Integrating multiple windows and document features for expert finding

Jianhan Zhu[1], Dawei Song, Stefan Rüger

Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom, MK7 6AA

Telephone: +44 (0)1908 652073 Fax: +44 (0)1908 653169

E-mail: {j.zhu, d.song, s.rueger} @open.ac.uk

ABSTRACT

Expert finding is a key task in enterprise search and has recently attracted lots of attention from both research and industry communities. The Text REtrieval Conference[2] (TREC) has organized an expert search task for 2005, 2006, and 2007. Given a search topic, a prominent existing approach is to apply some information retrieval (IR) system to retrieve top ranking documents, which will then be used to derive associations between experts and the search topic based on co-occurrences. However, we argue that expert finding is more sensitive to multiple document features that current expert finding systems insufficiently address, including: (1) multiple levels of associations between experts and search topics, (2) document internal structure and (3) document authority. We propose a novel approach which integrates the above three aspects as well as a query expansion technique in a two-stage model for expert finding. A systematic evaluation is conducted on the TREC2006 and TREC2005 expert search collections to test the performance of our approach and the effects of different aspects of document features and query expansion. These experimental results show that query expansion can dramatically improve expert finding performance. Both document internal structure and our novel multiple window based approach for taking into account multiple levels of associations improve expert finding over the direct use of a number of well-known IR models for both with and without query expansion.

Keywords

Expert finding, enterprise search

---

[1] Correspondence author.

[2] http://trec.nist.gov

1. INTRODUCTION

Expert finding is a key task in enterprise search and has recently attracted lots of attention. A typical user scenario is that one needs to learn about a subject and wants to talk to someone who knows about it as the first step. Another use case is that a project manager is trying to assemble a project team. Accordingly, Yimam-seid and Kobsa (2003) identified two main motives for expert finding, namely, as a source of information to answer the question “who knows about topic x?” and as someone who can perform a given organizational or social duty to answer the questions such as “how well does y know about topic x?”, “what else does y know?”, “how does y compare with others on topic x?” etc. They argued that manually developed expertise databases are labor intensive and often quickly out of date. On the other hand, much valuable and up-to-date expertise information often exists implicitly or explicitly in documents produced within the organization, e.g., emails, blogs, and web pages of individuals or groups, etc. Automating expert finding from these documents will provide a much cheaper way to gather useful and up-to-date expertise information.

The TREC enterprise track (Bailey et al. 2007; Craswell et al. 2006; Soboroff et al. 2007) has been the major forum for empirically comparing expertise modeling techniques. Since 2005, tremendous progress has been made in terms of expertise modeling, algorithms, and evaluation strategies. The goal of expert finding is to identify a list of people who are knowledgeable about a given topic. In contrary to traditional IR systems, the target of expert finding is people (named entity) instead of documents. This task is usually addressed by uncovering associations between people and topics (Craswell et al. 2006); commonly, co-occurrences of a person’s name with topic terms in the same context are assumed to be evidence of expertise. Essentially, the two most popular and well-performing types of approaches in TREC expert search task are profile-centric and document-centric approaches (Bailey et al. 2007; Craswell et al. 2006; Soboroff et al. 2007).

A prominent language modeling approach has been proposed by Balog et al. (2006). They distinguish between “Model 1”, which directly represents the knowledge of an expert from associated documents, and “Model 2”, which first locates documents on the topic and then finds the associated experts. Petkova and Croft [18] have further improved their models by proposing a proximity-based document representation for incorporating sequential information in text. Serdyukov and Hiemstra [19] propose a novel expert-centric language model for expert search.

However, all these language modeling approaches have not sufficiently considered the effect of document features in expert finding. As rich document features exist in an organizational intranet environment and are shown to be effective for document retrieval [8], it is timely to study the effect of document features in expert finding. We discuss the following document features that expert finding is potentially sensitive to.

1. Document internal structure. A document's internal structure can often be crucial in determining whether a person mentioned in the document is an expert on a topic that is also mentioned. For example, in a technical paper, the occurrence of a person's name in the author, content, reference, or acknowledgement section of the paper has different implications of the person's expertise on a topic. In a co-occurrence model, we can give different weights to text windows inside different sections of a document, e.g., give higher weight to text windows in the author section of a technical paper.

2. Multiple levels of associations in documents. In a co-occurrence model, the distance between occurrences of an expert and topic terms is a strong indicator of the expert's relevance to the topic. In traditional window-based association methods, a text window is set to measure the co-occurrences of the expert and query terms. Once the window size is set, it is fixed. However, in expert finding, there are associations between an expert and query terms on multiple levels, i.e., from phrase, sentence, paragraph, etc., up to document levels. All these levels of associations need to be considered in the co-occurrence model. In selecting window sizes, small window sizes often lead to high precision but low recall in finding experts, while large window sizes lead to high recall but low precision. Increased window sizes often lead to more coverage of associations while introducing noise. We can generally give higher weights to smaller windows than larger windows. In this paper, we propose to adopt a novel weighted multiple-sized-window based approach in the association discovery model.

3. Document authority. Some documents are more authoritative than the others in identifying people's expertise on a topic. Thus giving higher weights to these authorities than the other ordinary documents can potentially improve expert finding. We hypothesize that these authorities typically are linked more often by other documents and authoritative documents. We have used the PageRanks (Page et al. 1998; Brin & Page 1998) of documents to measure their authority.

In this paper, we propose to consider the above three aspects for more effective expert finding. Our approach is novel in the following aspects:

Firstly, to the best of our knowledge, this is the first attempt to use a weighted multiple-window approach in an information retrieval model for association discovery. We will carry out a systematic investigation of the effects of window combination in expert finding.

Secondly, we propose a novel approach which integrates multiple document features and query expansion in a unified way. We will study how different document features, and query expansion in combination with different IR models affect the expert finding performance.

Thirdly, although the effects of different retrieval models in document based retrieval have been extensively studied, little has been studied about the effects of these retrieval models in expert finding. In this paper, we will conduct a systematic investigation into the effects of different IR models in expert finding

Fourthly, we propose a novel query expansion technique for expert finding based on a well-known cognitive model called HAL (Hyperspace to Analogue Language) (Burgess et al. 1998). Query expansion is integrated with multiple document features in our unified expert finding approach.

Our experiments on the TREC W3C dataset show that our query expansion technique can dramatically improve expert finding performance, and the incorporation of document internal structure and multiple levels of associations out-performs the direct use of a number of well-regarded IR models combined with fix-sized-window-based association discovery. It is also worthy noting that our approach produces a better performance than our own runs, which are also the best performing runs, in the TREC2006 expert search task.

The remainder of the paper is organized as follows. In Section 2, we review the related work. We then discuss query expansion in Section 3. Our novel two stage model integrating three document features is presented in Section 4. The experimental results are reported in Section 5. Finally, we conclude and propose future work in Section 6.

2. RELATED WORK

Previous work related to our expert finding can be divided into the following four categories.

2.1 Corpus statistics based approaches

Given a search query, the Expert Finder (Maybury et al. 2001) works based on evidence such as frequency of documents published by an expert on the topic, contents of resumes, and co-occurrence of the expert and query terms in documents. The XperNet (Maybury et al. 2001) clusters experts with similar skills to form expert networks. Conrad and Utt (1994) used corpus-wise mutual information and phi-squared measures to discover associations between named entities. Although co-occurrences or corpus statistics carry useful information about inter-entity associations, they can suffer from topic-independence, i.e., insensitivity to document relevance to the search topic.

2.2 Link-analysis-based approaches

Campbell et al. (2003) used email content to find related emails to a given topic, from which they constructed a graph consisting of email senders and receivers. They applied the HITS (Hyperlink-Induced Topic Search) algorithm (Kleinberg 1998) to the graph in order to identify experts with high authority in an organization. However, this approach is limited only to datasets with explicit linkage information. Kolla and Vechtomova (2007) only used the email part of the TREC 2006 dataset for expert finding. Similar to Campbell et al. (2003)'s approach, they constructed a graph based on the email sender and receiver relations. However, their experimental results in terms of MAP and other performance measures based on outdegrees of experts and the HITS algorithm, respectively, are significantly lower than those of the other participants' systems which used the other parts of the TREC2006 dataset (Soboroff et al. 2007). Chen et al. (2007)'s work in TREC 2006 also shows that a two-stage model based on the whole dataset significantly outperforms the PageRank and HITS algorithm based on the email part of the whole dataset, respectively. We think the reason is that in an organizational environment, expertise information may be contained in documents of various format, the email dataset alone may not cover all expertise information for the domain. However, the above link analysis based approach can be integrated with other expert finding methods that exploit data other than emails.

2.3 Lexical-pattern-based approaches

Etzioni et al. (2004) used lexical patterns to discover relationships between terms from web documents. (Ciravegna 2001; Craven et al. 2000) used machine learning techniques to learn patterns from documents. Nenadic & Ananiadou (2006) proposed a hybrid method for identifying semantically related entities from biomedical literature based on lexical, syntactic, and contextual similarities between these entities. The advantage of these approaches is that it is possible to give a semantic interpretation to the relation which has been discovered between two entities. However a major disadvantage is that entity associations, for example, associations between people and areas of expertise, are often only expressed implicitly. For instance, if we search in Google for the terms "Tim Berners-Lee" and "Semantic Web" we find tens of thousands of pages where both terms appear, denoting a strong correlation between these two terms and suggesting the likelihood that Tim Berners-Lee is an expert in this topic. However, this does not necessarily mean that anybody has actually explicitly stated that Tim Berners-Lee is an expert on the semantic web in a form that is amenable to an approach based on lexical patterns. In other words these associations can often be implicit and can only be derived by statistical means. Furthermore many lexical patterns used in the aforementioned approaches are confined to a specific domain and therefore have limited applicability. On the other hand, the two-stage model in expert search is generic to any domain.

2.4 Information-retrieval-based approaches

The two most popular and well-performing types of approaches in TREC expert search task are basically profile-centric and document-centric approaches (Bailey et al. 2007; Craswell et al. 2006; Soboroff et al. 2007).

Profile centric approaches build the profile of an expert as a pseudo document by aggregating text segments relevant to the expert, e.g., context text windows of the expert in documents (Fu et al. 2006a; 2006b). Traditional document based retrieval models can be directly used for indexing and searching profiles of experts. The advantages of the profile centric approaches are that profiles can be significantly smaller than the original corpus, making retrieval of experts efficient, and these approaches can be integrated with expert profiling approaches (Balog & de Rijke 2007b).

Document-centric approaches are typically based on traditional document retrieval techniques, and can be generalized as a two-stage model. Firstly, in a *document relevance model*, we estimate the conditional probability $p(q|d)$, of the query topic q given a document d. Secondly, in an *association discovery model*, based on the assumption that terms co-occurring with an expert in the same context describe the expert, $p(q|d)$ is used to weight the evidence of co-occurrence of experts with terms in q in documents. The conditional probability $p(c|q)$ of an expert candidate c given a query q can be estimated by aggregating all the evidences in all the documents where c and terms in q co-occur. Query expansion techniques can be integrated with the document-centric approaches by firstly expanding the original query q for the expanded query q_e, secondly associating experts with terms in q_e, and finally weighting these co-occurrences by $P(c|q_e)$.

Document-centric approaches normally outperform profile-centric approaches (Soboroff et al. 2007) as the latter achieve efficiency at the expense of useful information in terms of internal document structure and high-level language features (Petkova & Croft 2006). Balog et al. (2006)'s work also shows that the document centric model outperforms the candidate centric model on the TREC dataset.

In contrast to the models by Balog et al. (2006), Petkowa and Croft (2007) and Serdyukov and Hiemstra (2008), which were discussed in the introduction, Cao et al. (2006) proposed a two-stage language model combining a document relevance and co-occurrence model. Fang et al (2007) derived a generative probabilistic model from the probabilistic ranking principle and extended it with query expansion and non-uniform candidate priors. We first proposed a novel multiple window based approach for integrating multiple levels of associations between experts and query topic in expert finding (Zhu et al. 2007).

A number of query expansion techniques are applied to expert finding (Balog et al. 2007a; Macdonald & Ounis 2007a; Petkova & Croft 2006). Information fusion techniques have also been applied to expert finding. Chu-Carroll et al. (2007) effectively used multiple agents for expert finding, and Maconald and Ounis (2007) presented a Bayesian belief network model for taking into account various types of evidence in expert finding.

In addition to the use of language models as the document relevance model in expert finding, other models such as the BM25 (Robertson et al. 1995), DFR (Divergence From Randomness) (Amati & van Rijsbergen 2002), and TF/IDF (Salton et al. 1983;

Salton & Buckley 1988) models etc. have also been used (Hu et al. 2006; Macdonald & Ounis 2006; Zhao & Lu 2007; Yao et al. 2006; Zhu et al. 2007). Little has been studied about the effects of different document relevance models in expert finding, which is a research problem we will address in this paper.

Expert finding can be generalized to retrieval of entities of other types in documents. The introduction of Entity Ranking Track in INEX 2007 on the Wikipedia dataset provides a platform for entity search evaluation (de Vries et al. 2007). Cheng et al. (2007) proposed an EntityRank algorithm integrating local co-occurrence and global access information for entity search into a probabilistic estimation of entity and query association, which is quite similar to the above two-stage expert finding approaches.

3. QUERY EXPANSION

A TREC expert finding topic looks like the follows:

```
<top>
<title>relationship cardinalities</title>
<description>A relevant expert will have knowledge in relationship cardinalities between roles in different
choreographies.</description>
<narrative>In the context of semantic web, the relationships between entities can have different cardinalities and roles.
The relevant expert will have an explicit knowledge of such choreographies. Experts in Semantic Web are not relevant
without explicit knowledge in choreographies.</narrative>
</top>
```

However, ordinary search engines users tend to employ short queries regardless of their target task (informational, navigational, or transactional) (Silverstein et al. 1999). Since an expert finding task is in some sense both informational (who knows about something) and navigational (where/how the searcher could find the experts), we have used only the title part, which normally consists of two to four terms, of these TREC expert finding topics to emulate real world users' expert finding queries.

As query expansion techniques have been successfully used in expert finding (Balog et al. 2007a; Macdonald & Ounis 2007a; Petkova & Croft 2006), we will also explore the effects of an automatic query expansion technique, namely, the HAL (Hyperspace to Analogue Language) based information flow model (Song & Bruza 2003), in expert finding. We chose this model is due to the reasons that HAL is cognitively compatible with human processing of text (Burgess et al. 1998), can effectively improve document retrieval on large scale datasets (Song & Bruza 2003), provides weights of terms that can be used in expert finding, and is able to weight terms by taking into account other sources of knowledge, such as weighting terms appearing in the description and narrative part of a TREC topic higher in query expansion.

We employed implicit relevance feedback in HAL based query expansion. We took the top 30 relevant documents returned by a document relevance model and removed HTML markups and stopwords from these 30 documents. A HAL semantic space (Burgess et al. 1998) is automatically constructed by moving a text window over these 30 documents by one term increment ignoring punctuation, sentence, and paragraph boundaries. All words within the window are considered co-occurring with each other with strengths inversely proportional to the distance between them. After traversing the documents, an accumulated co-occurrence matrix for all words in these documents is produced. An example of a normalized HAL vector for "cardinalities" is:

Cardinalities = < choreographies:0.13 relationship:0.12 semantic:0.11 web:0.11 services:0.09 roles:0.09 w3c:0.07….>

Based on the heuristic concept combination on the HAL space (Song & Bruza 2003), we obtain the combined HAL vector for the query title, e.g., "relationship cardinalities". After normalization, an example of a combined HAL vector for "relationship cardinalities" is:

Relationship $\oplus$ cardinalities = < choreographies:0.16 roles:0.14 semantic:0.09 services:0.08 web:0.05 w3c:0.02 ….>

We can observe how the weights of some dimensions have changed appropriately with respect to associations relevant to "cardinalities" in the context of "relationship". Weights for "choreographies" and "roles" increases, while weights of dimensions dealing with "cardinalities" in the general context, e.g., "semantic", "services", and "web", decrease.

We adapt the HAL space model to consider description and narrative parts of a topic by weighting dimensions of the combined vector, which appear in the description or narrative parts of the topic, higher. For example, the weights of the dimensions of the above combined vector appearing in the description or narrative parts of the topic are multiplied with a factor of 1.5 as:

Relationship $\oplus$ cardinalities = < choreographies:0.24 roles:0.21 semantic:0.135 services:0.08 web:0.075 w3c:0.02 ….>

Top 10 terms with the highest weights in the combined HAL vector of the query are used for query expansion, e.g., "relationship cardinalities" is expanded to "relationship cardinalities choreographies roles semantics services web …".

In the co-occurrence model for expert finding, we sometimes cannot expect that all the terms in the expanded query co-occur with a candidate in a text window of a document. Therefore, we may treat this as a query subset matching problem (Charikar et al. 2002). Given each text window, if a candidate co-occurs with a term in the query, we recursively add another term in the query and check whether the new query term also occurs in the text window. In this way, we will find the maximum subset of the query that co-occurs with the candidate in the text window. We assume that it is more likely for a candidate to be an expert when he/she co-occurs with more highly weighted terms in the query in text windows. Based on this assumption, we weight the evidence for the co-occurrence of a candidate and a query term in a text window by the weight of the query term, where we specify the weight for each original query term as 1.0, and the weight for an expanded query term as the term's weight in the combined HAL vector of the original query.

4. INTEGRATING DOCUMENT FEATURES

In this section, we present our novel two stage model integrating multiple document features. Firstly, we introduce the basic two stage model. Secondly, we propose our model for ranking supporting documents of each candidate. Thirdly, we discuss the document relevance model. Fourthly, we integrate document authority with document relevance. Fifthly, we propose a novel multiple window approach for taking into account multiple levels of associations in the co-occurrence model. Finally, the effect of document structure and window size in the co-occurrence model is analyzed.

4.1 Two stage model

Our models are instances of document-centric generative language modeling approaches to rank experts. The two-stage model consists of a document relevance model and a co-occurrence model. Formally, given a set of documents, *D*, a query *q*, and a set of candidates, *C*, we state the expert finding problem as "what is the probability of a candidate *c* in *C* being an expert given a query *q*?". Since *c* may co-occur with some but not all of the terms in *q* in a text window *w*, we give credit to the co-occurrence based on the importance of these terms in *q*. Suppose that *q* consists of a number of unique terms as $\{t_i\}$, $i \in (0, N)$ and *N* is the length of *q*. We assume that t_1, t_2 , …, t_N are independent of each other, and get:

$$P(c \mid q) = \frac{P(c, t_1, ..., t_N)}{P(q)} = \frac{P(c,t_1)P(c,t_2)...P(c,t_N)}{P(q)} = \frac{\prod_i P(t_i)P(c \mid t_i)}{P(q)},$$

Since $P(q)$ does not affect the ranking of candidates, the problem reduces to the estimation of $P(t_i)$ and $P(c| t_i)$. $P(t_i)$ is the prior probability of t_i. We estimate $P(t_i)$ as the weight of t_i in the combined HAL vector of the original query.

Therefore, the two-stage model is as follows:

$$P(c \mid t_i) = \sum_d P(c, d \mid t_i) = \sum_d P(d \mid t_i) P(c \mid d, t_i) \quad (1)$$

where *d* is a document, $P(d|t_i)$ is the document relevance model, and $P(c|d,t_i)$ is the co-occurrence model.

4.2 Supporting document model

Like in the TREC expert search task, we consider both retrieval with supporting documents and retrieval without supporting documents. For each expert, a number of supporting documents are retrieved. Experts without supporting documents are useful to searchers, however, as a consequence, the suggestion is not well founded.

When do not consider supporting documents, if a system retrieves a candidate that is judged to be an expert, the system will receive credit regardless of whether or not any supporting documents were retrieved. On the other hand, when consider supporting documents, if a system's retrieved supporting documents are judged to contain positive support for a true expert, the system will receive credit for retrieving it. However, if no positive supporting documents are retrieved for a candidate, the candidate was considered irrelevant.

Performance is measured by mean average precision (MAP)[3], which rewards systems that retrieve relevant experts highly ranked. We expect that the MAP of an expert finding system drops when the relevance is judged by considering supporting documents, compared with the case without considering supporting documents. This makes sense because human evaluators often take many factors into account in relevance judgments and their standards can often differ from what an IR relevance model judges a document as containing supporting evidence.

A document's support is estimated as the conditional probability $P(d|c,q)$ as

$$P(d \mid c,q) = \frac{P(d,c,q)}{P(c,q)} = \frac{\prod_i P(d,c,t_i)}{P(c,q)} = \frac{\prod_i P(c \mid d,t_i)P(d \mid t_i)P(t_i)}{P(c,q)} \tag{2}$$

4.3 Document relevance model

$P(d|t_i)$ consists of two parts, a content-based relevance score, $P_{content}(d|t_i)$, and a query-independent-page-authority-based score, where we have used the PageRank (Page et al. 1998; Brin & Page 1998) based score, i.e., $P_{PageRank}(d)$. Based on how Craswell et al. (2005) combined BM25 and PageRank scores, we estimate $P(d|t_i)$ by combining the two parts as follows:

$$P(d|t_i) \propto P_{content}(d|t_i) + P_{PageRank}(d)$$

We have experimented with different document relevance models in calculating the content-based relevance score. Language model (Metzler et al. 2004), a probabilistic IR model called BM25 (Robertson et al. 1995), and TF/IDF model (Salton et al. 1983; Salton & Buckley 1988) are used.

---

[3] AP for a topic is the average of the precision value obtained after each relevant expert is retrieved. MAP for a set of topics is the average value of the APs of all topics.

For the language model based relevance, we get:

$$P_{content}(d \mid t_i) = \frac{P(t_i \mid d)P(d)}{P(t_i)}$$

$p(t_i|d)$ is estimated by inferring a document language model θ_d for each document d such that

$$p(t_i \mid \theta_d) = p(t_i|\theta_d)^{n(t_i,q)} \qquad (3)$$

where $n(t_i,q)$ is the number of times t_i appears in q. We smooth the language model with the collection model and get:

$$p(t_i \mid \theta_d) = p(t_i \mid d) + \lambda p(t) \qquad (4)$$

where $p(t)$ is the maximum likelihood estimate of the term t_i given the background model, weighted with λ. We smooth the document model with the background model by setting λ to 0.05 in Equation 4.

We used the BM25 equation of Okapi (Robertson et al. 1995) to estimate the conditional probability as follows.

$$P_{content}(d \mid t_i) \propto w\frac{(k_1+1)tf}{K+tf}\frac{(k_3+1)qtf}{k_3+qtf} + k_2\frac{avdl-dl}{avdl+dl} \qquad (5)$$

where $w=log((N-n+0.5)/(n+0.5))$ is the IDF of t_i; N is the number of documents in the dataset; n is the number of documents where t_i appears; K is $k_1((1-b)+b*dl/avdl)$; k_1, b, k_2 and k_3 are parameters; tf is the frequency of t_i in d; qtf is the frequency of t_i in q; dl is the length of d; and $avdl$ is the average document length. Based on the suggested parameter values in Okapi (Robertson et al. 1995), we set the values of k_1, b, k_2 and k_3 as 1.4, 0.6, 0.0, and 8.0, respectively.

The TF/IDF model is used to estimate the conditional probability in Equation 6 as follows.

$$P_{content}(d \mid t_i) \propto coord \cdot \frac{1}{\sqrt{\sum_{t \in q} idf^2}} tf \cdot idf \cdot \frac{avdl}{dl} \qquad (6)$$

where *coord* is the number of query terms that are found in *d* divided by the total number of terms in the query, and *idf* is $1+log(N/(n+1))$.

4.4 Integrating document authority with document relevance

Craswell et al. (2005) proposed a method for combining PageRanks with BM25. Their experiments show that when applying a sigmoid transformation function to PageRank, the MAP of retrieval results on TREC2004 Web Track queries was largely increased. Their sigmoid transformation function is:

$$P_{PageRank}(d) \propto w\frac{PR^a}{k^a + PR^a} \qquad (7)$$

Here w is the weight for combining with $P_{content}(d|t_i)$, PR is the PageRank of d, and a and k are parameters. Based on the parameter settings used by Craswell et al. (2005), we set the values of w, a, and k as 1.8, 0.6, and 1.0, respectively, in Equation 7.

4.5 Modeling multiple levels of associations via multiple windows

Traditional association discovery approaches typically use one fixed window size for co-occurrences in text. Vechtomova et al. (2003) noticed the associations between terms are of multiple levels and used these associations for query expansion, but they have not combined these multiple levels of associations nor applied them to document or entity retrieval.

Based on the characteristics of expertise associations in documents, we propose a novel multiple-windows-based approach in the co-occurrence model, where we take a weighted sum of the association scores between an expert and a query term using different window sizes, respectively. The smaller windows are given higher weights and larger windows are given lower weights. This is consistent with the weighting scheme used in the HAL (Hyperspace Analogue to Language) model, where the weight of a term in a target term's vector is inversely proportional to the distance between them (Burgess et al. 1998). The difference of ours from HAL is that we propose to use multiple sized windows while the latter uses a fixed sized window only. We will show in the experimental evaluation section that the former produce better results.

Suppose that, in a document d, there are M occurrences of a candidate c as $\{c_k\}$ $(k=1,\ldots, M)$. We use L windows with incremental sizes, i.e., $\{W_j\}$ $(j=1,\ldots, L)$, for associating each candidate occurrence c_k with term t_i in d. For c_k, the smallest window in $\{W_j\}$, SW_k, which can enable c_k to co-occur with t_i in SW_k, is used to measure the association between c_k and t_i; if such a window does not exist, the association score between c_k and t_i is zero. For example, suppose that we use three windows $\{20, 40, 80\}$. If one occurrence of a candidate, c_k, does not co-occur with t_i within the 20-sized window but does co-occur with t_i within the 40-sized window, we use the window size 40 to measure their associations. Therefore, for different occurrences of candidates, different window sizes may be used for association discovery. This gives us more flexibility than the use of one fixed sized window only. Thus, in d, the association between c and t_i is a weighted sum of the association scores between all the occurrences of c with t_i, respectively, as follows:

$$P(c\,|\,d,t_i) \propto \sum_{\substack{k=1,\ldots,M \\ c_k \;\; and \;\; t_i \;\; co\text{-}occur \\ SW_k \;\; is \;\; the \;\; smallest}} P(SW(c_k), Section(c_k)) \cdot P(c_k\,|\,d,t_i,SW(c_k),Section(c_k)) \qquad (8)$$

Here $P(SW(c_k), Section(c_k))$ is based on the window size and the section where c_k occurs. Assuming that they are independent, we get:

$$P(SW(c_k), Section(c_k)) = P(SW(c_k))P(Section(c_k)) \qquad (9)$$

We can use different functions to estimate *P(SW(c_k))* and *P(Section(c_k)*, and compare their effects in expert finding. Generally, the smaller the window size, the higher the weight, and the weight is inversely proportional to the window size. The effects of window size and the section where c_k occurs in d will be discussed in the next section.

We extend the co-occurrence model proposed by Cao et al. (2006) to our multiple-window-based co-occurrence model and define $P(c_k|d,t_i,SW_k)$ as:

$$P(c_k \mid d, t_i, SW_k) \propto \mu \frac{pf(c, SW_k)}{pf_{total}(SW_k)} + \frac{1-\mu}{df_c} \sum_{d_i: c \in d_i} \frac{1}{n_c} \sum_{\substack{c_j : c_j \in d_i \\ c_j \text{ and } t_i \text{ co-occur} \\ SW_j \text{ is the smallest}}} \frac{pf(c, SW_j)}{pf_{total}(SW_j)} \tag{10}$$

where $pf(c,SW_k)$ is the frequency of c in window SW_k, $pf_{total}(SW_k)$ is the total frequency of candidates in SW_k, df_c is document frequency of c, n_c is the number of occurrences of c in d_i. We use a Dirichlet prior to smooth parameter μ:

$$\mu = \frac{pf_{total}(SW_k)}{pf_{total}(SW_k) + \kappa}$$

Here κ is the average of term frequency of all occurrences of all candidates inside all windows in the dataset.

4.6 Document internal structure and window size

Since documents on an organizational intranet often follow certain templates in formatting their contents, the template can be used to segment these documents into multiple sections. *P(Section(c_k))* in Equation 9 is decided by the importance of the section where c_k occurs in d. Generally, the more important the section where c_k occurs, the larger the value of *P(Section(c_k))*, e.g., give high importance value to the author section. We used the TREC2005 training topics[4] to train *P(Section(c_k))*. After training, we set *P(Section(c_k))* as 1.0, 7.5, 0.6, 0.2, 5.2, 1.2, 0.7, and 0.5 for candidate occurrences in the document body, author, acknowledgements, references, email sender, email receiver, email CC, and BCC sections, respectively.

P(SW(c_k)) in Equation 9 is determined by the window size. Generally, the larger the window size *SW(c_k)*, the smaller the value of *P(SW(c_k))*. We assume that *P(SW(c_k))* follows a Gaussian distribution function as used by Petkova & Croft (2007b) for combining co-occurrence models.

[4] http://trec.nist.gov/data/enterprise/05/ent05.expert.trainingtopics

5. EXPERIMENTAL EVALUATION

We have proposed incorporating three document features in a two stage model in expert finding. The aim of our evaluation is to test the effectiveness of different relevance models, the incorporation of three document features, namely, via PageRank, multiple sized windows and document internal structure, and query expansion in expert finding. We will use the basic two stage model as our baseline. The main questions we want to systematically investigate in our experiments are as follows:

1. What are the effects of query expansion?
2. Will PageRank improve expert finding with or without query expansion?
3. What are the effects of document relevance model (BM25, TF/IDF, or language model) in expert finding?
4. What are the effects of window size with or without query expansion?
5. What are the effects of multiple windows compared with single window, the number of windows, and window sizes with or without query expansion?
6. What are the effects of document internal structures with or without query expansion?

5.1 Data

The experiments were conducted using the TREC 2006 and 2005 enterprise track expert search task test collections. The dataset is a crawl of the W3C website in June 2004[5]. Table 1 illustrates the email lists (lists), web pages (www), wikis (esw), other pages (other), and personal web pages (people) part of the dataset[6].

The search target is 1092 W3C related people with their names and email addresses[7]. However, people are not always referred to by their exact full names making identifying occurrences of candidates a challenge. We employed rule based approach for automatically generating variants of people's names, e.g., given "Deborah L. McGuinness", the automatically generated variants are "Deborah McGuinness", "McGuinness, Deborah L.", and "McGuinness, D. L." etc, and other advanced named entity recognition techniques such as co-referencing, correspondence between firstnames and nicknames, e.g., "Michael" and "Mike", "Deborah" and "Deb" etc., and conventional correspondence between non-English and English letters, e.g., ë→e, ø→oe etc. Our experiments are based on the annotations of candidate occurrences created by us for the TREC Expert Search

[5] http://research.microsoft.com/users/nickcr/w3c-summary.html

[6] Since develop code part of the dataset mostly consists of programming code and was not very helpful in expert finding as shown by the other people's expert finding experiments , we have also excluded this part from expert finding.

[7] http://trec.nist.gov/data/enterprise/05/ent05.expert.candidates

participants[8], where candidates are recognized by full name, name variations, email addresses, and user ID, etc., using the Aho-Corasick matching algorithm (Aho & Corasick 1975). Our annotations have been widely used by the expert finding research communities (Petkova & Croft 2007b; Westerveld 2007). There are in total 1,662,024 occurrences of candidates in the dataset. A small number of candidates have a huge number of occurrences and the majority of candidates have a small number of occurrences, and the distribution of occurrences for 1092 candidates follows the Zipf's law.

Table 1: W3C collection size in numbers

Scope	Corpus size(gb)	Doc nums	Ave doc size(kb)
lists	1.855	198,394	9.8
www	1.043	45,975	23.8
esw	0.181	19,605	9.7
other	0.047	3,538	14.1
people	0.003	1,016	3.6

In TREC 2005, 50 search topics representing W3C working group names were used and experts were member of these groups. These ground-truth lists were not part of the collection and were used for creating relevance judgments with minimum effort (Craswell et al. 2006).

The TREC2006 expert search test collection consists of 49 search topics contributed by the participating groups. Based on the submitted runs, experts relevant to each topic were evaluated based on their corresponding supporting documents (Soboroff et al. 2007).

We removed HTML tags from the dataset, and used patterns such as regular expressions to segment the documents into multiple sections[9]. We used Lemur (http://www.lemurproject.org/) and Lucene (http://lucene.apache.org/) to index and search the dataset.

5.2 Results

Expert finding results are compared by manipulating different combinations of elements and parameters presented above. These elements and parameters are: QE (query expansion), document relevance model (BM25, TF/IDF, language model), PR (PageRank), SW (single window) or MW (multiple windows), window size, and IS (document internal structure). Results using the TREC2006 test collection are presented in Section 5.2.1 to 5.2.4, and TREC2005 test collection results are presented in Section 5.2.5.

[8] http://ir.nist.gov/w3c/contrib/

[9] Technical reports and academic papers in the dataset are well-structured. For example, typical sections of a technical report (http://www.w3.org/TR/2002/REC-xmldsig-filter2-20021108/) include title, authors and editors, abstract, table of contents, sections of the report, acknowledgements, references etc.

5.2.1 Effects of window size

The three document relevance models combined with a single-window-based association discovery approach are used as baselines, i.e., without employing query expansion, PageRank and internal structure. We have tested 31 window sizes (based on the number of terms and entities) ranging from 5 to 1100. The MAPs of the three baselines are shown in Figure 1. We can clearly see the similarity between all the three baselines. When window size increases, the MAPs increase very quickly at the beginning, then the increase slows down, and finally the MAPs reach a rather stable level at the window size of around 200. When the window size is over around 260, the MAPs of all three baselines stop increasing and even start to decrease slightly when supporting documents are not considered, while the MAPs of all three keep roughly unchanged or even increase slightly when considering supporting documents. This reflects that there are many levels of experts' associations with query topics, e.g., sentence, paragraph, and section levels etc. The increase of a small window size leads to many novel associations discovered with very little noise, resulting in rapidly increasing MAPs. When the window size becomes over 200, there are less novel associations discovered and noise slowly increases, potentially degrading MAP. On the other hand, supporting documents can help curb the noise, thus MAPs can still increase slightly, e.g., the BM25 baseline (with considering supporting documents) in Figure 1. This further confirms that even when the window size is very large, there is possibility of finding novel expertise associations.

Furthermore, from Figure 1 we can see that when the window size is above 200, MAPs of all the three baselines are on a rather stable level, respectively, when both consider and do not consider supporting documents. This shows the robustness of our approach and the results are not sensitive to window size selection. The robustness of our approach with respect to window size is evident when three document relevance models are combined with query expansion and/or document features (shown in Figure 2, 3, 4, 5, 6, and 7) as well. For each two-stage model (a combination of a document relevance model and QE, IS, and/or PR) with or without considering supporting documents in Figure 3 to 7, we performed *t* test (one-tail critical values for significance levels α =0.05 and 0.01, and degree of freedom =96) to compare the 49 APs of each run with window size over 200 of the two-stage model with the 49 APs of the best run of the two-stage model, and found that none of the decreases in MAP are statistically significant.

In Figure 1, overall the language model based expert finding baseline performs better than the other two models in terms of MAP at interpolated points for both with and without supporting documents. However, when we performed *t* test to compare the comparative effectiveness of the three baselines in expert finding, the differences between their MAPs are not statistically significant. It shows that all three document relevance models result in relative comparative expert finding results, i.e., expert finding is not very sensitive to the underlying document relevance model.

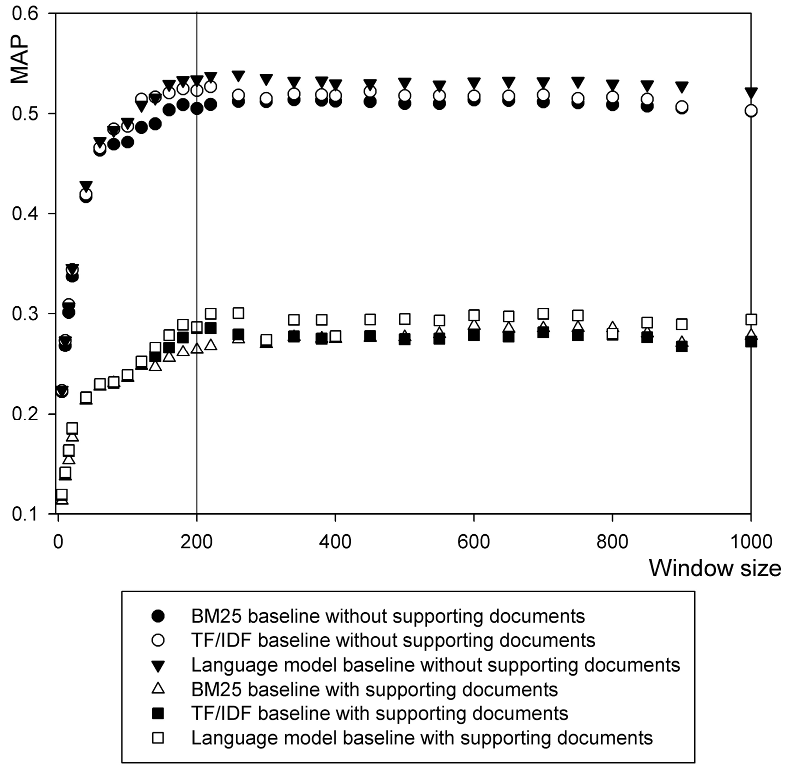

Figure 1: Comparison of three baselines using three document relevance models in terms of MAP versus window size

In terms of the other standard performance measures including R-precision, bpref, precision@5 (P@5), and precision@10 (P@10)[10], both R-precision and bpref follow the same trend as the three baselines shown in Figure 1, i.e., increase quickly with small incremental window sizes, and reach a rather stable level when the window size is above the window size of around 200. There is also no statistically significant difference between the R-precision and bpref values of the three baselines, respectively. The robustness of our approach observed on the MAP for three document relevance models combined with query expansion and/or document features (shown in Figure 3, 4, 5, 6, and 7) also holds for both the R-precision and bpref measures in these settings, respectively.

The assumption for our weighted multiple window approach is that close range co-occurrences often indicate more probable associations between candidates and query terms than longer range co-occurrences. Here we re-examine this assumption via the two precision oriented measures, i.e., P@5 and P@10. Figure 2 shows the P@5 values for different window sizes of three

---

[10] http://trec.nist.gov/pubs/trec15/appendices/CE.MEASURES06.pdf

baselines. We can see that, for each baseline with/without supporting documents, P@5 increases at the beginning until window size of around 60, stays at a rather stable level until the window size of around 600, and decreases slightly when the window size increases further. Comparing Figure 1 and 2, the trend of P@5 differs from that of MAP for three baselines in the following aspects:

In Figure 2, P@5 starts above 0.5 without supporting documents, and 0.3 with supporting documents, increase to over 0.7 with supporting documents, and around 0.5 with supporting documents, respectively. Therefore, the increase of P@5 is not as dramatic as the increase of MAP with respect to window sizes. P@5 reaches to a rather stable level at the window size of around 60 when consider supporting documents, and around 100 when do not consider supporting documents, while MAP reaches a stable level at the window size of around 200.

We think the trend of P@5 and its differences from that of MAP confirm our assumption that close range co-occurrences lead to more probably expertise associations than long range co-occurrences. The interaction of two main factors results in the trend we observe for P@5 in Figure 2. The first factor is the probability of associations between candidates and query terms, and the second factor is the introduction of novel associations between candidates and query terms. When the window size is small, the second factor dominates, i.e., although more noise is introduced in expertise associations, there are many new expertise associations discovered. Therefore, P@5 increases with small incremental window sizes. When the window size is medium, there are fewer new expertise associations discovered, to an extent that the two factors compensate the effect of each other, leading to rather stable level of P@5. When the window size is large, there are even fewer new expertise associations discovered, and introduction of new associations may not be able to compensate the effect of noise in associations, leading to slight decrease of P@5.

Precision @10 has similar trend as Precision @5 for all three baselines. There is no statistically significant difference between the three baselines in terms of their P@5 and P@10, respectively.

The trend of total number of relevant experts retrieved over all topics with respect to window sizes corroborates our above findings. The total number of relevant experts retrieved over all topics increases quickly when the window size starts at 5 and increases until around 60, and stays at a stable level when the window size is beyond 60.

In Section 5.2.4, we will show how our multiple window based approach can help improve the performance of the single window based approach.

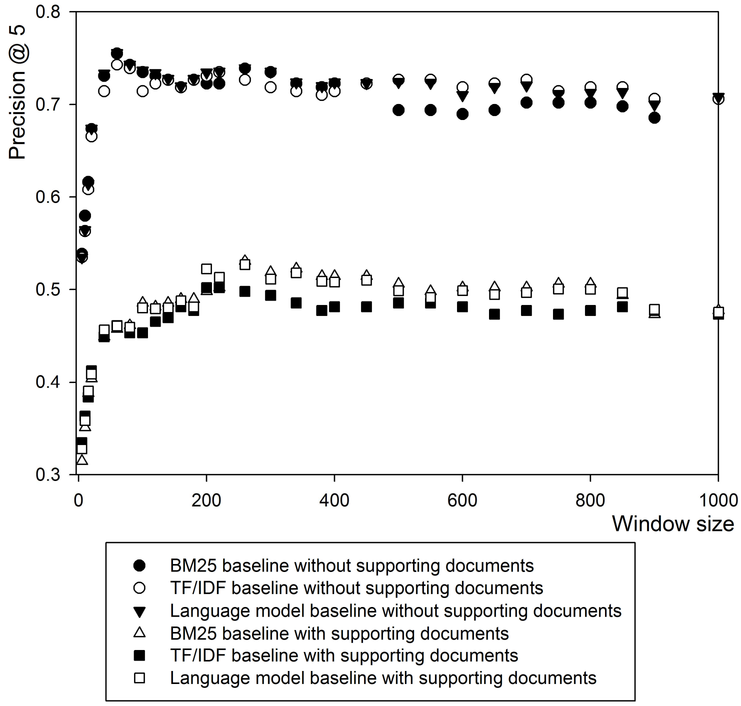

Figure 2: Three baselines using three document relevance models in terms of Precision@5 versus window size

The assumption for our weighted multiple window approach can be further verified by an anatomy of co-occurrences on different levels and their effects on expert finding. We divide a window size of 340 into gap windows of equal length of 20, i.e., 0 to 20, 20 to 40, 40 to 60, 60 to 80, and so on to 320 to 340. We count the total number of co-occurrences of experts and query terms for each gap window, and get performance measures such as MAP and P@5 etc for each gap window. If we divide these performance measure scores by their respective total number of co-occurrences of experts and query terms, the results can give us an idea of how much each co-occurrence for a gap window contributes to the effectiveness of expert finding on average. The higher the contribution, the more useful each co-occurrence is in expert finding, and vice versa. The results are shown in Table 2.

We can clearly see from Table 2 that the average MAP or P@5 score for each co-occurrence both with and without supporting documents consistently decreases when the distances between expert names and query terms increase. In particular, when do not consider supporting documents, the average MAP and P@5 for each co-occurrence of the gap window 320-340 decreases by around 50% from those of the gap window 0-20, respectively, and when consider supporting documents, the average MAP and P@5 for each co-occurrence of the gap window 320-340 decreases by over 85% from those of the gap window 0-20, respectively.

Results in Table 2 support our assumption that close range co-occurrences often indicate more probable associations between candidates and query terms than longer range co-occurrences. Longer range co-occurrences introduce more expertise association information at the expense of more noise, therefore, the average MAP or P@5 score for each co-occurrence decreases as the distances between experts and query terms increase.

Table 2: MAP and P@5 divided by the total number of co-occurrences of experts and query terms for each gap with and without supporting docs respectively. The MAP or P@5 gains based on the 0 to 20 gap window are calculated for each gap window respectively.

Gaps	Without supporting docs		With supporting docs	
	Avg. MAP for each co-occurrence ($\times 10^{-7}$)	Avg. P@5 for each co-occurrence ($\times 10^{-7}$)	Avg. MAP for each co-occurrence ($\times 10^{-7}$)	Avg. P@5 for each co-occurrence ($\times 10^{-7}$)
0-20	28.010	53.676	15.120	32.675
20-40	23.779 (-15.11%)	45.359 (-15.49%)	9.923 (-34.37%)	23.238 (-28.88%)
40-60	23.463 (-16.23%)	45.182 (-15.82%)	7.668 (-49.29%)	21.75 (-33.44%)
60-80	22.359 (-20.17%)	43.501 (-18.96%)	6.927 (-54.19%)	20.384 (-37.62%)
80-100	20.18 (-27.95%)	40.496 (-24.55%)	5.251 (-65.27%)	18.459 (-43.51%)
100-120	19.501 (-30.38%)	39.759 (-25.93%)	4.302 (-71.55%)	13.975 (-57.23%)
120-140	18.86 (-32.67%)	39.106 (-27.14%)	3.818 (-74.75%)	13.487 (-58.72%)
140-160	18.327 (-34.57%)	36.591 (-31.83%)	3.816 (-74.76%)	12.033 (-63.17%)
160-180	17.789 (-36.49%)	34.601 (-35.54%)	3.378 (-77.66%)	9.597 (-70.63%)
180-200	15.494 (-44.68%)	31.498 (-41.32%)	3.184 (-78.94%)	9.648 (-70.47%)
200-220	15.32 (-45.31%)	31.591 (-41.15%)	2.158 (-85.73%)	9.051 (-72.30%)
220-240	14.172 (-49.40%)	29.057 (-45.87%)	2.106 (-86.07%)	8.781 (-73.13%)
240-260	14.763 (-47.29%)	30.559 (-43.07%)	1.9 (-87.43%)	6.838 (-79.07%)
260-280	14.042 (-49.87%)	30.243 (-43.66%)	1.799 (-88.10%)	6.68 (-79.56%)
280-300	12.819 (-54.23%)	28.285 (-47.30%)	1.338 (-91.15%)	6.041 (-81.51%)
300-320	12.141 (-56.65%)	28.104 (-47.64%)	1.268 (-91.61%)	4.917 (-84.95%)
320-340	11.923 (-57.43%)	27.976 (-47.88%)	1.234 (-91.84%)	4.762 (-85.43%)

5.2.2 Effects of query expansion and document features

Query expansion is added together with the three document features to the baseline models to see whether they can help improve the performance. To judge their effectiveness, we first apply each of them to the BM25 baseline. In Figure 3, all the runs respectively enhanced by query expansion, PageRank, and document internal structures, show similar trends with the BM25 baseline in Figure 1 with respect to window sizes. The incorporation of query expansion, PageRank and internal structure individually in TF/IDF and language model baselines give analogous results.

Query expansion gives the biggest boost to the MAPs for both cases of taking and not taking into account supporting documents, due to the fact that many of the original topics are not very complete descriptions of the expert finding tasks, and the co-occurrences of automatically expanded terms and candidates provide additional evidences in expert finding. Considering internal structure also does increase the performance in both cases.

We used the PageRanks contributed by Danil Nemirovsky (http://ir.nist.gov/w3c/contrib/) and the Equation 7 for combining PageRank with the BM25 model. It is interesting to observe that PageRank slightly increases the MAPs when not considering supporting documents, but slightly hurts the performance when considering supporting documents. We have used another set of PageRanks contributed by SJTU (http://ir.nist.gov/w3c/contrib/) and the transformation function in Equation 7. Again, little improvement is achieved in MAPs by introducing PageRank.

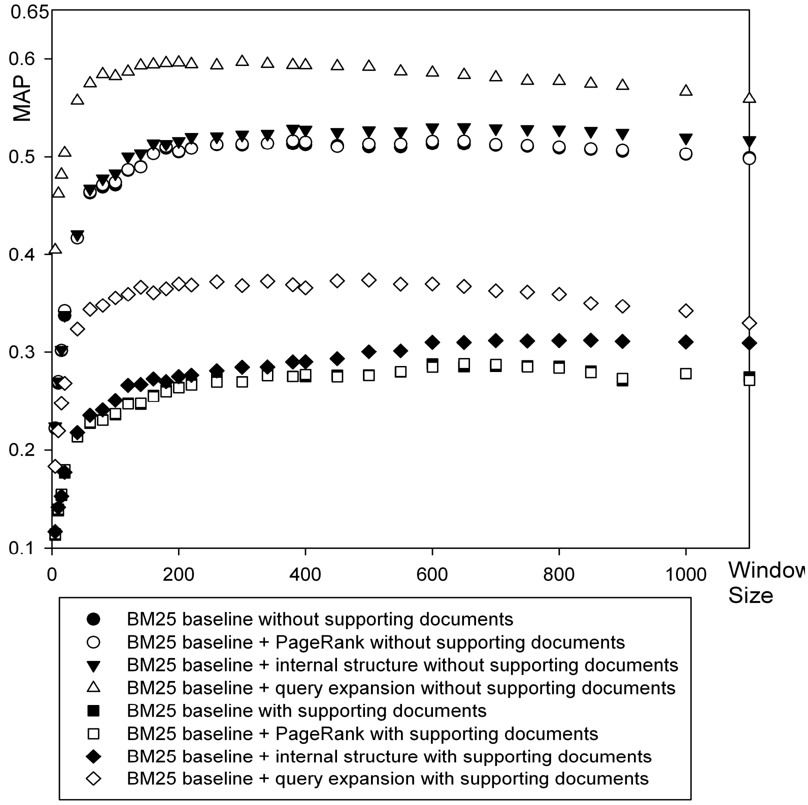

Figure 3: Effect of document features and query expansion in BM25 baseline in terms of MAP versus widow size

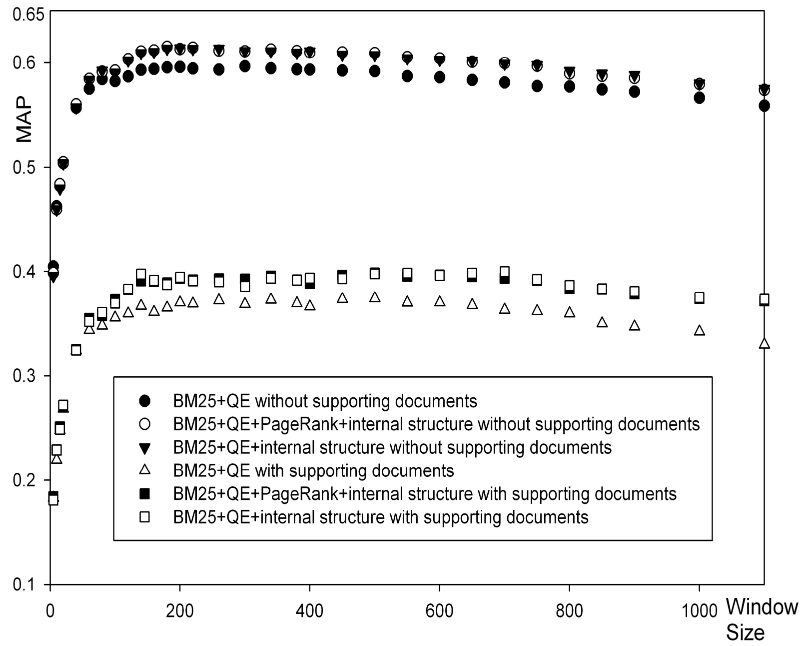

Figure 4: Effect of document features in query expansion enhanced BM25 baseline in terms of MAP versus window size

PageRank was also applied to the TF/IDF and language model baselines respectively where it degraded MAP slightly. In further work, we will further study the transformation functions of PageRank and train the parameters in Equations 7 to see whether PageRanks can improve different baseline models.

We have seen that query expansion helps increase MAP dramatically. Now we apply one or more document features to the query expansion enhanced BM25 baseline. The results are shown in Figure 4. It can be seen that internal structure can complement the effect of query expansion by helping further improve the MAPs of BM25+QE. The PageRank only slightly helps increase the MAPs when do not consider supporting documents and the window size is small. When considering supporting documents, the MAP scores stay roughly the same no matter whether PageRank is taken into account or not. However, the BM25 baseline integrated with query expansion, internal structure and PageRank together can lead to slightly better results than the other two models in Figure 4 when using a fixed window size.

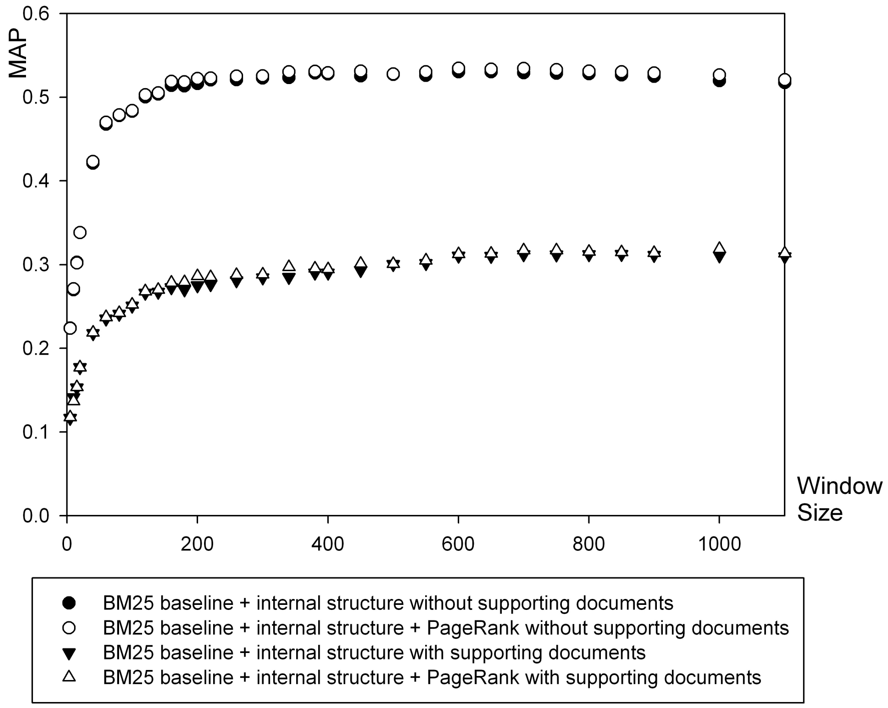

Figure 5: Effect of PageRank in internal structure enhanced BM25 baseline in terms of MAP versus widow size

To verify the effect of PageRank, firstly, we compare the BM25 baseline enhanced by internal structure with the BM25 baseline enhanced by both internal structure and PageRank in Figure 5. We can see that PageRank does not help improve MAP for some window sizes and only slightly help improve MAP in the other window sizes. Secondly, we compare the BM25 baseline enhanced by query expansion with the BM25 baseline enhanced by both query expansion and PageRank in Figure 6. We can see that PageRank does not help improve MAP when do not consider supporting documents, and even hurt the MAP when consider supporting documents.

Similar observations can be drawn from the integrations of query expansion enhanced TF/IDF and language model baselines with internal structure. However, as mentioned, PageRank does not help improve MAP significantly.

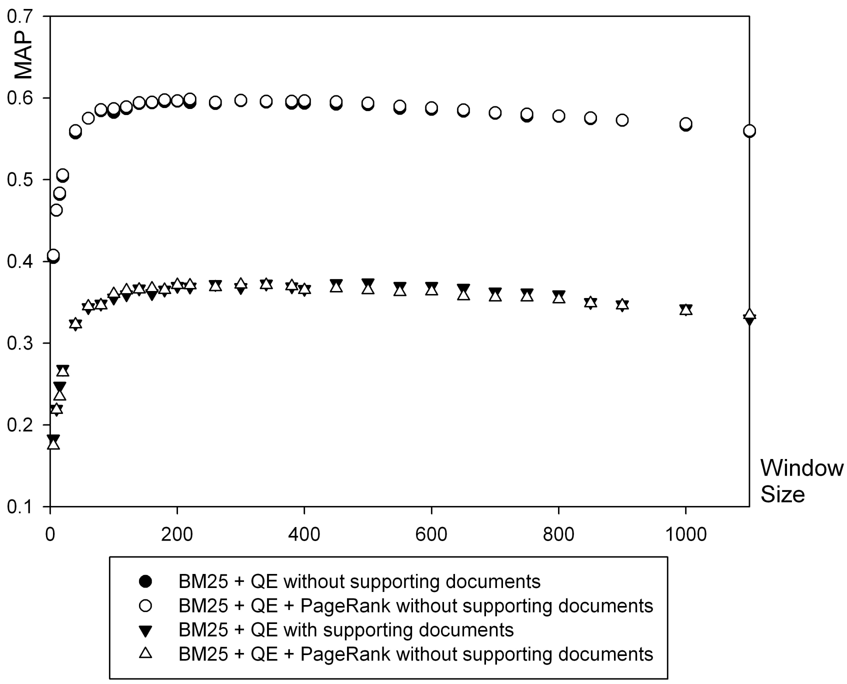

Figure 6: Effect of PageRank in query expansion enhanced BM25 baseline in terms of MAP versus window size

5.2.3 Effects of document relevance models

Figure 1 shows the effects of document relevance models in the baselines. The language model baseline performed slightly better than both the TF/IDF and BM25 baselines. Our assessment is that BM25 and TF/IDF models produce ranking scores which are not true probabilities while language model produces true probabilities which seem more suitable for linear combination, since in Equation 1 we linearly combine the ranking score of each document by multiplying it with the score from the co-occurrence model and aggregating the ranking score over a number of relevant documents.

In Figure 7, we compare the three document relevance models, BM25+PageRank, TF/IDF, and language model, integrated with both internal structure and query expansion, which are the best performing combinations for the three document relevance models in terms of MAP, respectively. Overall, the language model based approach performs slightly better than both the BM25 and TF/IDF based approaches in terms of MAP for both with and without supporting documents. However, there is no statistically significant difference among the performances of the three models when we did *t* test to compare them on all 31 window sizes.

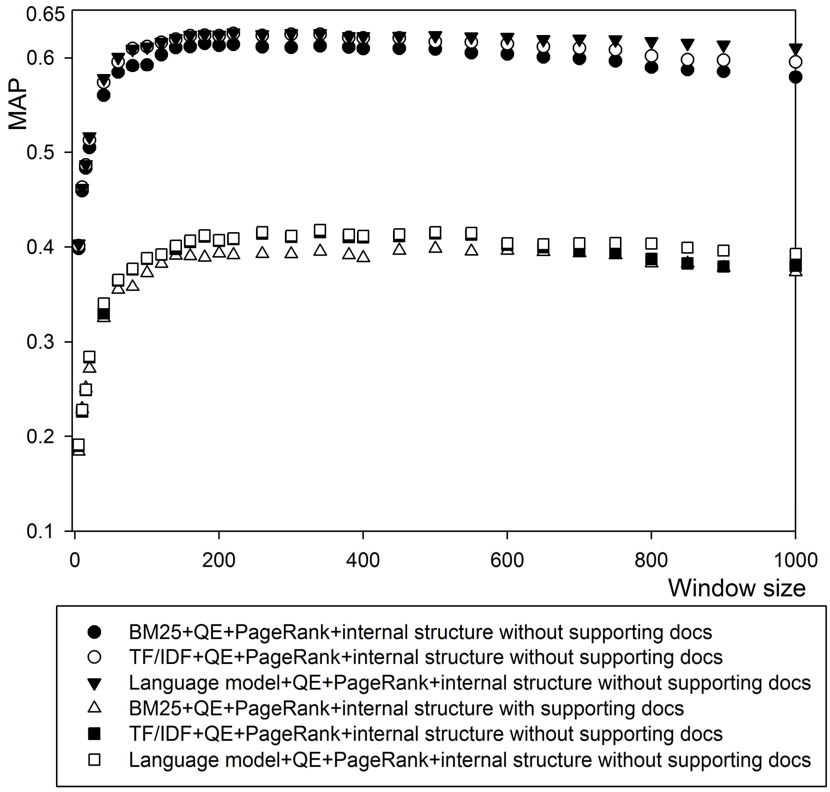

Figure 7: Comparison of BM25 +PageRank+QE+IS, TF/IDF+QE+IS, and language model+QE+IS in terms of MAP versus window size

In Table 3 and 4, we summarize the performance improvement of the three document relevance models integrated with document internal structure, PageRank, and/or query expansion over a BM25 baseline for with and without supporting documents, respectively. We selected six window sizes, i.e., 5, 20, 80, 200, 400, and 800, to represent a range of different levels of associations. We can see from Table 3 and 4 that query expansion boosts the performance of all three models dramatically, document internal structure helps improve the three models' performance, and PageRank does not significantly help improve the three models. In addition, document internal structure can complement query expansion enhanced models by improving their performance further. Language model based approach is the overall best performing one, and closely followed by the TF/IDF based approach.

Table 3: MAPs and MAP gains on the basis of the BM25 baseline model for three models integrated with document internal structure, PageRank, and query expansion when do not consider supporting documents. The highest MAP for a window size is in bold and underlined

Models	5	20	80	200	400	800
BM25	0.2220	0.3375	0.4692	0.5049	0.5123	0.5088
TF/IDF	0.2233 (0.59%)	0.3441 (1.96%)	0.4847 (3.30%)	0.5229 (3.57%)	0.5177 (1.05%)	0.5162 (1.45%)
LM	0.2244 (1.08%)	0.3448 (2.16%)	0.4854 (3.45%)	0.5235 (3.68%)	0.5182 (1.15%)	0.5165 (1.51%)
BM25+IS	0.2248 (1.3%)	0.3424 (1.45%)	0.4779 (1.85%)	0.5163 (2.26%)	0.5278 (3.03%)	0.5279 (3.75%)
TF/IDF+IS	0.2273 (2.39%)	0.345 (2.22%)	0.4898 (4.39%)	0.5376 (6.48%)	0.5391 (5.23%)	0.5333 (4.82%)
LM+IS	0.2281 (2.75%)	0.3455 (2.37%)	0.4905 (4.54%)	0.5382 (6.60%)	0.5394 (5.29%)	0.5339 (4.93%)
BM25+QE	0.4047 (82.30%)	0.504 (49.33%)	0.5842 (24.51%)	0.596 (18.04%)	0.5933 (15.81%)	0.5773 (13.46%)
TF/IDF+QE	0.4069 (83.29%)	0.5088 (50.76%)	0.5913 (26.02%)	0.594 (17.65%)	0.5922 (15.60%)	0.5776 (13.52%)
LM+QE	0.4074 (83.51%)	0.5105 (51.26%)	0.5921 (26.19%)	0.5962 (18.08%)	0.5943 (16.01%)	0.5792 (13.84%)
BM25+PR	0.2226 (0.27%)	0.3414 (1.2%)	0.4715 (0.49%)	0.5058 (0.18%)	0.5152 (0.57%)	0.5104 (0.31%)
TF/IDF+PR	0.223 (0.45%)	0.3409 (1.01%)	0.4798 (2.26%)	0.5148 (1.96%)	0.5144 (0.41%)	0.5095 (0.14%)
LM+PR	0.2236 (0.72%)	0.3417 (1.24%)	0.4807 (2.45%)	0.5152 (2.04%)	0.5157 (0.66%)	0.5102 (0.28%)
BM25+IS+PR	0.2237 (0.77%)	0.3408 (0.98%)	0.4756 (1.36%)	0.517 (2.40%)	0.5223 (1.95%)	0.5239 (2.97%)
TF/IDF+IS+PR	0.2267 (2.12%)	0.3438 (1.87%)	0.4873 (3.86%)	0.5382 (6.60%)	0.5339 (4.22%)	0.5295 (4.07%)
LM+IS+PR	0.2279 (2.66%)	0.3447 (2.13%)	0.4882 (4.05%)	0.5393 (6.81%)	0.5347 (4.37%)	0.5362 (5.39%)
BM25+QE+PR	0.4067 (83.20%)	0.5071 (50.25%)	0.5851 (24.70%)	0.5943 (17.71%)	0.5941 (15.97%)	0.575 (13.01%)
TF/IDF+QE+PR	0.4081 (83.83%)	0.5112 (51.47%)	0.5921 (26.19%)	0.5923 (17.31%)	0.5924 (15.64%)	0.5754 (13.09%)
LM+QE+PR	<u>**0.4087**</u> (84.10%)	0.5121 (51.73%)	0.5928 (26.34%)	0.5931 (17.47%)	0.5929 (15.73%)	0.5763 (13.27%)
BM25+QE+IS	0.3956 (78.20%)	0.5038 (49.27%)	<u>**0.6139**</u> (30.84%)	0.6139 (21.59%)	0.6108 (19.23%)	0.5921 (16.37%)
TF/IDF+QE+IS	0.4015 (80.86%)	0.5133 (52.09%)	0.6097 (29.94%)	0.6231 (23.41%)	0.6208 (21.18%)	0.602 (18.32%)
LM+QE+IS	0.4021 (81.13%)	0.5141 (52.33%)	0.6107 (30.16%)	<u>**0.6236**</u> (23.51%)	<u>**0.6214**</u> (21.30%)	<u>**0.6102**</u> (19.93%)
BM25+QE+PR+IS	0.3986 (79.55%)	0.5049 (49.60%)	0.5915 (26.07%)	0.6131 (21.43%)	0.6097 (19.01%)	0.5897 (15.90%)
TF/IDF+QE+PR+IS	0.4033 (81.67%)	0.5136 (52.18%)	0.6085 (29.69%)	0.6206 (22.92%)	0.6183 (20.69%)	0.5993 (17.79%)
LM+QE+PR+IS	0.4039 (81.94%)	<u>**0.5142**</u> (52.36%)	0.6087 (29.73%)	0.6212 (23.03%)	0.6201 (21.04%)	0.6078 (19.46%)

Table 4: MAPs and MAP gains on the basis of the BM25 baseline model for three models integrated with document internal structure, PageRank, and query expansion when consider supporting documents. The highest MAP for a window size is in bold and underlined

Models	5	20	80	200	400	800
BM25	0.1132	0.1763	0.2307	0.2641	0.2751	0.2856
TF/IDF	0.1189 (5.04%)	0.1849 (4.88%)	0.231 (0.13%)	0.2858 (8.22%)	0.2771 (0.73%)	0.2793 (-2.21%)
LM	0.1186 (4.77%)	0.1852 (5.05%)	0.2319 (0.52%)	0.2871 (8.71%)	0.2784 (1.20%)	0.2814 (-1.47%)
BM25+IS	0.1166 (3.00%)	0.1771 (0.45%)	0.2409 (4.42%)	0.275 (4.13%)	0.2902 (5.49%)	0.3121 (9.28%)
TF/IDF+IS	0.1251 (10.51%)	0.1936 (9.81%)	0.2548 (10.45%)	0.3053 (15.60%)	0.3041 (10.54%)	0.3056 (7.00%)
LM+IS	0.1247 (10.16%)	0.1942 (10.15%)	0.2557 (10.84%)	0.3104 (17.53%)	0.3055 (11.05%)	0.3067 (7.39%)
BM25+QE	0.183 (61.66%)	0.2682 (52.13%)	0.3479 (50.80%)	0.37 (40.10%)	0.3661 (33.08%)	0.3595 (25.88%)
TF/IDF+QE	0.1862 (64.49%)	0.2686 (52.35%)	0.3515 (52.36%)	0.3666 (38.81%)	0.3655 (32.86%)	0.3466 (21.36%)
LM+QE	0.1868 (65.02%)	0.2694 (52.81%)	0.3523 (52.71%)	0.3679 (39.30%)	0.3668 (33.33%)	0.3482 (21.92%)
BM25+PR	0.1143 (0.97%)	0.1797 (1.93%)	0.2302 (-0.22%)	0.2638 (-0.11%)	0.2771 (0.73%)	0.2836 (-0.70%)
TF/IDF+PR	0.1165 (2.92%)	0.1823 (3.40%)	0.2244 (-2.73%)	0.2741 (3.79%)	0.2731 -(0.73%)	0.2684 (-6.02%)
LM+PR	0.1173 (3.62%)	0.1831 (3.86%)	0.2254 (-2.30%)	0.2753 (4.24%)	0.2745 -(0.22%)	0.2703 (-5.36%)
BM25+IS+PR	0.1173 (3.62%)	0.1765 (0.11%)	0.238 (3.16%)	0.2749 (4.09%)	0.2887 (4.94%)	0.3038 (6.37%)
TF/IDF+IS+PR	0.1256 (10.95%)	0.1946 (10.38%)	0.2539 (10.06%)	0.3051 (15.52%)	0.3024 (9.92%)	0.3003 (5.15%)
LM+IS+PR	0.1263 (11.57%)	0.1951 (10.66%)	0.2547 (10.40%)	0.3061 (15.90%)	0.3028 (10.07%)	0.3018 (5.67%)
BM25+QE+PR	0.1844 (62.90%)	0.2664 (51.11%)	0.3488 (51.19%)	0.3684 (39.49%)	0.3631 (31.99%)	0.3476 (21.71%)
TF/IDF+QE+PR	0.1871 (65.28%)	0.2661 (50.94%)	0.3528 (52.93%)	0.3641 (37.86%)	0.3628 (31.88%)	0.3389 (18.66%)
LM+QE+PR	0.1878 (65.90%)	0.2672 (51.56%)	0.3535 (53.23%)	0.3652 (38.28%)	0.3642 (32.39%)	0.3405 (19.22%)
BM25+QE+IS	0.181 (59.89%)	0.2721 (54.34%)	0.3604 (56.22%)	0.3943 (49.30%)	0.3935 (43.04%)	0.3863 (35.26%)
TF/IDF+QE+IS	0.1899 (67.76%)	0.2836 (60.86%)	0.3767 (63.29%)	0.4065 (53.92%)	0.4103 (49.15%)	0.3875 (35.68%)
LM+QE+IS	0.1904 (68.20%)	0.2835 (60.81%)	<u>**0.3778**</u> (63.76%)	<u>**0.407**</u> (54.11%)	<u>**0.4107**</u> (49.29%)	<u>**0.3921**</u> (37.29%)
BM25+QE+PR+IS	0.1841 (62.63%)	0.2716 (54.06%)	0.3579 (55.14%)	0.3935 (49.00%)	0.3886 (41.26%)	0.3833 (34.21%)
TF/IDF+QE+PR+IS	0.191 (68.73%)	0.2833 (60.69%)	0.3722 (61.34%)	0.4032 (52.67%)	0.4054 (47.36%)	0.3815 (33.58%)
LM+QE+PR+IS	<u>**0.1918**</u> (69.43%)	<u>**0.2841**</u> (61.15%)	0.3736 (61.94%)	0.4057 (53.62%)	0.4063 (47.69%)	0.3885 (36.03%)

5.2.4 Effects of using multiple sized windows

Our multiple window based approach is based on the assumption that associations between candidates and query terms are of multiple levels, and small range associations are more likely to be accurate than long range associations. We experimented with combining multiple windows for with and without query expansion, and the results are shown in Table 5 to 9. We selected six window sizes, i.e., 5, 20, 80, 200, 400, and 800. In Table 5 to 9, we start with each individual window in a single window based expert finding approach, and explore all its combinations with one or more of the other five windows. Given each individual window, we report up to three top performing combinations for the window in terms of MAP and MAP gain. We observe from Table 5 to 9 the followings.

Firstly, it is impressive that multiple windows outperform single windows in terms of all three document relevance models (Table 5 to 9), query expansion enhanced models (Table 5 and 6), models without query expansion (Table 7, 8 and 9), models with supporting documents (Table 6 and 8), and models without supporting documents (Table 5, 7 and 9). In fact, our additional experiments show that multiple windows also produce higher MAPs in expert finding than all the 31 single windows used in Section 5.2.1 for the above integrated models, respectively.

Secondly, three windows largely outperform two windows, and four windows outperform three windows in terms of MAP. Most of the highest MAPs for a particular model are produced by four or five windows. However, four, five, and six windows perform comparatively.

Thirdly, six windows produce the highest MAPs only for two models in Table 6, and perform comparatively with four and six windows for the rest. Therefore, more windows do not necessarily lead to better performance.

Fourthly, certain window combinations consistently produce better results than other window combinations regardless of the document relevance model, document features, and query expansion. Our experiments on window combination can help us find these optimal window combinations.

Fifthly, both language model and TF/IDF based multiple windows outperform BM25 based multiple windows, and language model based multiple windows slightly outperform TF/IDF based multiple windows.

In Table 5 and 6, the highest MAP of 0.6559 (without supporting documents) is achieved by language model with query expansion and internal structure for the five window combination, i.e., 5, 20, 80, 200, and 400, and the highest MAP of 0.4545 (with supporting documents) is achieved again by language model with query expansion and internal structure for the six window combination, i.e., 5, 20, 80, 200, 400, and 800. These two highest MAP scores are both higher than those of our best runs in TREC2006 expert search task, which were also the best runs among all participating groups, respectively. In TREC2006 expert search task, the highest MAP of 0.6431 (without supporting documents) and 0.4421 were both achieved by our best run using the TF/IDF model, query expansion, and ten window sizes, i.e., 10, 28, 48, 88, 160, 280, 360, 660, 1200, and 3200. This further confirms that more windows do not necessarily lead to better results.

In Table 7 and 8, when do not consider query expansion, the highest MAP of 0.5673 (without supporting documents) is achieved by language model with internal structure for the four window combination, i.e., 20, 80, 200, and 800, and the highest MAP of 0.3505 (with supporting documents) is achieved by language model with internal structure for the five window combination, i.e., 20, 80, 200, 400, and 800. Our results have significantly outperformed the automatic runs reported by Macdonald & Ounis (2007)[11], and Petkova & Croft (2007a)[12], respectively.

Following our multiple window based approach (Zhu et al. 2007a), Petkova and Croft (2007b) presented a generative language modeling approach for expert finding which is based on estimating the joint distribution of terms and experts. Their experimental results also show that a step function, which is equivalent to our multiple window approach, has produced better retrieval results than both a triangle and Gaussian functions. In our approach, the integration of other document features such as document internal structure has further improved the performance of the multiple-window based approach.

In Table 5, BMPQI, TQI, and LMQI stand for BM25+PageRank, TF/IDF, and language model with query expansion, internal structure and without supporting documents, respectively.

Table 5: MAPs and MAP gains for BM25+PageRank+QE+IS, TF/IDF+QE+IS, and language model+QE+IS with different window combinations when do not consider supporting documents, and the highest MAP for a model is in bold and underlined

Base window	**Mode**	**BM25+PageRank+QE+IS**		**TF/IDF+QE+IS**		**Language model+QE+IS**	
		Windows	**MAP (MAP gain)**	**Windows**	**MAP (MAP gain)**	**Windows**	**MAP (MAP gain)**
5	1 window	5	0.3986	5	0.4015	5	0.4021
	2 windows	5, 400	0.6288 (57.8%)	5, 400	0.6407 (59.6%)	5, 400	0.6403 (59.2%)
		5, 200	0.6262 (57.1%)	5, 200	0.6379 (58.9%)	5, 200	0.6359 (58.2%)
		5, 800	0.6136 (53.9%)	5, 80	0.6182 (54.0%)	5, 80	0.6171 (53.5%)
	3 windows	5, 80, 400	0.6329 (58.8%)	5, 80, 400	0.6522 (62.4%)	5, 80, 400	0.6527 (62.3%)
		5, 200, 400	0.6325 (58.7%)	5, 200, 400	0.6489 (61.6%)	5, 200, 400	0.6522 (62.2%)
		5, 20, 400	0.6296 (58.0%)	5, 80, 200	0.6432 (60.2%)	5, 20, 400	0.6483 (61.2%)
	4 windows	5,80,200,400	0.6365 (59.7%)	5,80,200,400	**0.6536** (62.8%)	5,80,200,400	0.6558 (63.1%)
		5,80,200,800	0.6358 (59.5%)	5,20,80,400	0.6522 (62.4%)	5,20,80,400	0.6535 (62.5%)
		5,20,200,400	0.6349 (59.3%)	5,20,200,400	0.6499 (61.9%)	5,20,200,400	0.6512 (61.9%)
	5 windows	5,20,80,200, 400	0.6362 (59.6%)	5,20,80,200, 400	0.6535 (62.8%)	5,20,80,200, 400	**0.6559** (63.1%)
		5,80,200,400, 800	0.6355 (59.4%)	5,80,200,400, 800	0.6460 (60.9%)	5,80,200,400, 800	0.6503 (61.7%)
	6 windows	All windows	0.6360 (59.6%)	All windows	0.6463 (61.0%)	All windows	0.6524 (62.3%)
20	1 window	20	0.5049	20	0.5133	20	0.5141
	2 windows	20, 400	0.6290 (24.6%)	20, 400	0.6424 (25.2%)	20, 400	0.6421 (24.9%)
		20, 200	0.6272 (24.2%)	20, 200	0.6379 (24.3%)	20, 200	0.6408 (24.7%)
		20, 800	0.6158 (22.0%)	20, 800	0.6197 (20.7%)	20, 800	0.6232 (21.2%)

[11] Their highest MAP is 0.5210 when do not consider supporting documents.
[12] Their highest MAP is 0.5016 when do not consider supporting documents.

	3 windows	20, 80, 400	0.6342 (25.6%)	20, 80, 400	0.6508 (26.8%)	20, 80, 400	0.6535 (27.1%)
		20, 200, 400	0.6342 (25.6%)	20, 200, 400	0.6482 (26.3%)	20, 200, 400	0.6523 (26.9%)
		20, 200, 800	0.6296 (24.7%)	5, 20, 400	0.6426 (25.2%)	20, 200, 800	0.6436 (25.2%)
	4 windows	20,80,200, 800	**0.6366** (26.1%)	20,80,200, 400	0.6524 (27.1%)	20,80,200, 800	0.6546 (27.3%)
		20,80,200, 400	0.6363 (26.0%)	5,20,80,400	0.6522 (27.1%)	20,80,200, 400	0.6545 (27.3%)
		5,20,80, 400	0.6349 (25.7%)	5,20,200,400	0.6499 (26.6%)	5,20,80, 400	0.6535 (27.1%)
	5 windows	20,80,200, 400,800	0.6363 (26.0%)	5,20,80,200, 400	0.6535 (27.3%)	5,20,80,200, 400	**0.6559** (27.6%)
		5,20,80,200, 400	0.6362 (26.0%)	20,80,200, 400,800	0.6450 (25.7%)	20,80,200, 400,800	0.6450 (25.5%)
	6 windows	All windows	0.6360 (26.0%)	All windows	0.6463 (25.9%)	All windows	0.6524 (26.9%)
80	1 window	80	0.5915	80	0.6097	80	0.6107
	2 windows	80, 400	0.6345 (7.3%)	80, 400	0.6509 (6.8%)	80, 400	0.6542 (7.1%)
		80, 800	0.6260 (5.8%)	80, 200	0.6409 (5.1%)	80, 800	0.6439 (5.4%)
		80, 200	0.6253 (5.7%)	80, 800	0.6315 (3.6%)	80, 200	0.6357 (4.1%)
	3 windows	80, 200, 400	0.6365 (7.6%)	5, 80, 400	0.6522 (7.0%)	80, 200, 400	0.6550 (7.3%)
		20, 80, 400	0.6342 (7.2%)	80, 200, 400	0.6511 (6.8%)	20, 80, 400	0.6535 (7.0%)
		80, 200, 800	0.6327 (7.0%)	20, 80, 400	0.6508 (6.7%)	5, 80, 400	0.6527 (6.9%)
	4 windows	20,80,200, 800	**0.6366** (7.6%)	5,80,200,400	**0.6536** (7.2%)	5,80,200,400	0.6558 (7.4%)
		5,80,200,400	0.6365 (7.6%)	20,80,200, 400	0.6524 (7.0%)	20,80,200, 800	0.6546 (7.2%)
		20,80,200, 400	0.6363 (7.6%)	5,20,80,400	0.6522 (7.0%)	20,80,200, 400	0.6545 (7.2%)
	5 windows	20,80,200, 400,800	0.6363 (7.6%)	5,20,80,200, 400	0.6535 (7.2%)	5,20,80,200, 400	**0.6559** (7.4%)
		5,20,80,200, 400	0.6362 (7.6%)	5,80,200,400, 800	0.6460 (6.0%)	20,80,200, 400,800	0.6450 (5.6%)
	6 windows	All windows	0.6360 (7.5%)	All windows	0.6463 (6.0%)	All windows	0.6524 (6.8%)
200	1 window	200	0.6131	200	0.6231	200	0.6236
	2 windows	200, 400	0.6310 (2.9%)	200, 400	0.6442 (3.4%)	200, 400	0.6447 (3.4%)
		20, 200	0.6272 (2.3%)	80, 200	0.6409 (2.9%)	80, 800	0.6439 (3.3%)
		200, 800	0.6264 (2.2%)	20, 200	0.6379 (2.4%)	200, 800	0.6389 (2.5%)
	3 windows	80, 200, 400	0.6365 (3.8%)	80, 200, 400	0.6511 (4.5%)	80, 200, 400	0.655 (5.0%)
		20, 200, 400	0.6342 (3.4%)	5, 200, 400	0.6495 (4.2%)	80, 200, 800	0.6524 (4.6%)
		80, 200, 800	0.6327 (3.2%)	20, 200, 400	0.6482 (4.0%)	20, 200, 400	0.6523 (4.6%)
	4 windows	20,80,200, 800	**0.6366** (3.8%)	5,80,200,400	**0.6536** (4.9%)	5,80,200,400	**0.6559** (5.2%)
		5,80,200,400	0.6365 (3.8%)	20,80,200, 400	0.6524 (4.7%)	20,80,200, 400	0.6546 (5.0%)
		20,80,200, 400	0.6363 (3.8%)	5,20,200,400	0.6499 (4.3%)	5,20,200,400	0.6512 (4.4%)
	5 windows	20,80,200, 400,800	0.6363 (3.8%)	5,20,80,200, 400	0.6535 (4.9%)	5,20,80,200, 400	0.6559 (5.2%)
		5,20,80,200, 400	0.6362 (3.8%)	5,80,200,400, 800	0.6460 (3.7%)	5,80,200,400, 800	0.6503 (4.3%)
	6 windows	All windows	0.6360 (3.7%)	All windows	0.6463 (3.7%)	All windows	0.6524 (4.6%)
400	1 window	400	0.6097	400	0.6208	400	0.6214
	2 windows	80, 400	0.6345 (4.1%)	80, 400	0.6509 (4.8%)	80, 400	0.6542 (5.3%)
		200, 400	0.6310 (3.5%)	200, 400	0.6442 (3.8%)	200, 400	0.6447 (3.7%)
		20, 400	0.6290 (3.2%)	20, 400	0.6424 (3.5%)	20, 400	0.6421 (3.3%)
	3 windows	80, 200, 400	0.6365 (4.4%)	5, 80, 400	0.6522 (5.1%)	5, 80, 400	0.6527 (5.0%)
		20, 80, 400	0.6342 (4.0%)	80, 200, 400	0.6511 (4.9%)	80, 200, 400	0.6523 (5.0%)
		20, 200, 400	0.6342 (4.0%)	5, 200, 400	0.6495 (4.6%)	20, 200, 400	0.6523 (5.0%)
	4 windows	5,80,200,400	0.6365 (4.4%)	5,80,200,400	**0.6536** (5.3%)	5,80,200,400	0.6558 (5.6%)
		20,80,200, 400	0.6363 (4.4%)	5,20,80,400	0.6522 (5.1%)	5,20,80,400	0.6535 (5.2%)
		5,20,200,400	0.6349 (4.1%)	5,20,200,400	0.6499 (4.7%)	5,20,200,400	0.6512 (4.8%)
	5 windows	20,80,200, 400,800	0.6363 (4.4%)	5,20,80,200, 400	0.6535 (5.3%)	5,20,80,200, 400	**0.6559** (5.6%)
		5,20,80,200, 400	0.6362 (4.3%)	5,80,200,400, 800	0.6460 (4.1%)	20,80,200, 400,800	0.645 (3.8%)
	6 windows	All windows	0.6360 (4.3%)	All windows	0.6463 (4.1%)	All windows	0.6524 (5.0%)

800	1 window	800	0.5897	800	0.6020	800	0.6102
	2 windows	200, 800	0.6264 (6.2%)	200, 800	0.6349 (5.5%)	80, 800	0.6439 (5.5%)
		80, 800	0.6260 (6.2%)	80, 800	0.6315 (4.9%)	200, 800	0.6389 (4.7%)
		400, 800	0.6223 (5.5%)	400, 800	0.6295 (4.6%)	400, 800	0.6308 (3.4%)
	3 windows	80, 200, 800	0.6327 (7.3%)	80, 400, 800	0.6399 (6.3%)	80, 400, 800	0.6487 (6.3%)
		20, 200, 800	0.6296 (6.8%)	200, 400, 800	0.6396 (6.2%)	80, 200, 800	0.6482 (6.2%)
		200, 400, 800	0.6294 (6.7%)	80, 200, 800	0.6393 (6.2%)	200, 400, 800	0.6475 (6.1%)
	4 windows	20,80,200, 800	**<u>0.6366</u>** (8.0%)	5,80,200,800	0.6448 (7.1%)	5,80,200,800	0.6475 (6.1%)
		5,80,200,800	0.6358 (7.8%)	20,80,200, 800	0.6438 (6.9%)	20,80,200, 800	0.6464 (5.9%)
		5,20,200,800	0.6339 (7.5%)	5,80,400,800	0.6416 (6.6%)	5,20,200,800	0.6443 (5.6%)
	5 windows	20,80,200, 400,800	0.6363 (7.9%)	5,80,200,400, 800	0.6460 (7.3%)	20,80,200, 400,800	0.6450 (5.7%)
		5,80,200,400, 800	0.6355 (7.8%)	20,80,200, 400,800	0.6450 (7.1%)	5,80,200,400, 800	0.6441 (5.6%)
	6 windows	All windows	0.6360 (7.9%)	All windows	0.6463 (7.4%)	All windows	0.6524 (6.9%)

In Table 6, BMPQIS, TQIS, and LMQIS stand for BM25+PageRank, TF/IDF, and language model with query expansion, internal structure and with supporting documents, respectively.

Table 6: MAPs and MAP gains for BM25+PageRank+QE+IS, TF/IDF+QE+IS, and language model+QE+IS with different window combinations when consider supporting documents, and the highest MAP for a model is in bold and underlined

Base window	**Mode**	**BM25+PageRank+QE+IS (support)**		**TF/IDF+QE+IS (support)**		**Language model+QE+IS (support)**	
		Windows	**MAP (MAP gain)**	**Windows**	**MAP (MAP gain)**	**Windows**	**MAP (MAP gain)**
5	1 window	5	0.1841	5	0.1899	5	0.1904
	2 windows	5, 400	0.4180 (127.1%)	5, 400	0.4312 (127.1%)	5, 400	0.4317 (126.7%)
		5, 200	0.4124 (124.0%)	5, 200	0.4244 (123.5%)	5, 200	0.4261 (123.7%)
		5, 800	0.4118 (123.7%)	5, 800	0.4071 (114.4%)	5, 800	0.4163 (118.6%)
	3 windows	5, 200, 800	0.4316 (134.4%)	5, 80, 400	0.4457 (134.7%)	5, 80, 400	0.4426 (132.4%)
		5, 80, 800	0.4283 (132.6%)	5, 200, 400	0.4384 (130.9%)	5, 200, 400	0.4407 (131.4%)
		5, 400, 800	0.4280 (132.5%)	5, 200, 800	0.4374 (130.3%)	5, 200, 800	0.4398 (131.0%)
	4 windows	5,80,200,800	0.4425 (140.4%)	5,80,200,800	0.4517 (137.9%)	5,80,200,800	0.4532 (138.0%)
		5,20,200,800	0.4396 (138.8%)	5,20,80,400	0.4475 (135.7%)	5,20,80,400	0.4508 (136.7%)
		5, 20,80,800	0.4385 (138.2%)	5,20,200,800	0.4462 (135.0%)	5,20,200,800	0.4483 (135.5%)
	5 windows	5,20,80,200, 800	0.4431 (140.7%)	5,20,80,200, 800	0.4507 (137.3%)	5,20,80,200, 800	0.4524 (137.6%)
		5,80,200,400, 800	0.4412 (139.7%)	5,80,200,400, 800	0.4499 (137.0%)	5,80,200,400, 800	0.4516 (137.2%)
	6 windows	All windows	0.4438 (141.1%)	All windows	**<u>0.4527</u>** (138.4%)	All windows	**<u>0.4545</u>** (138.7%)
20	1 window	20	0.2716	20	0.2836	20	0.2835
	2 windows	20, 400	0.4223 (55.5%)	20, 400	0.4353 (53.5%)	20, 400	0.4357 (53.7%)
		20, 800	0.4200 (54.6%)	20, 200	0.4289 (51.2%)	20, 200	0.4296 (51.5%)
		20, 200	0.4181 (53.9%)	20, 800	0.4143 (46.1%)	20, 800	0.4193 (47.9%)
	3 windows	20, 400, 800	0.4328 (59.4%)	20, 80, 400	0.4462 (57.3%)	20, 80, 400	0.4483 (58.1%)
		20, 200, 800	0.4322 (59.1%)	20, 200, 800	0.4412 (55.6%)	20, 400, 800	0.4421 (56.0%)
		20, 80, 800	0.4317 (58.9%)	20, 200, 400	0.4398 (55.1%)	20, 200, 400	0.4414 (55.7%)
	4 windows	20,80,200, 800	**<u>0.4453</u>** (64.0%)	20,80,200, 800	0.4517 (59.3%)	20,80,200, 800	0.4529 (59.8%)
		5, 20,80,800	0.4385 (61.5%)	20,80,200, 400	0.4477 (57.9%)	20,80,200, 400	0.4492 (58.4%)
		20,200,400, 800	0.4379 (61.2%)	5,20,80,400	0.4475 (57.8%)	5, 20,80,800	0.4483 (58.1%)
	5 windows	20,80,200, 400,800	0.4431 (63.1%)	5,20,80,200, 800	0.4507 (59.0%)	5,20,80,200, 800	0.4524 (59.6%)
		5,20,80,200, 800	0.4431 (63.1%)	20,80,200, 400,800	0.4486 (58.2%)	20,80,200, 400,800	0.4480 (58.0%)
	6 windows	All windows	0.4438 (63.4%)	All windows	**<u>0.4527</u>** (60.0%)	All windows	**<u>0.4545</u>** (60.3%)
80	1 window	80	0.3579	80	0.3767	80	0.3778
	2 windows	80, 800	0.4253 (18.8%)	80, 400	0.4436 (17.8%)	80, 400	0.4448 (17.7%)
		80, 400	0.4235 (18.3%)	80, 800	0.4291 (13.9%)	80, 800	0.4336 (14.8%)

		80, 200	0.4095 (14.4%)	80, 200	0.4218 (12.0%)	80, 200	0.4312 (14.1%)
	3 windows	80, 200, 800	0.4339 (21.2%)	20, 80, 400	0.4462 (18.4%)	20, 80, 400	0.4483 (18.7%)
		20, 80, 800	0.4317 (20.6%)	5, 80, 400	0.4457 (18.3%)	5, 80, 400	0.4426 (17.2%)
		80, 400, 800	0.4311 (20.5%)	80, 200, 800	0.4452 (18.2%)	20, 80, 800	0.4421 (17.0%)
	4 windows	20,80,200, 800	**0.4453** (24.4%)	5,80,200,800	0.4517 (19.9%)	5,80,200,800	0.4532 (20.0%)
		5,80,200,800	0.4425 (23.6%)	20,80,200, 800	0.4517 (19.9%)	20,80,200, 800	0.4529 (19.9%)
		5, 20,80,800	0.4385 (22.5%)	20,80,200, 400	0.4477 (18.8%)	5, 20,80,800	0.4503 (19.2%)
	5 windows	20,80,200, 400,800	0.4431 (23.8%)	5,20,80,200, 800	0.4507 (19.6%)	5,20,80,200, 800	0.4524 (19.7%)
		5,20,80,200, 800	0.4431 (23.8%)	5,80,200,400, 800	0.4499 (19.4%)	20,80,200, 400,800	0.448 (18.6%)
	6 windows	All windows	0.4438 (24.0%)	All windows	**0.4527** (20.2%)	All windows	**0.4545** (20.3%)
200	1 window	200	0.3935	200	0.4065	200	0.4070
	2 windows	200, 800	0.4259 (8.2%)	200, 800	0.4337 (6.7%)	200, 800	0.4368 (7.3%)
		200, 400	0.4191 (6.5%)	200, 400	0.4308 (6.0%)	200, 400	0.4352 (6.9%)
		20, 200	0.4181 (6.3%)	20, 200	0.4289 (5.5%)	20, 200	0.4296 (5.6%)
	3 windows	80, 200, 800	0.4339 (10.3%)	80, 200, 800	0.4452 (9.5%)	80, 200, 800	0.4467 (9.8%)
		20, 200, 800	0.4322 (9.8%)	80, 200, 400	0.4436 (9.1%)	80, 200, 400	0.4428 (8.8%)
		5, 200, 800	0.4316 (9.7%)	20, 200, 800	0.4412 (8.5%)	20, 200, 800	0.4416 (8.5%)
	4 windows	20,80,200, 800	**0.4453** (13.2%)	5,80,200,800	0.4517 (11.1%)	5,80,200,800	0.4532 (11.4%)
		5,80,200,800	0.4425 (12.5%)	20,80,200, 800	0.4517 (11.1%)	20,80,200, 800	0.4529 (11.3%)
		5,20,200,800	0.4396 (11.7%)	20,80,200, 400	0.4477 (10.1%)	5,20,200,800	0.4483 (10.1%)
	5 windows	20,80,200, 400,800	0.4431 (12.6%)	5,20,80,200, 800	0.4507 (10.9%)	5,20,80,200, 800	0.4524 (11.2%)
		5,20,80,200, 800	0.4431 (12.6%)	5,80,200,400, 800	0.4499 (10.7%)	20,80,200, 400,800	0.448 (10.1%)
	6 windows	All windows	0.4438 (12.8%)	All windows	**0.4527** (11.4%)	All windows	**0.4545** (11.7%)
400	1 window	400	0.3886	400	0.4103	400	0.4107
	2 windows	80, 400	0.4235 (9.0%)	80, 400	0.4436 (8.1%)	80, 400	0.4448 (8.3%)
		20, 400	0.4223 (8.7%)	20, 400	0.4353 (6.1%)	20, 400	0.4357 (6.1%)
		400, 800	0.4209 (8.3%)	5, 400	0.4312 (5.1%)	5, 400	0.4317 (5.1%)
	3 windows	20, 400, 800	0.4328 (11.4%)	20, 80, 400	0.4462 (8.7%)	20, 80, 400	0.4483 (9.2%)
		80, 400, 800	0.4311 (10.9%)	5, 80, 400	0.4457 (8.6%)	5, 80, 400	0.4426 (7.8%)
		200, 400, 800	0.431 (10.9%)	80, 200, 400	0.4436 (8.1%)	20, 400, 800	0.4421 (7.6%)
	4 windows	20,200,400, 800	0.4379 (12.7%)	20,80,200, 400	0.4477 (9.1%)	5,20,80,400	0.4508 (9.8%)
		5,80,400,800	0.4367 (12.4%)	5,20,80,400	0.4475 (9.1%)	20,80,200, 400	0.4492 (9.4%)
		20,80,400, 800	0.4365 (12.3%)	20,200,400, 800	0.4472 (9.0%)	20,200,400,800	0.4477 (9.0%)
	5 windows	20,80,200, 400,800	0.4431 (14.0%)	5,80,200,400, 800	0.4499 (9.7%)	5,80,200,400, 800	0.4516 (10.0%)
		5,80,200,400, 800	0.4412 (13.5%)	20,80,200, 400,800	0.4486 (9.3%)	20,80,200, 400,800	0.448 (9.1%)
	6 windows	All windows	0.4438 (14.2%)	All windows	**0.4527** (10.3%)	All windows	**0.4545** (10.7%)
800	1 window	800	0.3833	800	0.3875	800	0.3921
	2 windows	200, 800	0.4259 (11.1%)	200, 800	0.4337 (11.9%)	200, 800	0.4368 (11.4%)
		80, 800	0.4253 (11.0%)	80, 800	0.4291 (10.7%)	80, 800	0.4336 (10.6%)
		400, 800	0.4209 (9.8%)	400, 800	0.4238 (9.4%)	400, 800	0.4321 (10.2%)
	3 windows	80, 200, 800	0.4339 (13.2%)	80, 200, 800	0.4452 (14.9%)	80, 200, 800	0.4467 (13.9%)
		20, 400, 800	0.4328 (12.9%)	20, 200, 800	0.4412 (13.9%)	20, 200, 800	0.4416 (12.6%)
		20, 200, 800	0.4322 (12.8%)	200, 400, 800	0.4406 (13.7%)	200, 400, 800	0.4411 (12.5%)
	4 windows	20,80,200, 800	**0.4453** (16.2%)	5,80,200,800	0.4517 (16.6%)	5,80,200,800	0.4532 (15.6%)
		5,80,200,800	0.4425 (15.4%)	20,80,200, 800	0.4517 (16.6%)	20,80,200, 800	0.4529 (15.5%)
		5,20,200,800	0.4396 (14.7%)	5,20,200,800	0.4462 (15.1%)	5,20,200,800	0.4483 (14.3%)
	5 windows	20,80,200, 400,800	0.4431 (15.6%)	5,20,80,200, 800	0.4507 (16.3%)	5,20,80,200, 800	0.4524 (15.4%)
		5,20,80,200, 800	0.4431 (15.6%)	5,80,200,400, 800	0.4499 (16.1%)	5,80,200,400, 800	0.4516 (15.2%)

	6 windows	All windows	0.4438 (15.8%)	All windows	**<u>0.4527</u>** (16.8%)	All windows	**<u>0.4545</u>** (15.9%)

In Table 7, BMPI, TI, and LMI stand for BM25+PageRank, TF/IDF, and language model with internal structure and without supporting documents, respectively.

Table 7: MAPs and MAP gains for BM25+PageRank+IS, TF/IDF+IS, and language model+IS with different window combinations when do not consider supporting documents, and the highest MAP for a model is in bold and underlined

Base window	**Mode**	**BM25+PageRank+IS**		**TF/IDF+IS**		**Language model+IS**	
		Windows	**MAP (MAP gain)**	**Windows**	**MAP (MAP gain)**	**Windows**	**MAP (MAP gain)**
5	1 window	5	0.2237	5	0.2273	5	0.2281
	2 windows	5, 800	0.5313 (137.5%)	5, 400	0.5476 (140.9%)	5, 400	0.5469 (139.8%)
		5, 400	0.529 (136.5%)	5, 200	0.5413 (138.1%)	5, 800	0.5411 (137.2%)
		5, 200	0.5149 (130.2%)	5, 800	0.5395 (137.4%)	5, 200	0.5393 (136.4%)
	3 windows	5, 80, 800	0.5424 (142.5%)	5, 200, 800	0.5589 (145.9%)	5, 200, 800	0.5563 (143.9%)
		5, 200, 800	0.5408 (141.8%)	5, 200, 400	0.5573 (145.2%)	5, 80, 400	0.5560 (143.7%)
		5, 400, 800	0.5386 (140.8%)	5, 80, 400	0.5563 (144.7%)	5, 200, 400	0.5540 (142.9%)
	4 windows	5,80,200,800	0.5495 (145.6%)	5,80,200,800	0.5650 (148.6%)	5,80,200,800	0.5654 (147.9%)
		5,80,400,800	0.5458 (144.0%)	5,80,200,400	0.5630 (147.7%)	5,80,200,400	0.5624 (146.6%)
		5,200,400,800	0.545 (143.6%)	5,20,200,800	0.5619 (147.2%)	5,80,400,800	0.5617 (146.2%)
	5 windows	5,80,200,400, 800	0.5507 (146.2%)	5,20,80,200, 800	**<u>0.5658</u>** (148.9%)	5,20,80,200, 800	0.5669 (148.5%)
		5,20,80,200, 800	0.5483 (145.1%)	5,20,80,200, 400	0.5629 (147.6%)	5,80,200,400, 800	0.5633 (147.0%)
	6 windows	All windows	0.5504 (146.0%)	All windows	0.5631 (147.7%)	All windows	0.5646 (147.5%)
20	1 window	20	0.3370	20	0.3450	20	0.3455
	2 windows	20, 800	0.5329 (58.1%)	20, 400	0.5497 (59.3%)	20, 400	0.5507 (59.4%)
		20, 400	0.5302 (57.3%)	20, 200	0.5448 (57.9%)	20, 800	0.5495 (59.0%)
		20, 200	0.5185 (53.9%)	20, 800	0.5447 (57.9%)	20, 200	0.5489 (58.9%)
	3 windows	20, 200, 800	0.5426 (61.0%)	20, 200, 800	0.5597 (62.2%)	20, 200, 800	0.5608 (62.3%)
		20, 80, 800	0.5422 (60.9%)	20, 80, 400	0.5581 (61.8%)	20, 80, 400	0.5597 (62.0%)
		20, 400, 800	0.5403 (60.3%)	20, 200, 400	0.5570 (61.4%)	20, 80, 800	0.5589 (61.8%)
	4 windows	20,80,200,800	0.5497 (63.1%)	20,80,200, 800	0.5654 (63.9%)	20,80,200, 800	**<u>0.5673</u>** (64.2%)
		20,200,400, 800	0.5474 (62.4%)	20,80,200, 400	0.5633 (63.3%)	20,80,200, 400	0.5648 (63.5%)
		20,80,400,800	0.546 (62.0%)	5,20,200,800	0.5619 (62.9%)	20,200,400, 800	0.5633 (63.0%)
	5 windows	20,80,200,400, 800	**<u>0.5508</u>** (63.4%)	5,20,80,200, 800	**<u>0.5658</u>** (64.0%)	5,20,80,200, 800	0.5669 (64.1%)
		5,20,80,200, 800	0.5483 (62.7%)	20,80,200, 400,800	0.5632 (63.2%)	20,80,200, 400,800	0.5647 (63.4%)
	6 windows	All windows	0.5504 (63.3%)	All windows	0.5631 (63.2%)	All windows	0.5646 (63.4%)
80	1 window	80	0.4756	80	0.4898	80	0.4905
	2 windows	80, 800	0.5436 (14.3%)	80, 400	0.5542 (13.1%)	80, 400	0.5552 (13.2%)
		80, 400	0.5362 (12.7%)	80, 800	0.5496 (12.2%)	80, 800	0.5501 (12.2%)
		80, 200	0.5212 (9.6%)	80, 200	0.5451 (11.3%)	80, 200	0.545 (11.1%)
	3 windows	80, 200, 800	0.5478 (15.2%)	80, 200, 800	0.5641 (15.2%)	80, 200, 800	0.5659 (15.4%)
		80, 400, 800	0.5452 (14.6%)	80, 200, 400	0.5637 (15.1%)	80, 200, 400	0.5647 (15.1%)
		5, 80, 800	0.5424 (14.0%)	20, 80, 400	0.5581 (13.9%)	80, 400, 800	0.559 (14.0%)
	4 windows	20,80,200,800	0.5497 (15.6%)	20,80,200, 800	0.5654 (15.4%)	20,80,200, 800	**<u>0.5673</u>** (15.7%)
		5,80,200,800	0.5495 (15.5%)	5,80,200,800	0.5650 (15.4%)	5,80,200,800	0.5654 (15.3%)
		20,80,400,800	0.546 (14.8%)	20,80,200, 400	0.5633 (15.0%)	20,80,200, 400	0.5648 (15.1%)
	5 windows	20,80,200,400, 800	**<u>0.5508</u>** (15.8%)	5,20,80,200, 800	**<u>0.5658</u>** (15.5%)	5,20,80,200, 800	0.5654 (15.3%)
		5,80,200,400, 800	0.5507 (15.8%)	20,80,200, 400,800	0.5632 (15.0%)	20,80,200, 400,800	0.5647 (15.1%)
	6 windows	All windows	0.5504 (15.7%)	All windows	0.5631 (15.0%)	All windows	0.5646 (15.1%)
200	1 window	200	0.5170	200	0.5376	200	0.5382
	2 windows	200, 800	0.5419 (4.8%)	200, 400	0.5579 (3.8%)	200, 800	0.5598 (4.0%)
		200, 400	0.5371 (3.9%)	200, 800	0.5578 (3.8%)	200, 400	0.5589 (3.8%)

		80, 200	0.5212 (0.8%)	80, 200	0.5451 (1.4%)	80, 200	0.545 (1.3%)
	3 windows	80, 200, 800	0.5478 (6.0%)	80, 200, 800	0.5641 (4.9%)	80, 200, 800	0.5659 (5.1%)
		200, 400, 800	0.544 (5.2%)	80, 200, 400	0.5637 (4.9%)	80, 200, 400	0.5647 (4.9%)
		20, 200, 800	0.5426 (5.0%)	20, 200, 800	0.5597 (4.1%)	20, 200, 800	0.5608 (4.2%)
	4 windows	20,80,200,800	0.5497 (6.3%)	20,80,200, 800	0.5654 (5.2%)	20,80,200, 800	**0.5673** (5.4%)
		5,80,200,800	0.5495 (6.3%)	5,80,200,800	0.5650 (5.1%)	5,80,200,800	0.5654 (5.1%)
		20,200,400, 800	0.5474 (5.9%)	20,80,200, 400	0.5633 (4.8%)	20,200,400, 800	0.5633 (4.7%)
	5 windows	20,80,200,400, 800	**0.5508** (6.5%)	5,20,80,200, 800	**0.5658** (5.2%)	5,20,80,200, 800	0.5669 (5.3%)
		5,80,200,400, 800	0.5507 (6.5%)	20,80,200, 400,800	0.5632 (4.8%)	20,80,200, 400,800	0.5647 (4.9%)
	6 windows	All windows	0.5504 (6.5%)	All windows	0.5631 (4.7%)	All windows	0.5646 (4.9%)
400	1 window	400	0.5223	400	0.5391	400	0.5394
	2 windows	200, 400	0.5371 (2.8%)	200, 400	0.5579 (3.5%)	200, 400	0.5589 (3.6%)
		80, 400	0.5362 (2.7%)	80, 400	0.5542 (2.8%)	80, 400	0.5552 (2.9%)
		400, 800	0.5348 (2.4%)	20, 400	0.5497 (2.0%)	400, 800	0.5511 (2.2%)
	3 windows	80, 400, 800	0.5452 (4.4%)	80, 200, 400	0.5637 (4.6%)	80, 200, 400	0.5647 (4.7%)
		200, 400, 800	0.544 (4.2%)	20, 80, 400	0.5581 (3.5%)	20, 80, 400	0.5597 (3.8%)
		20, 400, 800	0.5403 (3.4%)	5, 200, 400	0.5573 (3.4%)	80, 400, 800	0.559 (3.6%)
	4 windows	20,200,400, 800	0.5474 (4.8%)	20,80,200, 400	0.5633 (4.5%)	20,80,200, 400	0.5648 (4.7%)
		20,80,400,800	0.546 (4.5%)	5,80,200,400	0.5630 (4.4%)	20,200,400, 800	0.5633 (4.4%)
		5,80,400,800	0.5458 (4.5%)	20,200,400, 800	0.5602 (3.9%)	5,80,200,400	0.5624 (4.3%)
	5 windows	20,80,200,400, 800	**0.5508** (5.5%)	20,80,200, 400,800	0.5632 (4.5%)	20,80,200, 400,800	0.5647 (4.7%)
		5,80,200,400, 800	0.5507 (5.4%)	5,20,80,200, 400	0.5629 (4.4%)	5,20,80,200, 400	0.5637 (4.5%)
	6 windows	All windows	0.5504 (5.4%)	All windows	0.5631 (4.5%)	All windows	0.5646 (4.7%)
800	1 window	800	0.5239	800	0.5333	800	0.5339
	2 windows	80, 800	0.5436 (3.8%)	200, 800	0.5578 (4.6%)	200, 800	0.5598 (4.9%)
		200, 800	0.5419 (3.4%)	80, 800	0.5496 (3.1%)	400, 800	0.5511 (3.2%)
		400, 800	0.5348 (2.1%)	400, 800	0.5496 (3.1%)	80, 800	0.5501 (3.0%)
	3 windows	80, 200, 800	0.5478 (4.6%)	80, 200, 800	0.5641 (5.8%)	80, 200, 800	0.5659 (6.0%)
		80, 400, 800	0.5452 (4.1%)	20, 200, 800	0.5597 (5.0%)	20, 200, 800	0.5608 (5.0%)
		200, 400, 800	0.544 (3.8%)	5, 200, 800	0.5589 (4.8%)	80, 400, 800	0.559 (4.7%)
	4 windows	20,80,200,800	0.5497 (4.9%)	20,80,200, 800	0.5654 (6.0%)	20,80,200, 800	**0.5673** (6.3%)
		5,80,200,800	0.5495 (4.9%)	5,80,200,800	0.5650 (5.9%)	5,80,200,800	0.5654 (5.9%)
		20,200,400, 800	0.5474 (4.5%)	5,20,200,800	0.5619 (5.4%)	20,200,400, 800	0.5633 (5.5%)
	5 windows	20,80,200,400, 800	**0.5508** (5.1%)	5,20,80,200, 800	**0.5658** (6.1%)	5,20,80,200, 800	0.5669 (6.2%)
		5,80,200,400, 800	0.5507 (5.1%)	20,80,200, 400,800	0.5632 (5.6%)	20,80,200, 400,800	0.5647 (5.8%)
	6 windows	All windows	0.5504 (5.1%)	All windows	0.5631 (5.6%)	All windows	0.5661 (6.0%)

In Table 8, BMPIS, TIS, and LMIS stand for BM25+PageRank, TF/IDF, and language model with internal structure and with supporting documents, respectively.

Table 8: MAPs and MAP gains for BM25+PageRank+IS, TF/IDF+IS, and language model+IS with different window combinations when consider supporting documents, and the highest MAP for a model is in bold and underlined

Base window	**Mode**	**BM25+PageRank+IS (support)**		**TF/IDF+IS (support)**		**Language model+IS (support)**	
		Windows	**MAP (MAP gain)**	**Windows**	**MAP (MAP gain)**	**Windows**	**MAP (MAP gain)**
5	1 window	5	0.1173	5	0.1251	5	0.1247
	2 windows	5, 800	0.3195 (172.4%)	5, 800	0.3187 (154.8%)	5, 800	0.3260 (161.4%)
		5, 400	0.3032 (158.5%)	5, 400	0.3181 (154.3%)	5, 400	0.3249 (160.5%)
		5, 200	0.2838 (141.9%)	5, 200	0.3179 (154.1%)	5, 200	0.3203 (156.9%)
	3 windows	5, 80, 800	0.3297 (181.1%)	5, 200, 800	0.3383 (170.4%)	5, 200, 800	0.3398 (172.5%)

		5, 400, 800	0.3287 (180.2%)	5, 400, 800	0.3331 (166.3%)	5, 80, 800	0.3311 (165.5%)
		5, 200, 800	0.3267 (178.5%)	5, 80, 800	0.3329 (166.1%)	5, 400, 800	0.3283 (163.3%)
	4 windows	5,80,200,800	0.3345 (185.2%)	5,80,200,800	0.3475 (177.8%)	5,80,200,800	0.3499 (180.6%)
		5,20,400,800	0.3344 (185.1%)	5,20,200,800	0.3438 (174.8%)	5,20,200,800	0.3438 (175.7%)
		5, 20, 80, 800	0.3329 (183.8%)	5,20,80,800	0.3416 (173.1%)	5,20,400,800	0.3400 (172.7%)
	5 windows	5,20,80,200, 800	**<u>0.3359</u>** (186.4%)	5,20,80,200, 800	**<u>0.3489</u>** (178.9%)	5,20,80,200, 800	0.3503 (180.9%)
		5,20,80,400, 800	0.3347 (185.3%)	5, 80, 200, 400, 800	0.3445 (175.4%)	5,20,80,400, 800	0.3434 (175.4%)
	6 windows	All windows	0.3343 (185.0%)	All windows	0.3454 (176.1%)	All windows	0.3465 (177.8%)
20	1 window	20	0.1762	20	0.1936	20	0.1942
	2 windows	20, 800	0.3233 (83.5%)	20, 800	0.3267 (68.8%)	20, 800	0.3275 (68.6%)
		20, 400	0.3027 (71.8%)	20, 200	0.3221 (66.4%)	20, 400	0.3263 (68.0%)
		20, 200	0.284 (61.2%)	20, 400	0.3212 (65.9%)	20, 200	0.3245 (67.1%)
	3 windows	20, 80, 800	0.3307 (87.7%)	20, 200, 800	0.3398 (75.5%)	20, 80, 800	0.3411 (75.6%)
		20, 400, 800	0.3306 (87.6%)	20, 80, 800	0.3396 (75.4%)	20, 200, 800	0.3405 (75.3%)
		20, 200, 800	0.3303 (87.5%)	20, 400, 800	0.3353 (73.2%)	20, 400, 800	0.3333 (71.6%)
	4 windows	20,80,200,80 0	0.3354 (90.4%)	20,80,200, 800	0.3481 (79.8%)	20,80,200, 800	0.3479 (79.1%)
		20,80,400,80 0	0.3346 (89.9%)	5,20,200,800	0.3438 (77.6%)	5,20,200,800	0.3438 (77.0%)
		5,20,400,800	0.3344 (89.8%)	20, 200, 400 800	0.3417 (76.5%)	20,80,400,800	0.3423 (76.3%)
	5 windows	5,20,80,200, 800	**<u>0.3359</u>** (90.6%)	5,20,80,200, 800	**<u>0.3489</u>** (80.2%)	20,80,200, 400,800	**<u>0.3505</u>** (80.5%)
		5,20,80,400, 800	0.3347 (90.0%)	20,80,200, 400,800	0.3460 (78.7%)	5,20,80,200, 800	0.3503 (80.4%)
	6 windows	All windows	0.3343 (89.7%)	All windows	0.3454 (78.4%)	All windows	0.3465 (78.4%)
80	1 window	80	0.2380	80	0.2548	80	0.2557
	2 windows	80, 800	0.3285 (38.0%)	80, 800	0.3325 (30.5%)	80, 800	0.3349 (31.0%)
		80, 400	0.3057 (28.4%)	80, 400	0.3247 (27.4%)	80, 400	0.3287 (28.5%)
		80, 200	0.2812 (18.2%)	80, 200	0.3156 (23.9%)	80, 200	0.3207 (25.4%)
	3 windows	80, 200, 800	0.3324 (39.7%)	80, 200, 800	0.3457 (35.7%)	80, 200, 800	0.3435 (34.3%)
		20, 80, 800	0.3307 (38.9%)	20, 80, 800	0.3396 (33.3%)	20, 80, 800	0.3411 (33.4%)
		80, 400, 800	0.3305 (38.9%)	80, 400, 800	0.3377 (32.5%)	80, 400, 800	0.3407 (33.2%)
	4 windows	20,80,200,80 0	0.3354 (40.9%)	20,80,200, 800	0.3481 (36.6%)	5,80,200,800	0.3499 (36.8%)
		20,80,400,80 0	0.3346 (40.6%)	5,80,200,800	0.3475 (36.4%)	20,80,200, 800	0.3479 (36.1%)
		5,80,200,800	0.3345 (40.5%)	20,80,400, 800	0.3416 (34.1%)	20,80,400, 800	0.3423 (33.9%)
	5 windows	5,20,80,200, 800	**<u>0.3359</u>** (41.1%)	5,20,80,200, 800	**<u>0.3489</u>** (36.9%)	20,80,200, 400,800	**<u>0.3505</u>** (37.1%)
		5,20,80,400, 800	0.3347 (40.6%)	20,80,200, 400,800	0.3460 (35.8%)	5,20,80,200, 800	0.3503 (37.0%)
	6 windows	All windows	0.3343 (40.5%)	All windows	0.3454 (35.6%)	All windows	0.3465 (35.5%)
200	1 window	200	0.2749	200	0.3053	200	0.3104
	2 windows	200, 800	0.3261 (18.6%)	200, 800	0.3373 (10.5%)	200, 800	0.3379 (8.9%)
		200, 400	0.3049 (10.9%)	200, 400	0.3284 (7.6%)	200, 400	0.3295 (6.2%)
		20, 200	0.284 (3.3%)	20, 200	0.3221 (5.5%)	20, 200	0.3245 (4.5%)
	3 windows	80, 200, 800	0.3324 (20.9%)	80, 200, 800	0.3457 (13.2%)	80, 200, 800	0.3435 (10.7%)
		20, 200, 800	0.3303 (20.2%)	20, 200, 800	0.3398 (11.3%)	20, 200, 800	0.3405 (9.7%)
		5, 200, 800	0.3267 (18.8%)	5, 200, 800	0.3383 (10.8%)	5, 200, 800	0.3398 (9.5%)
	4 windows	20,80,200,80 0	0.3354 (22.0%)	20,80,200, 800	0.3481 (14.0%)	5,80,200,800	0.3499 (12.7%)
		5,80,200,800	0.3345 (21.7%)	5,80,200,800	0.3475 (13.8%)	20,80,200, 800	0.3479 (12.1%)
		5,20,200,800	0.3324 (20.9%)	5,20,200,800	0.3438 (12.6%)	5,20,200,800	0.3438 (10.8%)
	5 windows	5,20,80,200, 800	**<u>0.3359</u>** (22.2%)	5,20,80,200, 800	**<u>0.3489</u>** (14.3%)	20,80,200, 400,800	**<u>0.3505</u>** (12.9%)
		20,80,200,40 0, 800	0.3345 (21.7%)	20,80,200, 400,800	0.3460 (13.3%)	5,20,80,200, 800	0.3503 (12.9%)
	6 windows	All windows	0.3343 (21.6%)	All windows	0.3454 (13.1%)	All windows	0.3465 (11.6%)
400	1 window	400	0.2887	400	0.3041	400	0.3055
	2 windows	400, 800	0.3211 (11.2%)	200, 400	0.3284 (8.0%)	400, 800	0.3302 (8.1%)
		80, 400	0.3057 (5.9%)	400, 800	0.3267 (7.4%)	200, 400	0.3295 (7.9%)
		200, 400	0.3049 (5.6%)	80, 400	0.3247 (6.8%)	80, 400	0.3287 (7.6%)

	3 windows	20, 400, 800	0.3306 (14.5%)	80, 400, 800	0.3377 (11.0%)	80, 400, 800	0.3407 (11.5%)
		80, 400, 800	0.3305 (14.5%)	200, 400, 800	0.3365 (10.7%)	80, 200, 400	0.3382 (10.7%)
		5, 400, 800	0.3287 (13.9%)	80, 200, 400	0.3361 (10.5%)	200, 400, 800	0.3356 (9.9%)
	4 windows	20,80,400,800	0.3346 (15.9%)	20, 200, 400 800	0.3417 (12.4%)	20, 200, 400 800	0.3435 (12.4%)
		5,20,400,800	0.3344 (15.8%)	20,80,400, 800	0.3416 (12.3%)	5,20,80,400	0.3427 (12.2%)
		20,200,400, 800	0.3323 (15.1%)	5,20,80,400	0.3408 (12.1%)	20,80,400, 800	0.3423 (12.0%)
	5 windows	5,20,80,400, 800	0.3347 (15.9%)	20,80,200, 400,800	0.3460 (13.8%)	20,80,200, 400,800	**<u>0.3505</u>** (14.7%)
		20,80,200,400, 800	0.3345 (15.9%)	5, 80, 200, 400, 800	0.3445 (13.3%)	5,20,80,400, 800	0.3434 (12.4%)
	6 windows	All windows	0.3343 (15.8%)	All windows	0.3454 (13.6%)	All windows	0.3465 (13.4%)
800	1 window	800	0.3038	800	0.3056	800	0.3067
	2 windows	80, 800	0.3285 (8.1%)	200, 800	0.3373 (10.4%)	200, 800	0.3379 (10.2%)
		200, 800	0.3261 (7.3%)	80, 800	0.3325 (8.8%)	80, 800	0.3349 (9.2%)
		20, 800	0.3233 (6.4%)	400, 800	0.3267 (6.9%)	400, 800	0.3302 (7.7%)
	3 windows	80, 200, 800	0.3324 (9.4%)	80, 200, 800	0.3457 (13.1%)	80, 200, 800	0.3435 (12.0%)
		20, 80, 800	0.3307 (8.9%)	20, 200, 800	0.3398 (11.2%)	20, 80, 800	0.3411 (11.2%)
		20, 400, 800	0.3306 (8.8%)	20, 80, 800	0.3396 (11.1%)	20, 200, 800	0.3405 (11.0%)
	4 windows	20,80,200,800	0.3354 (10.4%)	20,80,200, 800	0.3481 (13.9%)	5,80,200,800	0.3499 (14.1%)
		20,80,400,800	0.3346 (10.1%)	5,80,200,800	0.3475 (13.7%)	20,80,200, 800	0.3479 (13.4%)
		5,80,200,800	0.3345 (10.1%)	5,20,200,800	0.3438 (12.5%)	5,20,200,800	0.3438 (12.1%)
	5 windows	5,20,80,200, 800	**<u>0.3359</u>** (10.6%)	5,20,80,200, 800	**<u>0.3489</u>** (14.2%)	20,80,200, 400,800	**<u>0.3505</u>** (14.3%)
		5,20,80,400, 800	0.3347 (10.2%)	20,80,200, 400,800	0.3460 (13.2%)	5,20,80,200, 800	0.3503 (14.2%)
	6 windows	All windows	0.3343 (10.0%)	All windows	0.3454 (13.0%)	All windows	0.3465 (13.0%)

5.2.5 Effects of test collections

In order to gain an impression of the effect of test collection, our multiple window based approach was tested on the TREC2005 test collection using the same parameters that we used on the TREC2006 test collection. We ran three models based approaches on the TREC2005 test collection without query expansion, i.e., using the title only. The results of three models, i.e., BM25+PageRank+IS (BMPI), TF/IDF+IS (TI), and language model+IS (LMI), using different window combinations (supporting documents are not considered in TREC2005 collection) are reported in Table 9.

Table 9 shows that our multiple-window-based approach is effective for the TREC2005 collection as well, and we have similar findings as we already have on the TREC2006 collection: Multiple windows outperform single windows in terms of all three models. All of the highest MAPs for a particular model are produced by five window combinations. Certain window combinations consistently perform better than other combinations regardless of the model used. Both language model and TF/IDF based multiple windows outperform BM25 based multiple windows.

Although we did not take advantage of any domain-specific knowledge (Craswell et al. 2006), we still got very competitive results compared with the others. We think that the nature of TREC2005 and TREC2006 collections differ. The former is based on ground-truth and the latter is based on user judgments, and the former may not fully reflect the reality on the ground.

Table 9: MAPs and MAP gains for BM25+PageRank+IS (BMPI), TF/IDF+IS (TI), and language model+IS (LMI) with different window combinations on TREC2005 collection, and the highest MAP for a model is in bold and underlined

Base window	Mode	BM25+PageRank+IS		TF/IDF+IS		Language model+IS	
		Windows	MAP (MAP gain)	Windows	MAP (MAP gain)	Windows	MAP (MAP gain)
5	1 window	5	0.1415	5	0.1465	5	0.1468
	2 windows	5, 80	0.2076 (46.7%)	5, 80	0.2109 (44.0%)	5, 80	0.2106 (43.5%)
		5, 200	0.2044 (44.5%)	5, 400	0.2100 (43.3%)	5, 400	0.2074 (41.3%)
		5, 400	0.2032 (43.6%)	5, 200	0.2084 (42.3%)	5, 200	0.2068 (40.9%)
	3 windows	5, 20, 80	0.2083 (47.2%)	5, 80, 200	0.2131 (45.5%)	5, 80, 200	0.2125 (44.8%)
		5, 80, 200	0.2076 (46.7%)	5, 80, 400	0.2124 (45.0%)	5, 20, 80	0.2117 (44.2%)
		5, 80, 400	0.207 (46.3%)	5, 20, 80	0.2123 (44.9%)	5, 80, 400	0.2093 (42.6%)
	4 windows	5,20,80,400	0.2084 (47.3%)	5,20,80,400	0.2145 (46.4%)	5,20,80,400	0.2142 (45.9%)
		5,20,80,200	0.2075 (46.6%)	5,80,200,400	0.2142 (46.2%)	5,80,200,400	0.2140 (45.8%)
		5,80,200,400	0.2074 (46.6%)	5,20,200,400	0.2119 (44.6%)	5,20,80,200	0.2119 (44.3%)
	5 windows	5,20,80,200, 400	<u>**0.2092**</u> (47.8%)	5, 20, 80, 200, 400	<u>**0.2150**</u> (46.8%)	5, 20, 80, 200, 400	<u>**0.2147**</u> (46.2%)
		5,20,80,200, 800	0.2075 (46.6%)	5, 20, 80, 200, 800	0.2089 (42.6%)	5, 20, 80, 200, 800	0.2088 (42.2%)
	6 windows	All windows	0.2068 (46.1%)	All windows	0.2111 (44.1%)	All windows	0.2107 (43.6%)
20	1 window	20	0.1716	20	0.1774	20	0.1782
	2 windows	20, 80	0.2072 (20.7%)	20, 80	0.2112 (19.1%)	20, 80	0.2125 (19.2%)
		20, 200	0.2025 (18.0%)	20, 400	0.2097 (18.2%)	20, 400	0.2119 (18.9%)
		20, 400	0.2022 (17.8%)	20, 200	0.2086 (17.6%)	20, 200	0.2097 (17.7%)
	3 windows	5, 20, 80	0.2083 (21.4%)	20, 80, 400	0.2128 (20.0%)	20, 80, 400	0.2124 (19.2%)
		20, 80, 200	0.2077 (21.0%)	20, 80, 200	0.2124 (19.7%)	5, 20, 80	0.2117 (18.8%)
		20, 80, 400	0.2066 (20.4%)	5, 20, 80	0.2123 (19.7%)	20, 80, 200	0.2114 (18.6%)
	4 windows	20,80,200, 400	0.2087 (21.6%)	5,20,80,400	0.2145 (20.9%)	5,20,80,400	0.2142 (20.2%)
		5,20,80,400	0.2084 (21.4%)	20,80,200, 400	0.2143 (20.8%)	5,20,200,400	0.2136 (19.9%)
		5,20,80,200	0.2075 (20.9%)	5,20,200,400	0.2119 (19.4%)	20,80,200, 400	0.2123 (19.1%)
	5 windows	5,20,80,200, 400	<u>**0.2092**</u> (21.9%)	5, 20, 80, 200, 400	<u>**0.2150**</u> (21.2%)	5, 20, 80, 200, 400	<u>**0.2147**</u> (20.5%)
		5,20,80,200, 800	0.2075 (20.9%)	20, 80, 200, 400, 800	0.2091 (17.9%)	20, 80, 200, 400, 800	0.2113 (18.6%)
	6 windows	All windows	0.2068 (20.5%)	All windows	0.2111 (19.0%)	All windows	0.2107 (18.2%)
80	1 window	80	0.2023	80	0.2074	80	0.2083
	2 windows	80, 200	0.2082 (2.9%)	80, 200	0.2135 (2.9%)	80, 200	0.2128 (2.2%)
		5, 80	0.2076 (2.6%)	80, 400	0.2126 (2.5%)	20, 80	0.2125 (2.0%)
		80, 400	0.2068 (2.2%)	20, 80	0.2112 (1.8%)	80, 400	0.2122 (1.9%)
	3 windows	5, 20, 80	0.2083 (3.0%)	80,200,400	0.2133 (2.8%)	5, 80, 200	0.2125 (2.0%)
		20, 80, 200	0.2077 (2.7%)	5, 80, 200	0.2131 (2.7%)	80,200,400	0.2125 (2.0%)
		5, 80, 200	0.2076 (2.6%)	20, 80, 400	0.2128 (2.6%)	20, 80, 400	0.2124 (2.0%)
	4 windows	20,80,200, 400	0.2087 (3.2%)	5,20,80,400	0.2145 (3.4%)	5,20,80,400	0.2142 (2.8%)
		5,20,80,400	0.2084 (3.0%)	20,80,200, 400	0.2143 (3.3%)	5,80,200,400	0.214 (2.7%)
		5,20,80,200	0.2075 (2.6%)	5,80,200,400	0.2142 (3.3%)	20,80,200, 400	0.2123 (1.9%)
	5 windows	5,20,80,200, 400	<u>**0.2092**</u> (3.4%)	5, 20, 80, 200, 400	<u>**0.2150**</u> (3.7%)	5, 20, 80, 200, 400	<u>**0.2147**</u> (3.1%)
		5,20,80,200, 800	0.2075 (2.6%)	20, 80, 200, 400, 800	0.2091 (0.8%)	20, 80, 200, 400, 800	0.2113 (1.4%)
	6 windows	All windows	0.2068 (2.2%)	All windows	0.2111 (1.8%)	All windows	0.2107 (1.2%)
200	1 window	200	0.1987	200	0.2056	200	0.2071
	2 windows	80, 200	0.2082 (4.8%)	80, 200	0.2135 (3.8%)	80, 200	0.2128 (2.8%)
		200, 400	0.2054 (3.4%)	200, 400	0.2096 (1.9%)	200, 400	0.2105 (1.6%)
		20, 200	0.2048 (3.1%)	20, 200	0.2086 (1.5%)	20, 200	0.2097 (1.3%)
	3 windows	20, 80, 200	0.2077 (4.5%)	80,200,400	0.2133 (3.7%)	5, 80, 200	0.2125 (2.6%)
		5, 80, 200	0.2076 (4.5%)	5, 80, 200	0.2131 (3.6%)	80,200,400	0.2125 (2.6%)
		5, 200, 400	0.2042 (2.8%)	20, 80, 200	0.2124 (3.3%)	20, 80, 200	0.2114 (2.1%)

	4 windows	20,80,200, 400	0.2087 (5.0%)	20, 80, 200, 400	0.2143 (4.2%)	5,80,200,400	0.214 (3.3%)
		5,20,80,200	0.2075 (4.4%)	5,80,200,400	0.2142 (4.2%)	20,80,200, 400	0.2123 (2.5%)
		5,80,200,400	0.2074 (4.4%)	5,20,200,400	0.2119 (3.1%)	5,20,80,200	0.2119 (2.3%)
	5 windows	5,20,80,200, 400	**0.2092** (5.3%)	5, 20, 80, 200, 400	**0.2150** (4.6%)	5, 20, 80, 200, 400	**0.2147** (3.7%)
		5,20,80,200, 800	0.2075 (4.4%)	20, 80, 200, 400, 800	0.2091 (1.7%)	20, 80, 200, 400, 800	0.2113 (2.0%)
	6 windows	All windows	0.2068 (4.1%)	All windows	0.2111 (2.7%)	All windows	0.2107 (1.7%)
400	1 window	400	0.1973	400	0.2030	400	0.2038
	2 windows	80, 400	0.2068 (4.8%)	80, 400	0.2126 (4.7%)	80, 400	0.2122 (4.1%)
		200, 400	0.2034 (3.1%)	5, 400	0.2100 (3.4%)	20, 400	0.2119 (4.0%)
		20, 400	0.2032 (3.0%)	20, 400	0.2097 (3.3%)	5, 400	0.2074 (1.8%)
	3 windows	5, 80, 400	0.207 (4.9%)	80,200,400	0.2133 (5.1%)	80,200,400	0.2125 (4.3%)
		20, 80, 400	0.2066 (4.7%)	20, 80, 400	0.2128 (4.7%)	20, 80, 400	0.2124 (4.2%)
		5, 200, 400	0.2042 (3.5%)	5, 80, 400	0.2124 (4.6%)	5, 80, 400	0.2093 (2.7%)
	4 windows	20,80,200, 400	0.2087 (5.8%)	5,20,80,400	0.2145 (5.7%)	5,20,80,400	0.2142 (5.1%)
		5,20,80,400	0.2084 (5.6%)	20, 80, 200, 400	0.2143 (5.6%)	5,80,200,400	0.214 (5.0%)
		5,20,80,200	0.2075 (5.2%)	5,80,200,400	0.2142 (5.5%)	20,80,200, 400	0.2123 (4.2%)
	5 windows	5,20,80,200, 400	**0.2092** (6.0%)	5, 20, 80, 200, 400	**0.2150** (5.9%)	5, 20, 80, 200, 400	**0.2147** (5.3%)
		20,80,200, 400,800	0.1998 (1.3%)	20, 80, 200, 400, 800	0.2091 (3.0%)	20, 80, 200, 400, 800	0.2113 (3.7%)
	6 windows	All windows	0.2068 (4.8%)	All windows	0.2111 (4.0%)	All windows	0.2107 (3.4%)
800	1 window	800	0.1843	800	0.1903	800	0.1918
	2 windows	80, 800	0.1955 (6.1%)	80, 800	0.2017 (6.0%)	80, 800	0.2042 (6.5%)
		200, 800	0.1941 (5.3%)	200, 800	0.2007 (5.5%)	200, 800	0.2035 (6.1%)
		400, 800	0.1934 (4.9%)	400, 800	0.1989 (4.5%)	400, 800	0.2015 (5.1%)
	3 windows	80, 200, 800	0.1975 (7.2%)	80,200,800	0.2055 (8.0%)	80,400,800	0.2102 (9.6%)
		20, 80, 800	0.1972 (7.0%)	80,400,800	0.2054 (7.9%)	80,200,800	0.2087 (8.8%)
		200, 400, 800	0.1964 (6.6%)	200,400, 800	0.2036 (7.0%)	200,400, 800	0.2067 (7.8%)
	4 windows	20, 80, 200, 800	0.2007 (8.9%)	20, 80, 200, 800	0.2084 (9.5%)	80, 200, 400, 800	0.2105 (9.7%)
		5,80,200,800	0.2003 (8.7%)	80, 200, 400, 800	0.2071 (8.8%)	20, 80, 200, 800	0.2095 (9.2%)
		5,20,80,800	0.1996 (8.3%)	5,80,200,800	0.207 (8.8%)	5,80,200,800	0.2088 (8.9%)
	5 windows	5,20,80,200, 800	0.2075 (12.6%)	20, 80, 200, 400, 800	0.2091 (9.9%)	20, 80, 200, 400, 800	0.2113 (10.2%)
		20,80,200, 400,800	0.2027 (10.0%)	5, 20, 80, 200, 800	0.2089 (9.8%)	5, 20, 80, 200, 800	0.2087 (8.8%)
	6 windows	All windows	0.2068 (12.2%)	All windows	0.2111 (10.9%)	All windows	0.2107 (9.9%)

6. CONCLUSIONS AND FUTURE WORK

We propose a novel approach to expert finding by integrating multiple document features and query expansion in a two-stage model. Our novel approach of integrating a query expansion technique (Song & Bruza 2003) in our expert finding approach results in significant improvement over a title only expert finding approach. Document internal structure largely helps improve expert finding performance. PageRank does not significantly help improve performance, and we will carry out further research to study the effect of document authority. Our large-scale experiments on the TREC2006 test collection show that our multiple window based approach has greatly improved the expert finding performance over three traditional IR models as document relevance models combined with the fix-sized-window-based approach. Language model and TF/IDF based approaches outperform a BM25 based approach for expert finding. Our multiple window based approach integrating document internal structure and query expansion has achieved outstanding results, even outperforming the best runs in the TREC2006 expert search task. Our experiments on the TREC2005 expert search test collection further show the effectiveness of our approach to other test collections.

Furthermore, we have applied our approach to the TREC2007 expert search test collection showing the effectiveness of our multiple-window based approach on the CSIRO (Australian Commonwealth Scientific and Research Organization) dataset (Zhu et al. 2008a). To test the effectiveness of our approach to generic entity search (expert is only one type of entity), we have applied our expert finding approach to generic entity search in the INEX 2007 Entity Ranking Track[13] (Zhu et al. 2008b). Expert finding can have many applications in a real world environment, in order to test the effectiveness of our approach to real organizations, we have developed an expert finding prototype for the Open University[14], which has attracted lots of real users and has effectively helped their expertise search tasks.

In the future, we will carry out the following work:

1. Research into how to better integrate document authority in our approach, e.g., to study the transformation functions and their parameters for integrating PageRank.
2. Formally investigate what form of document relevance model will lead to good results in expert finding.
3. Investigate more efficient and automatic methods for window combination optimization.
4. Investigate how to extend our approach for integrating multiple sources of information such as domain knowledge etc in expert finding.
5. Evaluate our approach on other datasets such as a crawl of the Tilburg University's intranet (Balog et al. 2007a) and our own university's intranet dataset.
6. Study how to extend our expert finding approach to generic entity search, and how to generalize our approach to more structured data by experimenting our approach on the INEX2007 Entity Ranking Track's Wikipedia dataset, which contains XML files with rich structural information.
7. Study how to extend expert finding to automatic discovery of expert association networks, which consist of their related entities, such as their colleagues, their other expertise, and projects they are involved in etc. Association networks can help users better understand these experts' background.

---

[13] http://inex.is.informatik.uni-duisburg.de/2007/xmlSearch.html

[14] http://library.open.ac.uk/research/findexp/index.cfm

7. Acknowledgements

We would like to thank the anonymous referees for their constructive comments that have greatly helped to improve the article.

8. REFERENCES

Aho, A.V., & Corasick, M.J. (1975) Efficient string matching: An aid to bibliographic search. Comm. of the ACM 18(6):333–340.

Amati, G., & van Rijsbergen, C. J. (2002) Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Trans. on Inf. Syst. 20(4): 357-389.

Bailey, P., Craswell, N., de Vries, A. P., & Soboroff, I. (2007) Overview of the TREC 2007 Enterprise Track (DRAFT). In Proc. of TREC 2007 Notebook.

Balog, K., Azzopardi, L., & de Rijke, M. (2006) Formal models for expert finding in enterprise corpora. In Proc. of SIGIR, 43-50.

Balog, K., Bogers, T., Azzopardi, L., de Rijke, M., & van den Bosch, A. (2007a) Broad expertise retrieval in sparse data environments. In Proc. of SIGIR 2007, 551-558.

Balog, K., and de Rijke, M. (2007b) Determining Expert Profiles (With an Application to Expert Finding). In Proc. of IJCAI 2007, 2657-2662.

Brin, S., and Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks 30(1-7): 107-117.

Burgess, C., Livesay, K., & Lund, K. (1998) Explorations in context space: words, sentences, discourse. Discourse Processes, 25(2&3):211-257.

Campbell, C. S., Maglio, P. P., Cozzi, A., & Dom, B. (2003) Expertise identification using email communications. In Proc. of ACM Conference on Information and Knowledge Management (CIKM).

Cao, Y., Liu, J., Bao, S., & Li, H. (2006) Research on Expert Search at Enterprise Track of TREC 2005. In Proc. of TREC 2005.

Charikar, M., Indyk, P., & Panigrahy, R. (2002) New Algorithms for Subset Query, Partial Match, Orthogonal Range Searching and Related Problems. In Proc. of the 29th International Colloquium on Automata Languages and Programming: 451 - 462

Chen, H., Shen, H., Xiong, J., Tan, S., and Cheng, X. (2007) Social Network Structure behind the Mailing Lists: ICT-IIIS at TREC 2006 Expert Finding Track. In Proc. of TREC 2006.

Cheng, T., Yan, X., and Chang, K. C-C (2007) EntityRank: Searching Entities Directly and Holistically. In Proc. of VLDB 2007: 387-398

Chu-Carroll, J., Averboch, G., Duboue, P., Gondek, D., Murdock, J.W., Prager, J., Hoffmann, P., & Wiebe, J. (2007) IBM in TREC 2006 Enterprise Track. In Proc. of the Fifteenth Text REtrieval Conference (TREC 2006), Gaithersburg, Maryland USA.

Ciravegna, F. (2001) Adaptive Information Extraction from Text by Rule Induction and Generalisation. In Proc. of IJCAI 2001.

Conrad, J.G., & Utt, M.H. (1994) A System for Discovering Relationships by Feature Extraction from Text Databases. In Proc. of SIGIR 1994: 260-270.

Craswell, N., Robertson, S. E., Zaragoza, H., & Taylor, M. J. (2005) Relevance weighting for query independent evidence. In Proc. of SIGIR 2005: 416-423.

Craswell, N., de Vries, A.P., & Soboroff, I. (2006) Overview of the TREC-2005 Enterprise Track. In Proc. of The Fourteenth Text REtrieval Conference (TREC 2005).

Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., & Slattery, S. (2000). Learning to Construct Knowledge Bases from the World Wide Web. Artificial Intelligence, 118(1-2): 69-113.

de Vries, A.P., Thom, J.A., Vercoustre, A-M, Craswell N., and Lalmas, M. (2007) INEX 2007 Entity Ranking Track Guidelines - V1. In Pre-Proc. of INitiative for the Evaluation of XML Retrieval.

Etzioni, O., Cafarella, M., Downey, D., Popescu, A., Shaked, T., Soderland, S., Weld, S., & Yates, A. (2004) Methods for Domain-Independent Information Extraction from the Web: An Experimental Comparison. In Proc. of AAAI 2004, pp. 391-398.

Fang, H., and Zhai, C. (2007) Probabilistic Models for Expert Finding. In Proc. of ECIR 2007: 418-430

Fu, Y., Yu, W., Li, Y., Liu, Y., Zhang, M., & Ma, S. (2006a) THUIR at TREC 2005: Enterprise Track. In Proc. of TREC 2005.

Fu, Y., Xiang, R., Zhang, M., Liu, Y., & Ma, S. (2006b) A PDD-Based Searching Approach for Expert Finding in Intranet Information Management. In Proc. of AIRS 2006: 43-53

Hatcher, E., & Gospodnetic, O. (2004) Lucene in Action. Manning Publications Co, ISBN: 1932394281.

Hu, G., Liu, J., Cao, Y., Li, H., Nie, J-Y, & Gao, J. (2006) A Supervised Learning Approach to Entity Search. In Proc. of AIRS: 54-66.

Kleinberg, J. (1998) Authoritative sources in a hyperlinked environment. In Proc. of Ninth Annual ACM-SIAM Symposium On Discrete Algorithms (SODA), pp. 668-677.

M. Kolla, & O. Vechtomova (2007) In Enterprise Search: Methods to Identify Argumentative Discussions and to Find Topical Experts. In Proc. of TREC 2006.

Macdonald, C., Ounis, I. (2006) Voting for candidates: adapting data fusion techniques for an expert search task. In Proc. of CIKM 2006: 387-396.

Macdonald, C., and Ounis, I. (2007a) Expertise drift and query expansion in expert search. In Proc. of CIKM 2007: 341-350.

Macdonald, C., & Ounis, I. (2007b) A Belief Network Model for Expert Search. In Proc. of 1st International Conference on Theory of Information Retrieval (ICTIR), 18 - 20 October 2007, Budapest, Hungary.

Maybury, M., D'Amore, R, & House, D. (2001) Expert Finding for Collaborative Virtual Environments. Communications of the ACM (CACM) 44(12): 55- 56. In Ragusa, J. and Bochenek, G. (eds). Special Section on Collaboration Virtual Design Environments.

Metzler, D., Lavrenko, V., & Croft, W.B. (2004) Formal multiple-bernoulli models for language modeling. In Proc. of SIGIR 2004: 540-541

Nenadic, G., & Ananiadou, S. (2006) Mining Semantically Related Terms from Biomedical Literature. ACM Transactions on Asian Language Information Processing (TALIP), 5(1), 22-43.

Page, L., Brin, S. , Motwani R., and Winograd, T. (1998) The PageRank citation ranking: bringing order to the Web. Computer Science Department, Stanford University, Technical Report.

Petkova, D., and Croft, W. B. (2006) Hierarchical Language Models for Expert Finding in Enterprise Corpora. In Proc. of IEEE International Conf. on Tools with Artificial Intelligence: 599-608

Petkova, D. and Croft, W. B. (2007a) UMass at TREC 2006: Enterprise Track. In Proc. of the Fifteenth Text REtrieval Conference (TREC 2006), Gaithersburg, Maryland USA.

Petkova, D. and Croft, W. B. (2007b) Proximity-based Document Representation for Named Entity Retrieval. To appear in Proc. of ACM Sixteenth Conference on Information and Knowledge Management (CIKM 2007), November 6-8, 2007, Lisboa, Portugal.

Robertson, S.E. (1990) On term selection for query expansion. Journal of Documentation 46, 359-364.

Robertson, S.E., Walker, S., Beaulieu, M.M., Gatford, M., & Payne, A. (1995): Okapi at TREC-4. In Proc. of the Fourth Text REtrieval Conference (TREC-04), 73-96.

Salton, G., Fox, E. A., and Wu, H. (1983) Extended Boolean information retrieval. Communications of ACM 26: 1022–1036.

Salton, G., and Buckley, C. (1988) Term-weighting approaches in automatic text retrieval. Information Processing & Management 24(5): 513–523

Serdyukov, P., and Hiemstra, D. (2008) Modeling documents as mixtures of persons for expert finding. In Proc. of ECIR 2008.

Silverstein, C., Henzinger, M.R., Marais, H., and Moricz, M. (1999) Analysis of a Very Large Web Search Engine Query Log. SIGIR Forum 33(1): 6-12.

Soboroff, I., de Vries, A.P., & Craswell, N. (2007) Overview of the TREC 2006 Enterprise Track. In Proc. of The Fifteenth Text REtrieval Conference (TREC 2006), Gaithersburg, Maryland USA.

Song, D., & Bruza, P.D. (2003) Towards Context Sensitive Informational Inference. Journal of the American Society for Information Science and Technology (JASIST), 52(4), 321-334.

Vechtomova, O., Robertson, S., & Jones, S. (2003) Query Expansion with Long-Span Collocates. Information Retrieval, 6(2), pp. 251-273.

Westerveld, T. (2007) Correlating Topic Rankings and Person Rankings to Find Experts. In Proc. of TREC2006.

Yao, C., Peng, B., He, J., & Yang, Z. (2006) CNDS Expert Finding System for TREC. In Proc. of TREC2005.

Yimam-Seid, D., & Kobsa, A. (2003) Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach. Journal of Organizational Computing & Electronic Commerce 13 (1).

Zhao, H., & Lu, W. (2007) Using Document Weight Combining Method for Enterprise Expert Mining. In Proc. of Intl. Conf. on Wireless Communications, Networking and Mobile Computing (WiCom): 3721-3723.

Zhu, J., Song, D., Rüger, S., Eisenstadt, M., & Motta, E., (2007) The Open University at TREC 2006 Enterprise Track Expert Search Task. In Proc. of the Fifteenth Text REtrieval Conference (TREC 2006), Gaithersburg, Maryland USA. (Invited for presentation at TREC2006)

Zhu, J., Song, D., & Rüger, S. (2008a) The Open University at TREC 2007 Enterprise Track Document Search and Expert Search Tasks. In Proc. of the Sixteenth Text REtrieval Conference (TREC 2007), Gaithersburg, USA. (Invited for presentation at TREC2007)

Zhu, J., Song, D., & Rüger, S. (2008b) Integrating Document Features for Entity Ranking. In Proc. of the INitiative for the Evaluation of XML Retrieval (INEX) 2007, Dagstuhl, Germany, Springer LNCS.