International Journal of Current Engineering and Technology
ISSN 2277 - 4106
© 2013 INPRESSCO. All Rights Reserved.
Available at http://inpressco.com/category/ijcet

Research Article

Performance and Effectiveness Examination of the IQE and AQE with Application on Arabic Content

Abdel Salam Obeidat[a], Sameh Ghwanmeh[a*], Ali Al-Ibrahim[a], Nidhal K. El-Omari[a] and Adel Mohammad[a]

[a]WISE University, Faculty of Information Technology, Amman, Jordan

Accepted 01 May 2013, Available online 01 August 2013, **Vol.3, No.3 (August 2013)**

Abstract

Literature search show that information retrieval (IR) systems on Arabic are little compared with English language. Additionally, IR systems face many problems when used with Arabic language, including, complexity and ambiguity. The performance and effectiveness of interactive query expansion (IQE) and automatic query expansion (AQE) represent a key towards improved IR systems. The performance and effectiveness of IQE compared with AQE have been examined via a series of search experiments on Arabic content. Compared with no query expansion, the experimental results showed that AQE provides enhanced performance and effectiveness, with 54% query improvement and average precision of 42.1. However, results revealed that IQE provides high performance and effectiveness compared with AQE, with 84% query improvement and average precision of 43.4.

Keywords: *Arabic language, Arabic content, evaluation of information retrieval, query expansion, IQE, AQE.*

1. Introduction

Searching and Retrieving Information is defined as the process of finding the relevant documents in a specific database based on the user requests. Statistical and other common methods have been used to implement the information retrieval (IR) systems. Usually, such IR systems contain select terms (phrases and others) from the searched documents and an indexed file to enable easy access to documents. The search process can be effective if the searched results contain the maximum number of related documents and minimum number of non-related documents (Hani. S et al,2006; Chang, Y.C et al,1997).

Adding additional query terms to existing query is called a query expansion and it is used to enhance the performance and correctness of the search process. This can be achieved by employing the user's related information created from the user assessment to the relevant documents. The expansion terms can be determined and ranked from those documents (Strzalkowski T. et al,1998). The created expansion terms can be used in the query by IQE or AQE (Al-Kharashi et al,1994). It is the user decision to select the appropriate terms in the expansion query (Robertson E, 1990; Fowkes, H et al,2000).

Arabic Language is a common language in the world. IR applications on Arabic are little compared with English language IR applications. Additionally, there is a shortage in the large test databases deals with Arabic content. IR systems face many problems when used with Arabic language, including, Orthographic variations, complexity, broken plurals, ambiguity, short vowels (Moukdad, H., 2001; Ruthven, I., 2006; Beaulieu, M,2004).

In this paper, the performance and effectiveness of IQE and AQE have 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 and IQE techniques. Performance evaluation process has been performed to reveal the best value of n in AQE that gives the optimal value of average precision for the whole query.

2. Research Methodology

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 offer a list of probable expansion terms. The wpq method (Al-Kharashi,1994; Efthimiadis, E., 1999) of ranking terms for query expansion has been employed in this research; this has been shown to provide acceptable results for both AQE and IQE as explained in (Strzalkowski, T,1998, Noaman, A,2012; Jinxi, X.,2002).

As the main objective of this research is to measure the performance and effectiveness of the IQE and AQE based on Arabic documents, the research methodology employs the algorithm presented in (Al-Kharashi,1994) for each query. The recall and precision values are calculated using a full-freezing method which is a standard method of

*Corresponding author: Sameh Ghwanmeh

measuring the performance of the query expansion technique (Koenemann, J.,1996; Magennis, M,1997; Efthimiadis, E., 1999).

Table 1: Arabic collection details

Number of documents	242.0
Number of queries used	14.0
Average words per query	2.40
Average number of relevant documents per query	43.4

The system is characterized as a computationally-intensive. That is why up to 15 terms for query expansion have been used. As shown in (Al-Kharashi,1994; El-Shishtawy,2012), the 15 terms are considered to be the finest for the query expansion. The proposed IQE/AQE system has been implemented using Microsoft Visual Basic 6.0. The system employs an executable code to remove the stop words in the normalization process (Jinxi, X,2002). AQE (collection dependent), AQE (query dependent) and IQE have been implemented.

3. System Evaluation

The system performance and its quality have been examined using an Arabic collection. A detailed performance comparison is conducted between IQE and AQE. For AQE, as explained in (Al-Kharashi,1994)and showed in Eq. 1, the optimal value of expansion terms, n, is determined based on the average precision (Eq. 1) calculated for each expanded query used per each collection. The n value ranges 1 to 15. Results revealed that the value n for AQE which provides the optimal value of the precision is 1. Average precision is calculated using Eq. 1 given by (Al-Kharashi,1994; Al-Shalabi,2004):

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

Where:

$\bar{P}(r)$ = Average precision at recall level r,
Nq = the number of queries used;
Pi(r) = the precision at recall level r for ith query.

4. Experimental Results and Discussion

Series of experimental tests have been conducted to measure the performance and effectiveness of IQE and AQE when applied to Arabic collection. Additionally, AQE has been compared with no query expansion. Detailed results of the comparison can be seen in Table 2. It can be seen from Table 2, that AQE outperforms the no query expansion.

Compared with no query expansion, results show that AQE provides enhanced performance and effectiveness, with 54% query improvement and average precision of 42.1. However, based on 15 expansion terms, results revealed that IQE provides high performance and effectiveness compared with AQE, with 84% query improvement and the highest average precision of 43.4 and high stability.

Table 2: Comparison between IQE and AQE when applied to Arabic

Query type	**Arabic collection % improvement**	**Arabic collection Average Precision**
No query expansion	-	39.1
AQE (collection dependent)	54%	42.1
IQE best	84%	43.4
IQE worst	0%	18.7

Conclusion

The performance and effectiveness of the IQE compared with the AQE have been examined via a series of search experiments on Arabic content. The experimental results revealed that the IQE technique is more effective. The results showed that the IQE is tending to detect the expansion terms more than AQE, especially with complex searches applied on Arabic content. The later result is comparable to the findings presented by (Abuali, A,2010).

Compared with no query expansion, the experimental results showed that AQE provides enhanced performance and effectiveness, with high query improvement and average precision. However, results discovered that IQE provides improved performance and effectiveness compared with AQE.

Acknowledgements

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

References

Hani, S. and Philip-Fai, C., 2006. English-Arabic Cross-Language Information Retrieval Based on Parallel Documents. *International Journal of Computer Processing of Oriental Languages*. 19(1): 21-37.

Strzalkowski, T., J. Wang and B. Wise, 1998. Summarization-based Query Expansion in Information Retrieval. *Coling*-ACL, pp: 1258-1264.

Al-Kharashi, I. and Evens, M., 1994. Comparing Words, Stems, and Roots as Index Terms in an Arabic Information Retrieval System. *Journal of the American Society for Information Science*, 45(8): 548-560.

Robertson, E., 1990. On term selection for query expansion. J. Documentation, 46 (4): 359-364.

Moukdad, H., 2001. Information Retrieval from Full-Text Arabic Databases: Can Search Engines Designed for English Do the Job. *International Journal of Libraries & Information Services*, 51(2): 63-74.

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, 213-220.

Noaman, A. and Al-Ghuribi, S., 2012. A new approach for Arabic text classification usinf light stemmer and probabilities. *International Journal of Academic Research*, 4(3): 114-121.

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.

Jinxi, X., Fraser, A. and R. Weischedel, 2002. Empirical Studies in Strategies for Arabic Retrieval.

Magennis, M., and van Rijsbergen, C. J, 1997. The potential and actual effectiveness of interactive query expansion. Proceedings of the 20th International ACM *SIGIR Conference on Research and Development in Information Retrieval*. pp 324-332.

Beaulieu, M. Experiments with interfaces to support query expansion, 2004 *Journal of Documentation*. 53. 1. pp 8-19.

Abuali, A., Al-Zoua'bi, L., Abu-Addose, H., Hamam, S., 2010. Bilingual Text-Based Image Retrieval Using PDA's (BTBIRUP). *International Review on Computers & Software.* 5(3): 302-308.

Baeza-Yates, R. and B. Ribeiro-Neto, 1999. Modern Information Retrieval. *Addison-Wesley*, pp: 178-212.

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.*

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.

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.

Fowkes, H. and M. Beaulieu, 2000. Interactive searching behaviour: Okapi experiment for trec 8. Proceedings of the *BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research*, Cambridge, 47-56.

El-Shishtawy, T. and El-Ghannam, F., 2012. An Accurate Arabic Root-Based Lemmatizer for Information Retrieval Purposes. *International Journal of Computer Science Issues (IJCSI)*, 9(1): 58-66.