

# Proximity Relevance Model for Query Expansion

Liana Ermakova  
Institut de Recherche en  
Informatique de Toulouse  
118 route de Narbonne  
Toulouse, France  
ermakova@irit.fr

Josiane Mothe  
Institut de Recherche en  
Informatique de Toulouse  
118 route de Narbonne  
Toulouse, France  
mothe@irit.fr

Elena Nikitina  
Perm State National Research  
University  
15 Bukireva st.  
Perm, Russia  
neyu@psu.ru

## ABSTRACT

Query expansion (QE) aims at improving information retrieval effectiveness by enhancing the query formulation. Because users' queries are generally short and because of the language ambiguity, some information needs are difficult to satisfy. Query reformulation and QE methods have been developed to face this issue. Pseudo relevance feedback (PRF) considers the top retrieved documents as relevant and uses their content in order to expand the initial query. Rather than considering feedback documents as a bag of words, it is possible to exploit term proximity information. Although there are some researches in this direction, the majority of them is empirical. The lack of theoretical works in this area motivated us to introduce a novel method integrated into the language model formalism that takes advantage of the remoteness of candidate terms for QE from query terms within feedback documents. In contrast to previous works, our approach captures the proximity directly and in terms of sentences rather than tokens. We show that the method significantly improves the retrieval performance on TREC collections especially for difficult queries.

## Keywords

Information Retrieval, Query Expansion, Pseudo Relevance Feedback, Term Proximity, Language Model, Relevance Model

## 1. INTRODUCTION

Information retrieval (IR) aims at retrieving the relevant documents according to a user's need. Concretely, a search engine computes a similarity between the user's query and the indexed documents. In the keyword based IR techniques, the documents that contain the query terms are retrieved and ordered according to their decreasing similarity with the query. Because real queries are short and because natural language is ambiguous such matches can be wrong or

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*SAC'16, April 4-8, 2016, Pisa, Italy*

Copyright 2016 ACM 978-1-4503-3739-7/16/04...\$15.00

<http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxx>

incomplete. The matching problem rises from the fact that the terms used by the authors of documents and the search engine users to represent a concept may be different. Another reason is that users express their needs using just a few words making the query difficult to 'be understood' by the system. To face these challenges, IR systems consider several strategies. One of them is to diversify the results. On the other hand, Query Expansion (QE) has driven many works in IR (see Carpineto's survey on QE[6]). QE aims at adding new terms to the initial query that will improve retrieval. QE is based on some knowledge, either extracted from external resources (e.g. WordNet), term collection distribution, user's profile (e.g. topics of interest), or relevance feedback. Local analysis or local feedback methods rely on the hypothesis that relevant documents contain terms that could be useful to reformulate an enhanced query. In the majority of previous works local context is viewed as an entire document presented as a bag of words and the proximity of terms is not captured. Although some researchers exploit term proximity in QE [5, 15, 18, 20], their works are rather empirical. The lack of theoretical works in this area motivated us to introduce a novel method integrated into the language model formalism that takes advantage of the remoteness of candidate terms for QE from query terms within feedback documents. In this paper, we introduce the formal model of query expansion exploiting term proximity.

As in the approaches based on the term proximity, we hypothesize that the closer a term is to a query term, the better the QE term candidate is. However, unlike the positional relevance model [12] which is a formal approach for QE extending the relevance Language Model (LM) [10], we believe that the suitability of the expansion candidates depends on rather their nature and the nature of the query terms than position within a document (e.g. synonyms usually do not co-occur within a sentence unlike other semantically related words). In contrast to [12], the proximity is captured directly rather than by weighting the positions within PRF.

We also put forward a hypothesis that it is more appropriate to estimate the distance not in terms of tokens, but rather in terms of sentences. This is motivated by the following facts:

- In linguistics a sentence is viewed as a minimal set of words that in principle tells a complete thought;
- Within a sentence, often words could be reordered

without meaning shift (e.g. paraphrasing, transformation between passive and active voices);

- Synonyms and associations are usually considered as good expansion candidates. However, synonyms usually do not co-occur within a sentence unlike other semantically related words.

Thus, our approach differs from the previous works by capturing the proximity directly and in terms of sentences rather than tokens.

The remainder of the paper is organized as follows. Section 2 observes the related work. Section 3 details the novel formal approach for QE. Section 3 presents the experimental framework. Finally, Section 4 concludes this paper and draws up some future works.

## 2. RELATED WORK

Automatic methods for QE were firstly proposed by Maron and Kuhns in 1960. QE techniques are either based on the analysis of a document collection [6] or they imply dictionary-based or ontology-based methods [3]. The results of the methods that need external resources to be used (dictionary-based or ontology-based methods, methods that uses other sources besides the original collection such as FAQ texts in question-answering systems) can highly depend on these resources. The analysis of the document collection can be either global (corpus analysis for the purpose of word relationships detection) [6] or local. The local analysis is related to the local feedback.

The use of relevance information for QE was first suggested by Rocchio [16] who defines the Relevance Feedback (RF) principle. Users are supposed to judge some of the retrieved documents and this feedback information is used then either to re-weight query terms or to expand the query with the most important terms from relevant documents. To avoid users' judgment that can be difficult to collect and to make the process fully automatic, Buckley *et al.* [4] suggested to consider the first initially retrieved documents as relevant, i.e. pseudo-relevance feedback (PRF). Many studies have shown that this method is efficient in average; however, it can lower results for some queries [6, 2, 7]. For example, it is most probable that for poor performing queries query expansion is helpless since it will be based on the first retrieved documents that are probably non-relevant documents. It is thus important to know in advance if QE will be helpful or on the contrary if it will degrade the results. Selective query expansion aims at making this decision [8]. Lee *et al.* propose a *resampling method* using top-retrieved document clustering [11]. Rather than focusing on how to select the best documents to use in PRF, some approaches focus on how to select the best terms to expand the initial query. Selecting the most appropriate terms from the relevant -or considered as such- documents is indeed a challenge [5, 12].

Xu and Croft use a feature selection based on co-occurrence of terms, considering that the best terms are the ones that co-occur with as many query terms as possible within the top-ranked documents or document passages [20]. Distinctly from Xu and Croft's approach that considers the distance

between the candidates and the query terms as binary (i.e. terms either co-occur within a text passage or not), in this paper we hypothesize that the dependence of the probability to find good candidates for QE on the distance is more sophisticated and that it should be considered at the sentence level.

Other empirical studies have shown that the term proximity is effective for selecting expansion terms. Cao *et al.* suggested a term classification method based on SVM to predict the usefulness of expansion term candidates [5] based on the term distribution, co-occurrence with query terms, and the proximity from them. Miao *et al.* proposed an extension of the Rocchio's approach by introducing a concept of proximity-based term frequency that focuses on the proximity of terms rather than positional information unlike the positional relevance model [15]. They provide 3 approaches for estimate the proximity-based term frequency, namely (1) moving window; (2) kernel-based and (3) Hyperspace Analogue to Language (HAL) methods. The approach of Miao *et al.* is rather empirical and is an elaboration of the TF-IDF model. Unlike [5, 15], we propose a theoretical reasoning of our approach.

Some works take into account only ordered or unordered n-grams within the window of N-terms [14, 17] capturing the proximity in binary sense. Tao and Zhai [18] explore only the proximity of query terms resting upon the hypothesis that in relevant documents query terms should be closer to each other. In contrast to the cumulative proximity expansions retrieval model [19] that does not require any co-occurrence statistics, we combined proximity and co-occurrence statistics within the language model formalism.

There is relatively little studies of formal models using positional heuristics for QE. The only formal approach for QE we are aware of is the positional relevance model [12] which is an extension of the relevance LM [10]. In the positional relevance model query likelihood is estimated as the product of the probabilities of the query terms in the position within pseudo-relevant documents. However, in this approach the term proximity is captured indirectly by weighting the positions within PRF.

Thus, the main contribution of this work is that it provides a novel formal LM for QE that directly captures the term proximity rather than by weighting term positions, and the distance is computed at sentence level.

## 3. METHOD DESCRIPTION

### 3.1 Proximity Relevance Model

The proposed method aims at selecting the most appropriate expansion terms for QE from the top-retrieved documents. Our approach is grounded on the following hypotheses:

1. A candidate term can expand not only a query term, but also a combination of query terms.
2. The terms lying in the neighborhood of query terms are closer related to them than the remote ones, and are better candidates for QE; however this dependence is not binary but rather it should be described as a more complex function.

- Since a sentence is a minimal set of words that in principle tells a complete thought, the distance should be estimated in terms of sentences rather than in terms of tokens. The probability to find semantically related words is the same sentence is usually higher. However, this probability depends on the nature of relationship (i.e. synonyms, antonyms, meronyms, associations etc.).

One of the most efficient and robust relevance model used for QE is the relevance LM that determines the probability  $P(w|Q)$  of observing a word  $w$  in the documents relevant to a particular information need expressed by a query  $Q$  [10]:

$$P(w|Q) \propto \sum_{d \in D} P(w|d)P(d) \prod_{i=1}^m P(q_i|d) \quad (1)$$

where  $Q = q_1, q_2, \dots, q_m$  is a query,  $q_i$  is the  $i$ -th term in  $Q$ ,  $P(d)$  is a prior of a document  $d$ , and  $D$  is a document set. Often, document priors  $P(d)$  are assumed to be uniform and in this case they can be ignored since they do not affect ranking.

In the relevance LM the probabilities are computed over the top documents from PRF. By the definition of conditional probability and since  $P(Q)$  does not depend on  $w$ :

$$P(w|Q) = \frac{P(w, Q)}{P(Q)} \propto P(w, Q) \quad (2)$$

In contrast to the relevance LM, we assume that considering the distance between a candidate term and query terms may improve the quality of QE. We hypothesize that good QE candidates in the neighborhood of query terms. Usually, the closer a term is to a query term, the better candidate it is. However, it is not a case of synonyms. Therefore, we introduce the random variable  $dist$  that expresses the probability to find a candidate term at some sentence distance from the query terms  $Q$ . Since  $P(w, Q)$  may be viewed as marginal over the variable  $dist$ , the general proximity relevance model can be expressed as:

$$P(w|Q) \propto \sum_{dist=0}^{\infty} P(w, dist, Q) \quad (3)$$

We enriched the relevance model by integrating the query term combinations into it. Thus, a word  $w$  can extend a query term combination  $Q_i \in \Omega = 2^Q \setminus \emptyset$  where  $2^Q$  is the power set of all query terms meeting the condition that  $(\forall i, j | i \neq j) : Q_i \not\subseteq Q_j$ . Since the events  $Q_i$  are mutually exclusive,

$$P(w, dist, Q) = \sum_{Q_i \in \Omega} P(w, dist, Q_i) \quad (4)$$

Thus, formula (3) can be rewritten as:

$$P(w|Q) \propto \sum_{Q_i \in \Omega} \sum_{dist=0}^{\infty} P(w, dist, Q_i) \quad (5)$$

Applying the chain rule,  $P(w, dist, Q_i)$  can be decomposed as:

$$P(w, dist, Q_i) = P(Q_i)P(dist|Q_i)P(w|dist, Q_i) \quad (6)$$

where  $P(Q_i)$  is the probability of the query term combination  $Q_i$ ,  $P(dist|Q_i)$  is the probability to find any expansion term at distance  $dist$  from  $Q_i$ , and  $P(w|dist, Q_i)$  is the probability to find the term  $w$  at distance  $dist$  from  $Q_i$ .  $P(dist|Q_i)$  may be viewed as a likelihood to see an expansion term at a specified distance depending on the nature of a query term combination  $Q_i$ .  $P(w|dist, Q_i)$  shows a likelihood to meet a specific term depending on the remoteness of a given  $Q_i$  i.e. it potentially captures the nature of the expansion candidate and its relationship with the query term (synonymy, meronymy, function etc.).

Substituting  $P(w, dist, Q_i)$  in (5) by (6), we obtain the final formula to estimate expansion candidate scores:

$$P(w|Q) \propto \sum_{Q_i \in \Omega} \sum_{dist=0}^{\infty} P(Q_i)P(dist|Q_i)P(w|dist, Q_i) \quad (7)$$

### 3.2 Estimation Details

The probability of a term combination  $Q_i = q_1, q_2, \dots, q_m$  is usually calculated as follows:

$$P(Q_i) = \prod_{j=1}^m P(q_j) \quad (8)$$

To avoid underflow, the probability is replaced by its logarithm [9]:

$$P(Q_i) \propto \sum_{j=1}^m \log(P(q_j) + 1) \quad (9)$$

Assuming that the probability to find any expansion term at distance  $dist$  from  $Q_i$  does not depend on  $Q_i$  we can simplify the calculation of  $P(dist|Q_i)$  by reducing it to  $P(dist)$ . The dependence of the nature of the query terms  $Q_i$  is the perspective of this paper.

The distributions of many quantities follow the power law, at least in their upper tail, especially in natural languages (e.g. Zipf's law). Although it is not exactly known why the power law holds for most languages, the explanation may be statistical or related to the principle of least effort, i.e. interlocutors do not want to work any harder than necessary to reach understanding. We hypothesize that the principle of least effort holds also for topic development within a text. Thus, a topic within a text is expanded in the neighboring context and we assume that the distribution of the words used for it follows the power law. Thereby, the probability to find an expansion candidate for a topic expressed by query terms should also fit the power law.

The probability of  $P(w|dist, Q_i)$  is estimated as the frequency of observing the term  $w$  at distance  $dist$  from  $Q_i$ :

$$P(w|dist, Q_i) \approx \frac{count(w|dist, Q_i)}{\sum_{k=1}^{|W|} count(w_k|dist, Q_i)} \quad (10)$$

In this paper the distance means the remoteness from the closest query term or their combination  $Q_i$ . Since we compute the distance in terms of sentences and the combinations of the query terms are considered only within a sentence, the remoteness does not depend on the length of the query term combination.

The set of the query term combinations  $Q_i \in \Omega = 2^Q \setminus \emptyset | (\forall i, j | i \neq j) : Q_i \not\subseteq Q_j$  does not lead to the exponential complexity of the algorithm since we consider only query term combinations within a sentence and we ignore embedded combinations. Thus, the computation of the query term combinations has a linear time over the number of tokens in the PRF.

Smoothing the probability  $P(w|dist, Q_i)$  by the collection probability of the candidate term  $P_c(w)$  gives:

$$P_s(w|dist, Q_i) = \lambda P(w|dist, Q_i) + (1 - \lambda)P_c(w) \quad (11)$$

where  $P_s(w|dist, Q_i)$  is a smoothed probability and  $\lambda$  is a smoothing parameter.

Dividing the equation by  $(1 - \lambda)P_c(w)$  we obtain the final ranking score of the expansion candidate terms:

$$score(w) = \frac{\lambda P(w|dist, Q_i)}{(1 - \lambda)P_c(w)} + 1 \quad (12)$$

## 4. EXPERIMENTS

### 4.1 Experimental Framework

The evaluation was performed on two TREC datasets:

- Robust TREC;
- WT10G.

Robust TREC set is a "pure" collection since the documents have almost the same format and there is no spam. In contrast, WT10G is a snapshot of the web with real documents in HTML format, some of which are spam. Robust TREC set consists of 249 topics, TREC Disk4&5 (except CR data) and relevance judgments. WT10G is 10GB subset of the web snapshot and contains more than 1.6 million of documents. There are 98 topics with relevance judgments.

The system performance was evaluated by several measures implemented in `trec_eval`<sup>1</sup> software provided by the TREC community for evaluating an ad hoc retrieval run, given the results file and a standard set of judged results. In this study, we report the following measures:

- Mean Average Precision (MAP) over all queries;
- Binary preference (BPREF);
- Normalized discounted cumulative gain (NDCG).

MAP may be viewed as one of the main measures since it has very good discrimination and stability [13]. Binary preference computes a preference of whether judged relevant documents have higher rank than judged non-relevant documents. Thus, BPREF does not treat non-assessed documents as non-relevant while MAP does. This is important for large collections where the probability of retrieving non-assessed documents is higher. NDCG is suitable for non-binary judgments.

For comparison purpose we used several PRF methods, namely DRF (Divergence from Randomness) models implemented

<sup>1</sup>[http://trec.nist.gov/trec\\_eval/](http://trec.nist.gov/trec_eval/)

in Terrier platform<sup>2</sup>, a state-of-the-art open source search engine developed at the University of Glasgow. The DFR models are based on the assumption that informative words are relatively more frequent in relevant documents than in others [1]. During QE the best-scored terms from the top-ranked documents are extracted. Terms are ranked using one of the DFR weighting model. We compare our system (PRM\_SNT) with the following DFR models:

- baseline presented by InL2c1.0 model without any query expansion which is the default model in Terrier and based on  $TF - IDF$  measure with  $L2$  term frequency normalization (InL2c1.0);
- Kullback-Leibler divergence model (KL);
- Chi-square divergence model (CS);
- Bose-Einstein 1 model (Bo1);
- Bose-Einstein 2 model (Bo2).

All these systems used InL2c1.0 model for PRF.

Moreover, we compared our method with RM3 model implemented in Indri, a search engine from the Lemur project mainly built on the language modeling information retrieval<sup>3</sup>. RM3 is an Indri's adaptation of Lavrenko and Croft's relevance models [10]. RM3 is a well-known relatively strong baseline.

### 4.2 Details of the Implemented System

Our approach requires PRF. In order to obtain preliminary ranking we used Terrier with the following parameters: words are stemmed using Porter's algorithm, as a retrieval model we applied *InL2c1.0*. *InL2* demonstrates better performance at many recall levels and in average precision than traditional retrieval models such as BM25 [1].  $L2$  normalization is less sensitive to document length. The sentence chunking was performed by Stanford CoreNLP<sup>4</sup>.

In our experiments we assumed that the probability  $P(dist)$  of a candidate to occur at the given distance follows the power law. We set the limit of distance  $MaxDist = 9$  sentences and thus we calculated the  $P(dist)$  as:

$$P(dist) = \begin{cases} \frac{1}{(MaxDist+2)^{0.5}} & \text{if } dist > MaxDist \\ \frac{1}{(dist+1)^{0.5}} & \text{if } dist \leq MaxDist \end{cases} \quad (13)$$

The smoothing parameter  $\lambda$  was set to 0.3. This parameter will be learnt and optimized in future work.

In order to test the hypothesis that it is preferable to estimate the distance in terms of sentences rather than tokens, we compared our approach with the same method in which the distance was calculated at word level (PRM\_W). The *MaxDist* parameter was also set to be 9 sentences, the es-

<sup>2</sup>[terrier.org/](http://terrier.org/)

<sup>3</sup><http://www.lemurproject.org/>

<sup>4</sup>[nlp.stanford.edu/software/corenlp.shtml](http://nlp.stanford.edu/software/corenlp.shtml)

timation of the probability  $P(dist)$  was slightly different:

$$P(dist, wdist) = \begin{cases} \frac{1}{((MaxDist+1) \times avgSntLen+1)^{0.5}} & \text{if } dist > MaxDist \\ \frac{1}{(wdist+1)^{0.5}} & \text{if } dist \leq MaxDist \end{cases} \quad (14)$$

where  $wdist$  is a word distance,  $avgSntLen$  is an average sentence length.

### 4.3 Results

Table 1 reports the results obtained for the Robust and WT10G data sets. All systems were based on PRF of 5 documents from which 10 best scored terms were extracted. Our system based on sentence-level distance (PRM\_SNT) demonstrated the best results according to all metrics for both data collections. We performed the Student’s t-test to verify the statistical significance of the difference of the results obtained by our method and the baseline (this test is applicable since the performance results fit a normal distribution according to the Pearson’s  $\chi^2$ -test). We also compared our results with those of the best approach implemented in Terrier, namely Bo1 (although KL is slightly better on the Robust data set, it has much lower results on WT10G). The differences with the baseline, Bo1 and RM3 marked by \*, + and # respectively are significant at the level  $p < 0.05$ .

Considering the Robust collection, in comparison with Bo1 our method PRM\_SNT showed better results for 128 queries and lower performance for 119 queries. Our method outperformed Bo1 for 80 (63% of all improved results) difficult queries ( $MAP(baseline) < 0.25$ ) and for 43 (34%) very difficult queries ( $MAP(baseline) < 0.1$ ). Among ameliorated results 10 (8%) of queries were simple ( $MAP(baseline) > 0.5$ ). Thus, we can conclude that PRM\_SNT is better than the state-of-the-art QE model even in case of difficult queries. The degradation of results in comparison with the baseline was observed in 79 cases. Among the latter for 88% of queries the degradation of the results relative to the baseline without QE was observed for all QE methods; this feature leads us to conclude that either these queries should not be expanded or the methods based on co-occurrence are not suitable. PRM\_W showed worse results than PRM\_SNT and KL according to *BPREF* but outperformed other DFR models. PRM\_W is much better than the baseline and it is comparable with the DFR QE models according to other metrics. On Robust data set RM3 performed worse than the DFR QE models. Both our systems significantly exceeded RM3 by all metrics.

For the WT10G data set our system was better than Bo1 for 53 queries. PRM\_SNT was worse than the baseline for 32 queries and among them for 27 queries (84%) any applied QE method worsened the results. It lowers performance compared to Bo1 for 41 queries. This allows to draw a conclusion that our method may be significantly improved by selective QE since it has lower results than the DFR models mainly for queries that should not be expanded at all. Word-based PRM surpassed all DFR models but it was inferior to PRM\_SNT. In case of WT10G, RM3 was comparable

with DFR QE approaches but remained significantly lower than the method proposed in this paper.

For both test collections, for 89% of queries improved by Bo1, our system outperformed the DFR models.

## 5. CONCLUSION

QE is a powerful technique in IR, though the terms that are added can bias a query and thus decrease both recall and precision. In this paper we proposed a novel approach for query expansion by incorporating term proximity information into the LM formalism. The method is based on PRF, but it differs from previous researches in several ways:

- it is formalized within LM;
- the term proximity is captured directly, and not by weighting term positions;
- the distance is computed in terms of sentences from the query terms and its combinations.

We evaluated our method on TREC Robust collection as well as on WT10G. Our system demonstrated the best results among the state-of-the-art QE method implemented in such search engines as Terrier and Lemur according to all metrics for both data collections. Experiment results showed that the proposed method is significantly better than other PRF-based QE approaches (DFR QE models, RM3) as well as the baseline. Major improvement was observed for difficult and very difficult queries. Our method has lower results than the DFR models mainly for queries that should not be expanded. Therefore, the proposed approach may be significantly improved by selective QE.

For both test collections, our method grounded on sentence-level distance estimation outperformed the word-based one. This fact allows concluding that distance measuring in terms of sentences is more preferable than in terms of tokens.

One of the most promising directions of further research is to differentiate the probability distribution of the distances depending on query terms.

## 6. ACKNOWLEDGMENTS

The authors would like to thank the French Embassy in Russia for funding this project (bourse de thèse en cotutelle).

## 7. REFERENCES

- [1] G. Amati. *Probability Models for Information Retrieval Based on Divergence from Randomness: PhD Thesis*. University of Glasgow, 2003.
- [2] G. Amati, C. Carpineto, and G. Romano. Query difficulty, robustness, and selective application of query expansion. *Advances in Information Retrieval*, page 127–137, 2004.
- [3] J. Bhogal, A. Macfarlane, and P. Smith. A review of ontology based query expansion. *Inf. Process. Manage.*, 43(4):866–886, July 2007.

**Table 1: Results**

		InL2c1.0	KL	CS	Bo1	Bo2	RM3	PRMLW	PRMLSNT
Robust	MAP	0.2407	0.2829	0.263	0.2822	0.2801	0.2602	0.2795	<b>0.2884</b> *+ #
	BPREF	0.2506	0.2807	0.2635	0.28	0.2782	0.2585	0.28	<b>0.2863</b> *+ #
	NDCG	0.5124	0.5566	0.5329	0.5561	0.5549	0.5268	0.5549	<b>0.5614</b> *+ #
WT10G	MAP	0.1894	0.2121	0.2013	0.2179	0.2174	0.2142	0.2247	<b>0.2345</b> *+ #
	BPREF	0.1895	0.2016	0.2009	0.2107	0.2071	0.2084	0.2153	<b>0.2275</b> *+ #
	NDCG	0.4624	0.4888	0.4497	0.4927	0.4885	0.4699	0.5005	<b>0.5076</b> *+ #

- [4] C. Buckley. Automatic query expansion using SMART : TREC 3. In *Proceedings of The third Text REtrieval Conference (TREC-3)*. NIST Special Publication 500-226, page 69–80. National Institute of Standards and Technology, 1995.
- [5] G. Cao, J.-Y. Nie, J. Gao, and S. Robertson. Selecting good expansion terms for pseudo-relevance feedback. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 243–250. ACM, 2008.
- [6] C. Carpineto and G. Romano. A survey of automatic query expansion in information retrieval. *ACM Computing Surveys*, 44(1):1–50, Jan. 2012.
- [7] C. Chen, H. Chunyan, and Y. Xiaojie. Relevance feedback fusion via query expansion. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 03*, WI-IAT '12, pages 122–126, Washington, DC, USA, 2012. IEEE Computer Society.
- [8] S. Cronen-Townsend and W. B. Croft. Quantifying query ambiguity. page 104–109, Mar. 2002.
- [9] D. Hiemstra. Language models. In T. Özsu and L. Liu, editors, *Encyclopedia of Database Systems*, pages 1591–1594. Springer Verlag, Berlin, 2009.
- [10] V. Lavrenko and W. B. Croft. Relevance based language models. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 120–127. ACM, 2001.
- [11] K. S. Lee, W. B. Croft, and J. Allan. A cluster-based resampling method for pseudo-relevance feedback. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 235–242. ACM, 2008.
- [12] Y. Lv and C. Zhai. Positional relevance model for pseudo-relevance feedback. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, page 579–586. ACM, 2010.
- [13] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [14] D. Metzler and W. B. Croft. Latent concept expansion using markov random fields. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 311–318. ACM, 2007.
- [15] J. Miao, J. X. Huang, and Z. Ye. Proximity-based rocchio’s model for pseudo relevance. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 535–544. ACM, 2012.
- [16] J. Rocchio. Relevance feedback in information retrieval. In *The SMART Retrieval System*, page 313–323. 1971.
- [17] F. Song and W. B. Croft. A general language model for information retrieval. In *Proceedings of the 1999 ACM SIGIR Conference on research and development in Information Retrieval*, pages 279–280. ACM, 1999.
- [18] T. Tao and C. Zhai. An exploration of proximity measures in information retrieval. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 295–302. ACM, 2007.
- [19] J. Vuurens and A. de Vries. Distance matters! cumulative proximity expansions for ranking documents. *Information Retrieval*, 17(4):380–406, 2014.
- [20] J. Xu and W. B. Croft. Query expansion using local and global document analysis. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 4–11. ACM, 1996.