Combining global and local semantic contexts for improving biomedical information retrieval

Duy Dinh, Lynda Tamine

Combining global and local semantic contexts for improving biomedical information retrieval

Duy Dinh, Lynda Tamine

University of Toulouse,
118 route de Narbonne, 31062 Toulouse, France
{dinh,lechani}@irit.fr

Abstract. In the context of biomedical information retrieval (IR), this paper explores the relationship between the document's global context and the query's local context in an attempt to overcome the term mismatch problem between the user query and documents in the collection. Most solutions to this problem have been focused on expanding the query by discovering its context, either *global* or *local*. In a global strategy, all documents in the collection are used to examine word occurrences and relationships in the corpus as a whole, and use this information to expand the original query. In a local strategy, the top-ranked documents retrieved for a given query are examined to determine terms for query expansion. We propose to combine the document's global context and the query's local context in an attempt to increase the term overlap between the user query and documents in the collection via document expansion (DE) and query expansion (QE). The DE technique is based on a statistical method (IR-based) to extract the most appropriate concepts (global context) from each document. The QE technique is based on a blind feedback approach using the top-ranked documents (local context) obtained in the first retrieval stage. A comparative experiment on the TREC 2004 Genomics collection demonstrates that the combination of the document's global context and the query's local context shows a significant improvement over the baseline. The MAP is significantly raised from 0.4097 to 0.4532 with a significant improvement rate of +10.62% over the baseline. The IR performance of the combined method in terms of MAP is also superior to official runs participated in TREC 2004 Genomics and is comparable to the performance of the best run (0.4075).

Key words: Term Mismatch, Concept Extraction, Document Expansion, Query Expansion, Biomedical Information Retrieval

1 Introduction

The effectiveness of an Information Retrieval (IR) system is influenced by the degree of term overlap between user queries and relevant documents. When a user searches an information in the collection, (s)he may formulate the query using another expression to mention the same information in the document. This causes the *term mismatch* problem yielding poor search results retrieved by IR systems [1].

In this paper, we focus on addressing the term mismatch problem in the biomedical domain. In this latter, biomedical documents contain many different expressions or term variants of the same concept such as synonyms ('cancer', 'tumor' are synonyms of the concept 'neoplasm'), abbreviations ('AMP' stands for 'Adenosine Monophosphate'), lexical variations such as differences in case, singular-plural inflection, etc. A natural solution to alleviate the term mismatch problem in biomedical IR is to use concepts in ontologies as means of normalizing the document vocabulary. Many works have been focused on both concept-based *Query expansion* (QE) [1–6] and/or *Document expansion* (DE) [3,6–8].

The principle goal of QE is to increase the search performance by increasing the likelihood of the term overlap between a given query and documents that are likely to be relevant to the user information need. Current approaches of QE can be subdivided into two main categories: *global analysis* [1–3,5,6,9] and *local analysis* [10–13]. Global techniques aim to discover word relationships in a large collection such as Web documents [9] or external knowledge sources like Wordnet [2], MeSH [5,6] or UMLS [1,3]. Local techniques emphasize the analysis of the top ranked documents retrieved for a query [10–12].

Similarly, DE can help to enhance the semantics of the document by expanding the document content with the most informative terms. This technique has been used recently in the context of textual document IR [7,8] as well as in the context of biomedical IR [3,6]. The difference between DE and QE is basically the timing of the expansion step. In DE, terms are expanded during the indexing phase for each individual document while in QE only query terms are expanded at the retrieval stage.

In the work presented in this paper, we explore the impact of using concepts from MeSH (global context) for a concept-based document and query representation enhanced with techniques of DE and QE. Our contributions are outlined through the following key points:

1. We propose a novel IR-based concept identification method for selecting the most representative concepts in each document. Concepts are then used to expand the document content to cope with the term mismatch problem.
2. We propose to combine the document's global context and the query's local context in an attempt to increase the term overlap between the user's query and documents in the collection through both DE and QE. The DE technique, which is classified as global, is based on our concept identification method using a domain terminology namely MeSH. The QE technique, which is classified as local, is based on a blind feedback approach using the top-ranked expanded documents obtained in the first retrieval stage.

The remainder of this paper is structured as follows. We introduce the related work in Section 2. Section 3 presents the techniques involved for combining query and document expansion. Experiments and results are presented in section 4. We conclude the paper in section 5 and outline research directions for future work.

2   Related work

The term mismatch problem between user queries and documents has been the focus of several works in IR for almost 40 years [9, 10]. One of the earliest studies on QE was carried out by Sparck Jones [9] who clustered words based on co-occurrence in documents and used those clusters to expand the query. Since then, a number of studies have been undertaken and are divided into two main approaches: *global analysis* [2, 3, 5] and *local analysis* [10–12]. In a global strategy, all documents in the collection are used to examine word occurrences and relationships in the corpus as a whole, and use this information to expand any particular query. For example, the global context QE technique presented in [2] explored the lexical-semantic links in Wordnet in order to expand hierarchically related terms to the original query, but reported results have not been positive somehow due to term ambiguity. In a local strategy, the top-ranked documents retrieved for a given query q are examined to determine terms for QE. The local context QE presented in [10] involves generating a new query in an iterative way by taking into account the difference between relevant and non-relevant document vectors in the set of the retrieved ones: at each iteration, a new query is automatically generated in the form of a linear combination of terms from the previous query and terms automatically extracted from both relevant and irrelevant documents. Empirical studies (e.g., [11–13]) have demonstrated that such approach is usually quite effective. Moreover, in their technique, also known as pseudo-relevance feedback or blind feedback, they assume that the top-ranked documents (e.g., top 10, 20 ones) are relevant and could be used for QE.

Similar to QE methods, DE or also document smoothing, can be classified as either global [3, 6] or local [7, 8]. While the global DE made use of domain specific knowledge sources, the local DE focused on the analysis of the sub-collection. For instance, authors in [8] proposed a local DE method for expanding the document with the most informative terms from its neighbor documents, which are identified using a cosine similarity between each document and its neighbors in the collection. Another work in [7] proposed to smooth the document with additional terms collected from the top-ranked documents w.r.t the original document by using three different term selection measures: Term Selection Value [14], Kullback-Leibler Divergence [12], and BM25 [15]. Both of these works are similar in the way that they proposed to combine the local context DE with the local context analysis QE to better improve the IR effectiveness.

In the biomedical domain, several works have been done using both QE [1, 4, 5, 13] and/or DE techniques [3, 6]. The work in [13] adapted the local analysis QE approach for evaluating the IR performance on searching MEDLINE documents. Their approach relies on a blind feedback by selecting the best terms from the top-ranked documents. Candidate terms for QE are weighted using the linear combination of the within-query term frequency and the inverse document frequency according to whether the term appears in the query and/or the document. Using a global approach, the work in [5] investigated QE using MeSH to expand terms that are automatically mapped to the user query via the Pubmed's Automatic Term Mapping (ATM) service, which basically maps untagged terms

from the user query to lists of pre-indexed terms in Pubmed's translation tables (MeSH, journal and author). Authors in [1] exploited several medical knowledge sources such MeSH, Entrez gene, SNOMED, UMLS, etc. for expanding the query with synonyms, abbreviations and hierarchically related terms identified using Pubmed. Furthermore, they also defined several rules for filtering the candidate terms according to each knowledge source. Differently from the latter, the work in [6] combined both QE and DE using the MeSH thesauri to retrieve medical records in the ImageCLEF 2008 collection. More concretely, they combined an IR-based approach of QE and DE for a conceptual indexing and retrieval purpose. For each MeSH concept, its synonyms and description are indexed as a single document in an index structure. A piece of text, the query to the retrieval system, is classified with the best ranked MeSH concepts. Finally, identified terms denoting MeSH concepts are used to expand both the document and the query. Authors in [4] proposed a knowledge-intensive conceptual retrieval by combining both the global context (i.e., concepts in ontologies) using the Pubmed ATM service and the local context (top-ranked documents) of the query. They reported an improvement rate of about 23% over the baseline.

This paper examines the utility of DE/QE for resolving the term mismatch problem in biomedical IR. Compared to previous works, the major contributions of this work are twofold. First, differently from the work in [6], we propose a novel IR-based concept extraction method by estimating concept relevance for a document by combining both document-to-concept matching degree and document-concept correlation. Second, unlike previous works [1,3–6], which only focus on QE/DE using the global context (UMLS, MeSH, etc.) or only QE/DE using the local context (corpus-based) [7, 8] or even only QE using both the local and global context [4], we aim to point out that the combination of the document's global context (MeSH) and the query's local context (top-ranked documents) may be a source evidence to improve the biomedical IR effectiveness.

3   Combining global and local contexts for biomedical IR

Our retrieval framework is made up of two main components detailed below: (1) global context document expansion and (2) local context query expansion. We integrate them into a conceptual IR process as the combination of the global and local semantic contexts for improving the biomedical IR effectiveness.

3.1   Global context document expansion

In our DE approach, each document is expanded with preferred terms denoting concepts[1] identified using an IR-based document-to-concept mapping method. In other words, given a document, the mapping leads to the selection of the most relevant MeSH concepts using a content-based similarity measure. Furthermore,

---

[1] In MeSH, each concept is defined by one preferred term (e.g., 'Skin Neoplasms') and many non-preferred terms (e.g., 'Tumors of the Skin', 'Skin Cancer', etc.)

in order to take into account the importance of the word order while matching a concept entry, which can be both a preferred or non-preferred term, to a bounded multi-word term located in a window issued from a document, we propose to leverage the content-based similarity between the document and concept entries using a rank correlation based matching. Our basic assumption behind concept relevance is that a list of document words is more likely to map a concept that (1) both shares a maximum number of words either among its preferred or non-preferred terms; (2) the words tend to appear in the same order so to cover the same meaning. For example, the phrase 'Skin cancer represents the most commonly diagnosed disease, surpassing lung, breasts, ...' should be mapped to 'Skin cancer' rather than 'Lung cancer', 'Breast cancer' or 'Cancer'.

Our strategy, which is based on ranking concepts extracted from documents using a combined score, involves three steps detailed below: (1) computing a content-based matching score, (2) computing a rank correlation based score, (3) selecting an appropriate number of terms denoting concepts for document expansion by ranking candidate concepts according to their combined score.

1. **Computing a content-based matching score**. According to our IR based approach, the top-ranked relevant concepts issued from MeSH are assigned to the document. Formally, we compute for each concept vector C a content-based cosine similarity w.r.t the document D, denoted $Sim(C, D)$, as follows:

$$Sim(C, D) = \frac{\sum_{j=1}^{N_c} c_j * d_j}{\sqrt{\sum_{j=1}^{N_c} c_j^2} * \sqrt{\sum_{j=1}^{N_c} d_j^2}} \tag{1}$$

where N_c is the total number of concepts in MeSH, d_j is the weight of word w_j in document D computed using an appropriate weighting schema, c_j is the weight of word w_j in concept C computed using the BM25 weighting schema [15]:

$$c_j = tfc_j * \frac{log\frac{N-n_j+0.5}{n_j+0.5}}{k_1 * ((1-b) + b * \frac{cl}{avgcl}) + tfc_j} \tag{2}$$

where tfc_j is the number of occurrences of word w_j in concept C, N is the total number of concepts in MeSH, n_j is the number of concepts containing at least one occurrence of word w_j in its textual fields, cl is the length of concept C, i.e. the total number of distinct words in C, and $avcl$ is the average concept length in the MeSH thesaurus, k_1, and b are tuning parameters.

2. **Computing a rank correlation coefficient**. The candidate concepts extracted from step 1 are re-ranked according to a correlation measure that estimates how much the word order of a MeSH entry is correlated to the order of words in the document. For this aim, we propose to measure the word order correlation between the concept entry and the document both represented by word position vectors. Formally, the correlation measure is computed using the Spearman operator as follows: let document $D = (w_{d_1}, w_{d_2}, \ldots, w_{d_L})$ be the ranked word based vector according to the average position of related occurrences

in document D, i.e., w_{d_i} is the document word in D such that $p\bar{o}s(occs(w_{d_i})) < p\bar{o}s(occs(w_{d_{i+1}}))$ $\forall i = 1\ldots L-1$, where $occs(w_{d_i})$ is the set of positions of word w_{d_i} in document D, L is the total number of unique words in document D. Similarly, let $E = (w_{e_1}, w_{e_2}, \ldots, w_{e_{L'}})$ be the ranked word based vector according to the average position of related occurrences in concept entry E, where L' is the concept entry length. We denote the set of words in D as $words(D) = \{w_{d_1}, w_{d_2}, \ldots w_{d_L}\}$ and in concept entry E as $words(E) = \{w_{e_1}, w_{e_2}, \ldots, w_{e_{L'}}\}$.

First, in order to avoid false rank bias, when measuring the word order correlation, a portion of the document window bounded by the first and last word occurrences shared by the concept entry E and the document is captured and normalized as $D_w = (w_{d_w}, w_{d_{w+1}}, \ldots, w_{d_W})$, where $words(D_w) \subset words(D)$, $w_{d_w} \in words(E), w_{d_{W+1}} \notin words(E)$. Afterwards, the Spearman correlation is used to compute the word rank correlation between words in D and E:

$$\rho(E, D) = 1 - \frac{6*\sum_i^T [rank(w_i, D_w) - rank(w_i, E)]^2}{T*(T^2 - 1)} \qquad (3)$$

where $rank(w_i, D_w)$ (resp. $rank(w_i, E)$) is the word order of word w_i according to $p\bar{o}s(occs(w_{d_i}))$ in D_w (resp. E), $T = |words(D_w) \cap (words(E))|$ is the number of shared words between document D and concept entry E. We simply assume that the rank of an absent word in D_w or E is assigned a default value $r_0 > T$. The correlation coefficient $\rho(E, D)$ allows measuring the degree of agreement between two word rankings in E and D_w. The value of $\rho(E, D)$ lies between -1 and 1 according to the agreement between two rankings. In order to consider each significant entry separately, we practically compute:

$$\rho(C, D) = Max_{E \in Entries(C)} \ \rho(E, D) \qquad (4)$$

where $Entries(C)$ refers to both preferred or non-preferred terms.

3. ***Selecting the candidate concepts for document expansion***. Finally the content based similarity and the correlation between concept C and document D are combined in order to compute the overall relevance score $Rel(C, D)$:

$$Rel(C, D) = (1 + Sim(C, D)) * (1 + \rho(C, D)) \qquad (5)$$

The N top-ranked concepts with highest scores are selected as candidate concepts of document D. Preferred terms are used to expand the document content. Document terms are then weighted using the state-of-the-art BM25 model [15].

3.2   Local context query expansion

The local context QE applies a blind-feedback technique to select the best terms from the top-ranked expanded documents in the first retrieval stage. In this expansion process, terms in the top-returned documents are weighted using a particular Divergence From Randomness (DFR) term weighting model [12]. In our work, the Bose-Einstein statistics [12] is used to weight terms in the expanded query q^e derived from the original query q. Formally:

$$weight(t \in q^e) = tfq_n + \beta * \frac{\text{Info}_{\text{Bo1}}}{\text{MaxInfo}} \qquad (6)$$

where

- $tfq_n = \frac{tfq}{\max_{t \in q} tfq}$: the normalized term frequency in the original query q,
- $\text{MaxInfo} = \arg\max_{t \in q^e} \max \text{Info}_{\text{Bo1}}$,
- Info_{Bo1} is the normalized term frequency in the expanded query induced by using the Bose-Einstein statistics, that is:

$$
\begin{aligned}
\text{Info}_{\text{Bo1}} &= -\log_2 \text{Prob}(\text{Freq}(\text{t}|\text{K})|\text{Freq}(\text{t}|\text{C})) \\
&= -\log_2(\tfrac{1}{1+\lambda}) - \text{Freq}(\text{t}|\text{K}) * \log_2(\tfrac{\lambda}{1+\lambda})
\end{aligned}
\tag{7}
$$

where Prob is the probability of obtaining a given frequency of the observed term t within the topmost retrieved documents, namely K; C is the set of documents in the collection; $\lambda = \frac{\text{Freq}(\text{t}|\text{C})}{\text{N}}$, with N is the number of documents in the collection, $\beta = 0.4$. The number of top-ranked documents and the number of terms expanded to the original query are tuning parameters.

4 Experimental evaluation

4.1 Experimental data set

We used the TREC Genomics 2004 collection [16], which is a 10-year subset (1994-2003) of the MEDLINE database, under the Terrier IR platform [17] for validating our conceptual IR approach. Some statistical characteristics of the collection are depicted in Table 1. However, human relevance judgments were merely made to a relative small pool, which were built from the top-precedence run from each of the 27 participants. Our prototype IR system only indexes and searches all human relevance judged documents, i.e. the union of 50 single pools that contains total 42,255 unique articles' titles and/or abstracts. We did not use the set of manually annotated MeSH concepts provided by human experts in our system, but which we referred to as the manual DE task.

In our experiments described later, we used the latest version of MeSH released in 2010, which consists of 25,588 main headings and also over 172,000 entry terms that assist in finding the most appropriate MeSH concepts.

For measuring the IR effectiveness, we used *P@10, P@20* representing respectively the mean precision values at the top 10, 20 returned documents and *MAP* representing the *Mean Average Precision* calculated over all topics.

Table 1: TREC Genomics 2004 test collection statistics

Number of documents	4.6 millions
Average document length	202
Number of queries	50
Average query length	17
Average number of relevant docs/query	75

4.2   Experimental design

The purpose of this investigation is to determine the utility of the combination of the global context of the document and the local context of the query. Hence, we carried out two series of experiments: the first one is based on the classical indexing of title and abstract articles using the state-of-the-art weighting scheme BM25 [15], as the baseline, denoted *BM25*. The second one concerns our indexing and retrieval approach and consists of five scenarios:

1. the first one concerns the document expansion using concepts[2] manually assigned by human experts, denoted DE^{manual} or simply DE^m,
2. the second one concerns the document expansion using concepts identified by the combination of the cosine content-based similarity and the Spearman rank correlation between word occurrences in the document and concept entries (see section 3.1, formula 5), denoted $DE^{combination}$ or simply DE^c,
3. the third one concerns the query expansion using the blind feedback technique applied on the original document (title and abstract) without DE (see section 3.2, formula 6), denoted QE.
4. the fourth one concerns the combination of both QE and the manual DE^m method, denoted $QE + DE^m$,
5. the last one concerns our method which relies on the combination of both QE and our automatic DE^c strategy as described above, denoted $QE+DE^c$.

4.3   Results and discussion

First, we aim to measure the impact of the number of expanded terms on the IR effectiveness by tuning the number of terms denoting concepts expanded to the document for DE and the number of terms from the top-ranked documents expanded to the original query for QE. Second, we will measure the IR effectiveness using the optimal number of expanded terms for QE and/or DE. Finally, we compare our IR results to the official ones in TREC 2004 Genomics Track.

Impact of the number of expanded terms. For automatic DE, we tuned the number of candidate concepts from 0 to 50, with a step of 5. Query terms in the expanded documents are weighted using the state-of-the-art BM25 model [15]. Table 2 shows the MAP results achieved by our document expansion method, namely DE^c. The results show that our document expansion method achieves the best MAP (0.4118) when expanding with $N = 5$ terms denoting concepts to the document content. We observed that the query space (i.e, 50 ad hoc topics) usually contains synonyms (non-preferred terms, e.g., 'FANCD2', 'ache', 'breast cancer', etc.) of medical concepts while the document space has been adjusted using their preferred terms (e.g., Fanconi Anemia Complementation Group D2 Protein, Pain, Breast Neoplasms, etc.). Therefore, we believe that picking up some related terms from the expanded documents returned for each topic would

---

[2] only preferred terms are used for document expansion

better increase the term overlap between the expanded query and the expanded documents in the collection. We retain $N = 5$ for the experiments described in the next section.

Table 2: Document expansion: P@10, P@20, MAP results by varying N from 0 to 50

N	P@10	P@20	MAP
0	**0.5920**	0.5380	0.4097
5	0.5780	**0.5390**	**0.4118**
10	**0.5920**	0.5280	0.4031
15	0.5840	0.5330	0.3999
20	0.5660	0.5290	0.4032
25	0.5660	0.5250	0.4007
30	0.5600	0.5150	0.3975
35	0.5400	0.5070	0.3918
40	0.5340	0.5020	0.3882
45	0.5300	0.4940	0.3855
50	0.5280	0.4880	0.3835

For automatic QE, the number of expanded terms and the number of selected documents are tuned from 5 to 25 with a step of 5. Query terms in the original documents are weighted using the state-of-the-art BM25 model [15]. Table 3 depicts the MAP results of 50 ad hoc topics. The best results are obtained at 20 terms and 20 top-ranked documents. Therefore, we retain 20 terms and 20 documents for the next experiments described later.

Table 3: Query expansion: MAP results by varying the number of expanded terms/docs

Nb. terms / Nb. docs	5	10	15	**20**	25
5	0.4369	0.4347	0.4455	0.4422	0.4440
10	0.4204	0.4232	0.4286	0.4289	0.4332
15	0.4357	0.4407	0.4431	0.4463	0.4428
20	0.4373	0.4395	0.4454	**0.4475**	0.4467
25	0.4347	0.4403	0.4429	0.4448	0.4473

Retrieval effectiveness. At this level, we aim to study the impact of the combination of the global context of the document and the local context of the query. For this purpose, we measure the IR effectiveness of five retrieval scenarios described in section 4.2. We now present and discuss the experimental results.

Table 4 shows the IR effectiveness of 50 ad hoc topics. According to the results, we observe that both the manual and automatic DE methods (DE^m, DE^c) slightly outperform the baseline in terms of MAP, but both of these methods

Table 4: IR effectiveness in terms of $P@10, P@20, MAP$ (%change)

	P@10	**P@20**	**MAP**
$BM25$	**0.5920**	0.5380	0.4097
DE^m	0.5900 (-00.34)	0.5370 (-00.19)	0.4139 (+01.03) †††
DE^c	0.5780 (-02.36)	0.5390 (+00.19)	0.4118 (+00.51)
QE	0.5720 (-03.38)	0.5430 (+00.93)	0.4475 (+09.23)
$QE + DE^m$	0.5320 (-10.14)††	0.5220 (-02.97)	**0.4567** (+11.47)
$QE + DE^c$ (our method)	0.5860 (-01.01)	**0.5470** (+01.67)	0.4532 (+10.62) †††

Paired sample t-test: † significant ($p < 0.05$), †† very significant ($p < 0.01$), and †††
extremely significant ($p < 0.001$).

do not improve the IR performance in terms of P@10 and P@20. The difference
between these two methods is about the number of terms expanded to the docu-
ment, a dozen for the manual DE [16] and five for the automatic DE. Automatic
QE using related terms, which may denote domain concepts or not, from top-
ranked documents improves better the MAP. As argued in this study, we can see
that the combination of the local context of the query and the global context of
the document is helpful for improving much better the IR performance in terms
of MAP. Indeed, as depicted in Table 4, the combination of the QE and DE^m
shows a gain of +11.47% in terms of MAP over the baseline. However, the preci-
sion values of this combination are dramatically decreased. The reason could be
explained as follows: in general, long queries (17 terms in average) are enough
to describe the user information need, therefore expanding *related terms* to a
long query may improve the recall but not the precision. Furthermore, expanded
terms in the document are preferred terms while the ones in reformulated query
may be terms denoting domain concepts or not. Therefore, the query space and
document space are not correctly adjusted for increasing the term overlap. Our
document expansion method for detecting domain concepts revealing the doc-
ument subject matter(s) (the global context of the document) enhanced with
the local context of the query allows to retrieve more relevant documents than
the baseline as well as document expansion alone or query expansion alone. The
highest MAP value of our method $QE + DE^c$ is obtained at 0.4532 with an
improvement rate of +10.62% even though the precision values are slightly dif-
ferent compared to the baseline. As shown in Table 4, the paired-sample T-test
($M = 4.35\%, t = 3.5291, df = 49, p = 0.0009$) shows that the combination of the
global context of the document and the local context of the query, i.e. $QE + DE^c$
method, is extremely statistically significant compared to the baseline.

Comparative evaluation. We further compare the IR performance of our best
automatic retrieval method ($QE + DE^c$) to official runs in TREC Genomics
2004. Table 5 depicts the comparative results of our best run with official runs
of participants in the TREC 2004 Genomics Track. The results show that the
precision values (P@10, P@20) of our best run are better than the third run
but lower than the two first runs. However, the MAP of our run is much bet-

ter than the first run. As shown in Figure 1, our best indexing and retrieval method ($QE + DE^c$) outperforms the best run submitted to TREC Genomics 2004 (MAP=0.4075) with a gain of +11.21%. Thus, we conclude that conceptual indexing and searching in conjunction with an efficient way of identifying appropriate concepts representing the semantics of the document as well as of the query would significantly improve the biomedical IR performance.

Table 5: The comparison of our best run with official runs participated in TREC 2004 Genomics Track. Runs in TREC are ranked by Mean Average Precision (MAP)

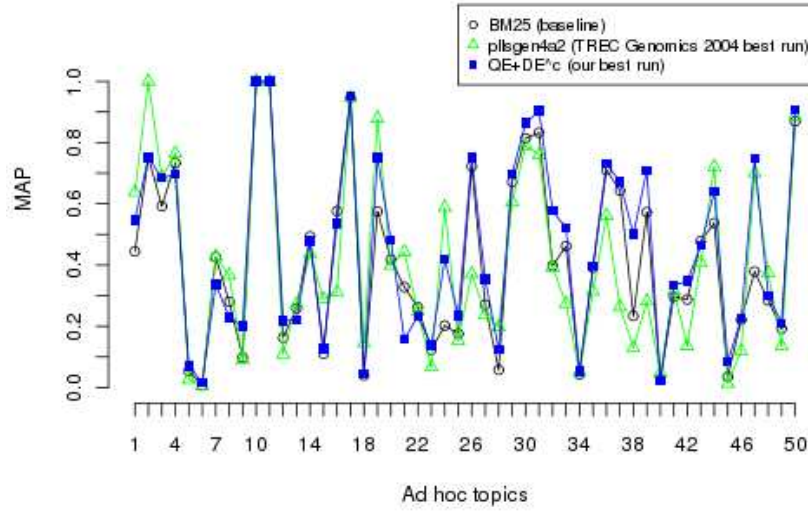| Run	P@10	P@20	MAP
pllsgen4a2 (the best)	0.6040	0.5720	0.4075
uwmtDg04tn (the second)	**0.6240**	**0.5810**	0.3867
pllsgen4a1 (the third)	0.5700	0.5430	0.3689
PDTNsmp4 (median)	0.4056	0.4560	0.2074
edinauto5 (the worst)	0.0360	0.0310	0.0012
QE+DE^c (our best run)	0.5860	0.5470	**0.4532**

Fig. 1: The comparison in terms of MAP of our runs (Combination approach and the baseline) to the official best run in TREC 2004 Genomics Track on 50 ad hoc topics

5   Conclusion

In this paper, we have proposed a novel IR method for combining the global context DE and the local context QE. Our automatic DE relies mainly on turning the concept mapping into a concept retrieval task by means of concept relevance scoring. The QE technique relies on the selection of related terms from the top-ranked documents. The results demonstrate that our IR approach shows a significant improvement over the classical IR as well as DE or QE as alone. The performance of our approach is also significantly superior to the average of official runs in TREC 2004 Genomic Track and is comparable to the best run.

For future work, we plan to improve the precision of our concept extraction method, which will integrate the concept centrality and specificity. We believe that these two factors allow to overcome the limits of the bag-of-words based similarity by leveraging lexical and semantic contents of candidate concepts.

References

1. Stokes, N., Li, Y., Cavedon, L., Zobel, J.: Exploring criteria for successful query expansion in the genomic domain. Information Retrieval **12**(1) (2009) 17–50
2. Voorhees, E.M.: Query expansion using lexical semantic relations. In: SIGIR'94 Conference on Research and Development in Information Retrieval. (1994) 61–69
3. Le, D.T.H., Chevallet, J.P., Dong, T.B.T.: Thesaurus-based query and document expansion in conceptual indexing with umls. In: RIVF'07. (2007) 242–246
4. Zhou, W., Yu, C.T., *et. al.*: Knowledge-intensive conceptual retrieval and passage extraction of biomedical literature. In: SIGIR. (2007) 655–662
5. Lu, Z., Kim, W., Wilbur, W.J.: Evaluation of query expansion using mesh in pubmed. Information Retrieval **12**(1) (2009) 69–80
6. Gobeill, J., Ruch, P., Zhou, X.: Query and document expansion with medical subject headings terms at medical imageclef 2008. In: CLEF 2008. (2009) 736–743
7. Billerbeck, B., Zobel, J.: Document expansion versus query expansion for ad-hoc retrieval. In: the 10th Australasian Document Comput. Symp. (2005) 34–41
8. Tao, T., Wang, X., *et. al.*: Language model information retrieval with document expansion. In: Association for Computational Linguistics. (2006) 407–414
9. Sparck Jones, K.: Automatic Keyword Classification for Information Retrieval, London: Butterworths (1971)
10. Rocchio, J. In: Relevance Feedback in Information Retrieval. (1971) 313–323
11. Xu, J., Croft, W.B.: Query expansion using local and global document analysis. In: Conference on Research and Development in Information Retrieval. (1996) 4–11
12. Amati, G.: Probabilistic models for Information Retrieval based on Divergence from Randomness. PhD thesis, University of Glasgow (2003)
13. Abdou, S., Savoy, J.: Searching in medline: Query expansion and manual indexing evaluation. Information Processing Management **44**(2) (2008) 781–789
14. Robertson, S.E., Walker, S.: Okapi/keenbow at trec-8. In: TREC-8. (1999) 151–162
15. Robertson, S.E., Walker, S., Hancock-Beaulieu, M.: Okapi at trec-7: Automatic ad hoc, filtering, vlc and interactive. In: TREC-7 Proceedings. (1998) 199–210
16. Hersh, W., Bhuptiraju, R.: Trec 2004 genomics track overview. In: The Thirteenth Text Retrieval Conference (TREC 2004)
17. Ounis, I., Lioma, t.: Research directions in terrier. Novatica Special Issue on Web Information Access, Ricardo Baeza-Yates et al. (Eds), Invited Paper (2007)