEXUAL SEXUALITAT SEXUELLE SEXUELLES
6	¬P&Q	0.005652	FEMINIS FEMINISM
7		0.004986	HOMOSEXUALITY
8		0.004283	MASTURBATION
9	P&Q	0.003922	ROLE ROLES ROLLE ROLLEN
10		0.00378	INTERCOURSE INTERCOURSES
11		0.003707	LIBERAL
12		0.003591	MAN MEN
13		0.00356	FRAU FRAUEN WOMAN WOMEN
14	P&¬Q	0.003453	EINFLUA EINFLUSSEN INFLUENCE INFLUENCING INFLUENCES
15		0.003453	ORGASM
16	¬P&Q	0.00328	PSYCHOSEXUAL
17		0.003242	OCCUPATION
18		0.003032	IDENTITY
19		0.002815	PARTNER PARTNERN PARTNERS PARTNERSCHAFT PARTNERSCHAFTLICH
20		0.002813	FEMALE FEMALES WEIBLICHEN
56	P&Q	0.001462	BEHAVIOR BEHAVIORAL BEHAVIORS VERHALTEN VERHALTENS
195	P&¬Q	0.000579	AUF ON
225	P&¬Q	0.000448	VON FROM VOM
439	P&¬Q	0.000008	OF AUS

The terms are ranked according to their weights. The first 20 ranks are given completely as well as all terms that occurred in the record. (For the classification see section 5)

Evaluation of a single Simulation

The terms of each record were grouped into three classes:

1. *P&Q-terms*: the terms, that appeared in the problem description and in the query
2. *P&¬Q-terms*: the terms, that appeared in the problem description but not in the query
3. *¬P&Q-terms*: the terms that did not appear in the problem description but in the query

The result of a simulation can be given as the mean ranks of the terms of these three classes.

Another way to evaluate the selection of search terms an overlap measure. Saracevic and Kantor (1988) used such a measure to show that the agreement between searchers is rather low: For 40 search problems they computed an overlap between the terms used by two searchers for the same problem with the formula:

$$O(R_1, R_2) = \frac{\#(Q(R_1) \cap Q(R_2))}{\#Q(R_1)}$$

where $Q(R_i)$ is the set of query terms used in search R_i and $\#$ is the number of elements of a set (p. 203). They got a mean overlap value of 0.27.

In a similar way one can compute an overlap between the query terms of searchers and the same number of terms with the highest activities using the formula:

$$O(Q, R) = \frac{\#(Q \cap \{r(1), ..., r(\#(Q))\})}{\#(Q)}$$

with $r(i)$ denoting the term with activity rank i.

Figure 3: Overall results of the simulation

Calibration sample:

#DOC	P&Q	¬P&Q	P&¬Q	OVL
246,889	18.5	155.6	194.4	0.39
136.887	18.4	159.8	186.8	0.38
75,000	19.4	165.9	165.2	0.40
40,000	21.8	183.5	147.8	0.38
20,000	21.3	186.3	164.4	0.38

Test sample:

#DOC	P&Q	¬P&Q	P&¬Q	OVL
246,889	18.9	172.2	203.7	0.41
136.887	18.5	174.7	196.3	0.39
75,000	18.5	197.7	170.2	0.40
40,000	21.5	211.6	150.8	0.37
20,000	20.1	228.8	178.1	0.37

#DOC is the number of documents in the subsample of the data base used to calculate the co-occurrence data. The vocabulary and all parameters were the same in all simulations. As a consequence the minimum frequency of a term in the data base could be less than 40 for simulations based on less than the complete data base. The simulation could not be run with smaller numbers of documents (which would have been comparable to the number of documents in the other document test collections used by Peat and Willett (1991)) because some terms of the vocabulary did not occur in any of those smaller subsamples of documents.

Evaluation of the Model

To control the performance of the simulation the set of records was divided at random into two samples: One half, the *calibration sample*, was used to check out good parameter values, the other half, the *test sample*, was kept aside to test the model. With the calibration sample and the co-occurrence data of the whole corpus a set of parameter values was chosen, such that the mean activity rank of the terms in the query (P&Q and ¬P&Q) was low under the additional restriction that the mean activity rank of the ¬P&Q-terms was lower than that of the P&¬Q-terms.

Results

Table 3 gives the results of the simulations. First it can be observed, that the mean ranks of the query terms are much better than they could be expected by chance (This would be a mean of 436). Second the results for the two samples are quite similar. This shows that the model has some general validity. The largest differences can be found for the new terms selected by the searcher. The most stable results can be found for the P&Q-terms, that the searcher took from the request.

Under decreasing number of documents in the corpus used to extract the co-occurrences the same pattern can be observed: The mean ranks of the new terms increase most, while the mean ranks of the P&Q-terms are again quite stable. This supports the assumption that for query expansion the use of a large corpus is of great importance. Another reasons for this behavior could be that the set ¬P&Q is the smallest of the three sets. Hence the statistical basis for the estimations is small. But the selection of new terms is also the most demanding of the three tasks because there is no input to these terms from the problem description. The complete weight has to come from the associations. That makes it probably the one that is most sensitive to a reduction of the corpus from which the co-occurrences are taken.

The overlap values are quite stable. That is probably due to the large number of terms in P&Q compared to ¬P&Q. All overlap values are larger than those found by Saracevic and Kantor (1988). Although it is difficult to compare the results, because they are based on different material and procedures, it looks as if the results from the simulations were at least as good as they could be expected from another human searcher.

Altogether the study gives some evidence, that the use of co-occurrence data for query expansion can lead to useful results, if the model in which the data is used is designed appropriately and the corpus from which the co-occurrences are taken is large enough.

References

Belkin, N. J., & Vickery, A. (1985). Interaction in information systems. A review of research from document retrieval to knowledge-based systems. Library and Information Research Report 35, The British Library Board.

Crestani, F., & Van Rijsbergen, C. J. (1995). Information retrieval by logical imaging. *Journal of Documentation 51*(1), 3-17.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science 41*(6), 391-407.

Ferber, R. (1992). Vorhersage der Suchwortwahl von professionellen Rechercheuren in Literaturdatenbanken durch assoziative Wortnetze. In *Mensch und Maschine – Informationelle Schnittstellen der Kommunikation. (Proceedings ISI '92)* (1992), H. H. Zimmermann, H.-D. Luckhardt, & A. Schulz, Eds., Universitätsverlag Konstanz, 208–218.

Ferber, R., Wettler, M., & Rapp, R. (1995). An associative model of word selection in the generation of search queries. *Journal of the American Society for Information Science (JASIS) 46*(9), 685-699.

Fidel, R. (1984). Online searching styles. a case-study-based model of searching behavior. *Journal of the American Society for Information Science 35*, 211-221.

Fidel, R. (1991a). Searchers' selection of search keys: I. The selection routine. *Journal of the American Society for Information Science 42*, 490-500.

Fidel, R. (1991b). Searchers' selection of search keys: II. Controlled vocabulary or free-text searching. *Journal of the American Society for Information Science 42*, 501-514.

Fidel, R. (1991c). Searchers' selection of search keys: III. Searching styles. *Journal of the American Society for Information Science 42*, 515-527.

Gale, W. A., & Church, K. W. (1990). Poor estimates of context are worse than none. In *DARPA Speech and Natural Language Workshop* (Hidden Valley, PA, 1990), 283-287.

Giuliano, V. E., & Jones, P. E. (1963). Linear associative information retrieval. In *Vistas in Information Handling*, P. W. Howerton & D. C. Weeks, Eds., vol. 1. Spartan Books, Washington D. C., Washington, D.C., ch. 2, 30-54.

Glöckner-Rist, A. (1993). *Suchfragen im Information Retrieval*. Universitätsverlag Konstanz.

Grefenstette, G. (1992). Use of syntactic context to produce term association lists for text retrieval. In *Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (1992), ACM SIGIR, 89-97.

Jing, Y., & Croft, W. B. (1994). An association thesaurus for information retrieval. In *Proceedings of the RIAO 94* (1994), vol. 1, 146-160.

Jones, W. P., & Furnas, G. W. (1987). Pictures of relevance: A geometric analysis of similarity measures. *Journal of the American Society for Information Science 38(6),* 420-442.

Peat, H. J., & Willett, P. (1991). The limitations of term co-occurrence data for query expansion in document retrieval systems. *Journal of the American Society for Information Science 42*(5), 378-383.

Rapp, R. (1991). Ein statistisches Modell für die maschinelle Sprachverarbeitung. *Grundlagenstudien aus Kybernetik und Geisteswissenschaft 32*(4), 163-176.

Rapp, R. (1993). A statistical model to verify verbal material. In *Sprache-Kommunikation-Informatik. Akten des 26. Linguistischen Kolloquiums.* (1993), J. Darski & Z. Vetulani, Eds., 543-548.

Rijsbergen, C. J. v., Harper, D. J., & Porter, H. F. (1981). The selection of good serach terms. *Information Processing and Management 17*, 77-91.

Ruge, G. (1992). Experiments on linguistically-based term associations. *Information Processing and Management 28*(3), 317-332.

Salton, G., & Buckley, C. (1988). On the use of spreading activation methods in automatic information retrieval. In *Proceedings of the eleventh Annual International Conference on Research and Development in Information Retrieval* (1988), ACM, 147-160.

Saracevic, T., & Kantor, P. (1988). A study of information seeking and retrieving. III searchers, searches and overlap. *Journal of the American Society for Information Science 3*(39), 197–216.

Van Rijsbergen, C. J. (1977). A theoretical basis for the use of co-occurrence data in information retrieval. *Journal of Documentation 33*(2), 106 - 119.

Wettler, M., Rapp, R., & Ferber, R. (1993). Freie Assoziationen und Kontiguitäten von Wörtern in Texten. *Zeitschrift für Psychologie 201*, 99-108.

Willett, P. (1985). An algorithm for the calculation of exact term discrimination values. *Information Processing and Management 21*(3), 225-232.