

Interactive and Automatic Query Expansion: A Comparative Study with an Application on Arabic

¹Ghassan Kanaan, ¹Riyad Al-Shalabi, ²Sameh Ghwanmeh and ¹Basel Bani-Ismael
¹Arab Academy for Banking and Financial Sciences, Amman-Jordan
²Computer Engineering Department, Yarmouk University, Irbid-Jordan

Abstract: Much attention has been paid to the relative effectiveness of Interactive Query Expansion (IQE) versus Automatic Query Expansion (AQE). This research has been shown that automatic query expansion (collection dependent) strategy gives better performance than no query expansion. The percentage of queries that are improved by AQE strategy is 57% with average precision equal to 43.2. Compared against AQE (collection dependent) strategy, IQE gives better average precision than AQE strategy. The percentage of queries that are improved by best IQE decision is 86% with average precision equal to 44.1. Evaluation process reveals that the value of n in AQE strategy that gave the optimal value of average precision for the whole query set is equal to one.

Key words: Arabic language, evaluation of ir, interactive query expansion, query expansion

INTRODUCTION

Information Retrieval (IR) is a task of selecting documents from a database in response to a user's query and ranking them according to relevance. This has been usually accomplished using statistical methods (often coupled with manual encoding) that:

- Select terms (words, phrases and other units) from documents that are deemed to best represent their content
- Create an inverted index file (or files) that provide an easy access to documents containing these terms

A subsequent search process attempts to match preprocessed user queries against term based representations of documents in each case determining a degree of relevance between the two, which depends upon the number and types of matching terms. A search is successful if it can return as many as possible documents which are relevant to the query, with as few as possible non-relevant documents. In addition, the relevant documents should be ranked ahead of non-relevant ones^[1].

Query expansion techniques aim to improve a user's search by adding new query terms to an existing query. A standard method of performing query expansion is to use relevance information from the user; those documents a user has assessed as containing relevant information. The content of these relevant documents can be used to form a set of possible

expansion terms, ranked by some measure that describes how useful the terms might be in attracting more relevant documents^[2]. All or some of these expansion terms can be added to the query either by the user-interactive query expansion (IQE)-or by the retrieval system-Automatic Query Expansion (AQE)^[3]. One argument in favor of AQE is that the system has access to more statistical information on the relative utility of expansion terms and can make a better selection of which terms to add to the user's query. The main argument in favor of IQE is that interactive query expansion gives more control to the user. As it is the user who decides the criteria for relevance in a search, then the user should be able to make better decisions on which terms are likely to be useful^[4].

In this research the potential effectiveness of interactive query expansion has been examined. A chain of experiments were carried out using 242 Arabic abstracts From the Saudi Arabian National Computer Conference. The experiments have been conducted to provide a clear comparison between AQE (collection dependent) strategy and IQE techniques. Evaluation process has been performed to reveal the best value of n in AQE strategy that gives the optimal value of average precision for the whole query.

Automatic query expansion strategies: In^[3] the author discusses three AQE techniques. These techniques act as baseline performance measures for comparing AQE and IQE:

Collection independent expansion: A common approach to AQE is to add a fixed number of terms, n , to each query.

Collection dependent expansion: when using a specific test collection we can calculate a better value of n ; one that is specific to the test collection used. To calculate n , for each collection, we compared the average precision over all the queries used in each collection after the addition of the top n expansion terms, where n varies from 1 to 15. The value of n that gives the optimal value of average precision for the whole query set was taken to be the value of n for each query in the collection. These values could not be calculated in an operational environment, where knowledge of all queries submitted is unknown. However, it gives a stricter AQE baseline measure as the value of n is optimal for the collection used.

Query dependent expansion: The collection dependent expansion strategy adds a fixed number of terms to each query within a test collection. This is optimal for the entire query set, but may be sub-optimal for individual queries. This means some queries may give better retrieval effectiveness for greater or smaller values of n . The query dependent expansion strategy calculates which value of n is optimal for individual queries. This may be implemented in an operational retrieval system by, for example, setting a threshold on the expansion term weights.

Why arabic language?: Previous studies show that Arabic language is one of the most widely used languages in the world, yet there are relatively few studies on the retrieval of Arabic documents in the literature. Furthermore, the lack of a realistically large test corpus has been a problem in past studies on Arabic retrieval^[5]. In this research work we will explore a few strategies for the retrieval of Arabic documents, using the recently available TREC Arabic corpus for evaluation. Arabic is a challenging language for Information Retrieval (IR) for a number of reasons. The following problems prevent to make exact keyword ineffective for Arabic retrieval:

- Orthographic variations are prevalent in Arabic; certain combinations of characters can be written in different ways. For example, sometimes in glyphs combining HAMZA or MADDA with ALEF the HAMZA or MADDA is dropped, rendering it ambiguous as to whether the HAMZA or MADDA is present, examples: نأنا-أنا
- Arabic language has a very complex morphology

- Broken plurals are common. Broken plurals are somewhat like irregular English plurals except that they often do not resemble the singular form as closely as irregular plurals resemble the singular in English. Because broken plurals do not obey normal morphological rules, existing stemmers do not handle them
- Arabic words are often ambiguous due to the tri-literal root system. In Arabic, a word is usually derived from a root, which usually contains three letters. In some derivations one or more of the root letters may be dropped, rendering many Arabic words highly ambiguous with one another
- Short vowels are omitted in written Arabic, examples: مر-ي مر-ا Synonyms are widespread, perhaps because Arabic speakers appreciate variety in expression as part of a good writing style

Document pre-processing: Previous research work^[6] introduced the document preprocessing which is a procedure which can be divided mainly into five text operations (or transformations):

- Lexical analysis of the text with the objective of treating digits, hyphens, punctuation marks and the case of letters
- Elimination of stopwords with the objective of filtering out words with very low discrimination values for retrieval purposes
- Stemming of the remaining words with the objective of removing affixes (i.e., prefixes and suffixes) and allowing the retrieval of documents containing syntactic variations of query terms (e.g., connect, connecting, connected, etc.)
- Selection of index terms to determine which words/stems (or groups of words) will be used as indexing elements. Usually, the decision on whether a particular word will be used as an index term is related to the syntactic nature of the word
- Construction of term categorization structures such as thesaurus, or extraction of structure directly represented in the text

MATERIALS AND METHODS

The research experiments were carried out on the Arabic collection, details of which are given in Table 1. For each query, the top 10 retrieved documents are used to provide a list of possible expansion terms. The wpq method of ranking terms for query expansion has been employed in this research; this has been shown to give good results for both AQE and IQE^[2,7].

Table1: Collection statistics

Number of documents	242.0
Number of queries used	014.0
Average words per query	002.4
Average number of relevant documents per query	043.4

Equation 1 has been used to calculate the weight of the term using wpq, where the value rt = the number of seen relevant documents containing term t , nt = the number of documents containing t , R = the number of seen relevant documents for query q , N = the number of documents in the collection^[2].

$$wpqt = \log \frac{rt/(R - rt)}{(nt - rt)/(N - nt - R + rt)} \cdot \left(\frac{rt}{R} - \frac{nt - rt}{N - R} \right) \quad (1)$$

The research methodology employed in this research involved the following steps for each query:

- Rank the documents using a standard $tf*idf$ weighting to obtain an initial ranking of the documents using the cosine similarity (Eq. 2 and 3)

$$w_{i,j} = tf \times idf = tf \times \log \frac{N}{n_i} \quad (2)$$

$$\begin{aligned} \text{sim}(d_j, q) &= \frac{\bar{d}_j \cdot \bar{q}}{|\bar{d}_j| \times |\bar{q}|} \\ &= \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \end{aligned} \quad (3)$$

- Use the relevant documents in the top 10 retrieved documents to obtain a list of possible expansion terms, using the wpq formula to rank the expansion terms
- Using the top 15 expansion terms. Each of these 15 terms represents a possible IQE decision that could be made by a user
- Using each combination of expansion terms, add the combination to the original query and use the new query to rank the documents, again using $tf*idf$ and the cosine similarity
- Calculate the recall-precision values for each version of the query. We use a full-freezing approach by which only re-rank the unseen documents; those not use to create the list of expansion terms. This is a standard method of assessing the performance of a query expansion technique based on relevance information^[8,10]

Only the top 15 expansion terms for query expansion have been used, as this is a computationally

intensive method of creating possible queries. In a real interactive situation users may be shown more terms than this. However, it does allow us to concentrate on those terms that are considered by the system to be the best for query expansion^[3].

The proposed Interactive/Automatic Query Expansion system has been implemented using Microsoft Visual Basic 6.0. The system uses an executable code to remove the stop words in the normalization process^[9]. The following query expansion strategies have been implemented:

- AQE (collection dependent)
- AQE (query dependent)
- IQE

System evaluation: Evaluation is a key part of this research. For each possible IQE decision we can assess the effect of making this decision on the quality of the expanded query. A comparison is conducted to compare between the possible IQE decisions against AQE (collection dependent) strategy^[3]. In AQE strategy, we need to calculate an optimal value of n which is specific to the test collection used. To calculate n , we compared the average precision (Eq. 4) over all the queries used in each collection after the addition of the top n expansion terms, where n varied from 1 to 15. The value of n that gives the optimal value of average precision for the whole query set has been taken to be the value of n for each query in the collection.

Average precision:

$$\bar{P}(r) = \frac{\sum_{i=1}^{N_q} P_i(r)}{N_q} \quad (4)$$

Where:

$\bar{P}(r)$ = Average precision at recall level r ,

N_q = The number of queries used;

$P_i(r)$ = The precision at recall level r for i -th query.

In the evaluation process we find the value of n in AQE (collection dependent) strategy that gave the optimal value of average precision for the whole query set is equal to one.

RESULTS AND DISCUSSION

We have examined the potential effectiveness of IQE against AQE (collection dependent) strategy. The effect of query expansion against no query expansion has been studied. Table 2 shows details about the

Table 2: AQE baseline and example of IQE decisions

Baseline	Arabic collection
No expansion	42.3
Collection dependent	57% 43.2
IQE best	86% 44.1
IQE worst	0% 19.8

comparison between the AQE baseline against no query expansion; the performance of the original query with no additional query terms.

Specifically, we compare how many queries in the collection give higher average precision versus no query expansion; the percentage of queries that are improved by AQE strategy. Also included in Table 2, the average precision figures given by applying the techniques. It can be seen, AQE strategy is more likely improves a query than harm it. We can compare AQE strategy against possible IQE decisions. In row 4 of Table 2, it can be seen that the percentage of queries improved and average precision obtained, when using the best IQE decision for each query. This set of figures gives the best possible results on the collection when using query expansion. This is the highest potential performance of IQE using the top 15 expansion terms.

Comparing the performance of the best IQE decision against the AQE strategy, it can be seen that IQE has the potential to be the most stable technique overall. It also has the potential to be the most effective query expansion technique as it gives the highest overall average precision. However this is only a potential benefit, it may not be easy for a user to select such an optimal set of terms. For example, in row 5 of Table 2, it shows the effect if a user was consistently making the worst IQE decisions possible, i.e., always choosing the combination of expansion terms that gave the lowest average precision of all possible decisions. Even though a user is unlikely to always make such poor decisions, these decisions are being made on terms selected from the top 15 expansion terms. So, although IQE can be effective it is a technique that needs to be applied carefully.

CONCLUSION

We have examined the potential effectiveness of interactive query expansion. The experimental results initially provide a comparison between AQE (collection dependent) strategy and IQE techniques. IQE has the potential to be an effective technique compared with AQE. One of the main claims for IQE is that searchers can be more adept, than the system, at identifying good expansion terms. This may be particularly true for certain types of search, e.g. Fowkes and Beaulieu^[12] showed that searchers preferred IQE when dealing with

complex query statements. Subjects may also be better at targeting specific aspects of the search, i.e. focusing on parts of their information need. However, the analyses presented here show that the potential benefits of IQE may not be easy to achieve.

ACKNOWLEDGMENTS

The authors gratefully acknowledge and highly appreciate the financial support and the remarkable resources provided by Yarmouk University, Irbid, Jordan.

REFERENCES

1. Strzalkowski, T., J. Wang and B. Wise, 1998. Summarization-based Query Expansion in Information Retrieval. *Coling-ACL*, pp: 123-146.
2. Robertson, E., 1990. On term selection for query expansion. *J. Documentation*, 46 (4): 359-364.
3. Ruthven, I., 2006. Re-examining the potential effectiveness of interactive query expansion. *Proceedings of the 26th ACM SIGIR Conference on Research and Development in Information Retrieval*, pp: 213-220.
4. Koenemann, J. and N. Belkin, 1996. A case for interaction: A study of interactive information retrieval behaviour and effectiveness. *Proceedings of the Human Factors in Computing Systems Conference, Zurich*, pp: 205-212.
5. Fraser, J. and R. Weischedel, 2002. *Empirical Studies in Strategies for Arabic Retrieval*.
6. Baeza-Yates, R. and B. Ribeiro-Neto, 1999. *Modern Information Retrieval*. Addison-Wesley, pp: 178-212.
7. Efthimiadis, E., 1999. User-choices: A new yardstick for the evaluation of ranking algorithms for interactive query expansion. *Inform. Processing Manage.*, 31 (4): 605-620.
8. Chang, Y., C. Cirillo and J. Razon, 1997. Evaluation of feedback retrieval using modified freezing, residual collection & test and control groups. *The SMART retrieval system-experiments in Automatic Document processing*, pp: 355-370.
9. Al-Shalabi, R., G. Kanaan, J. Jaam, A. Hasnah and E. Hilat, 2004. Stop-word removal algorithm for arabic language. *Proceedings of 1st International Conference on Information and Communication Technologies: From Theory to Applications, Damascus*, pp: 545-550.
10. Fowkes, H. and M. Beaulieu, 2000. Interactive searching behavior: Okapi experiment for trec 8. *Proceedings of the BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research, Cambridge*, pp: 47-56.