Query Length and its Impact on Arabic Information Retrieval Performance

Suleiman Mustafa and Reham Bany-Younis

Yarmouk University, Jordan

smustafa@yu.edu.jo

ABSTRACT

This paper reports the results of investigating the impact of query length on the performance of Arabic retrieval. Thirty queries were used in the investigation, each of which was phrased in three different types of length: short, medium, and longer, giving ninety different queries. A Corpus of one thousand documents on herbal medication was used and expert judgments were used to determine document relevance to each query. The main finding of this research is that using shorter queries improves both precision and recall. Due to the absence of other results to compare with and the lack of agreement on how length affects retrieval, it has been concluded that the results should be viewed in light of the type of dataset used and how queries were formulated and categorized.

KEYWORDS: Arabic Information Retrieval, Query Length, Retrieval Performance

1. INTRODUCTION

Arabic language is ranked as the seventh largest language on the Internet. However, it has been the fastest growing language in the last decade in terms of users. Given the current growth rate of Internet penetration among the Arabic speaking population, Arabic users should have the fourth largest user population on the Internet by 2020 (Darwish, 2014). This gives a special importance to the language and emphasizes the need for effective IR approaches for enabling effective search of Arabic documents.

The major issue addressed by researchers and practitioners in the field of information

retrieval (IR) has not changed since the sixties of the last century. It is locating the relevant information in the available data sources within an acceptable amount of time. What has primarily changed since then is that the volume of digital information available online has increased exponentially.

With the invention of the Internet and the web, search engines have introduced new ideas and techniques for improving the process of retrieving information. Recall and precision have been always the main measures of assessing the performance of information retrieval systems. The main challenge of these systems lies in the fact that the searching takes place in sources of information that are characterized by unstructured nature since they usually come in the form of natural language text. A considerable amount of research efforts have been devoted to developing more powerful IR models, such as the Boolean model, the Space Vector Model (SVM), and the probabilistic model (Raman et al., 2012; Saini et al, 2014).

However, while a large body of research has accumulated over the last five decades which addresses the issues of retrieving information from English documents, little attention has been directed towards Arabic IR until recently. According to Moukdad, *(2004*), economic, cultural and security issues are all contributing to the international interest in the development of Arabic-supporting technologies. A number of techniques and algorithms have suggested in the literature for dealing with Arabic documents, especially in the area of term conflation and indexing. With the increasing growth of Arabic documents on the Web, several data mining techniques have also been tested and their results reported in the literature.

Each language has its own characteristics that affect how information and queries are represented in the IR system. How an information need, which represents a cognitive state of some IR user, is expressed in the form a query in natural language and how it is matched with the index terms of the IR system play a major role in the performance IR systems. But, despite the differences between various languages and various IR models, every IR system should maintain three basic operations: document representation, query formulation, and a process for matching queries with documents (Hiemstra et al., 2009).

Different users represent their information requests in different ways, but may be looking for the same information or have the same information need. Assume, for example, that several users may be interested in searching for information about the "causes of cancer". One user might represent it as "cancer", another as "cancer disease", a third as "causes of cancer", and a fourth as "causes of cancer disease". According to the process of query matching in the IR system, we expect that the four query forms will provide different results. Even if the users know how to express their information needs in natural language they frequently face a “language” barrier while trying to convert this knowledge into a few exact terms to formulate an adequate search query (Taksa & Amanda, 2007).

Query length has generally been regarded as a query-specific constant that does not affect document ranking (Lv, 2015). On the other hand, it has been reported that longer queries are more effective at retrieving useful documents than short keyword queries (Belkin et al., 2002). In many cases longer queries may be desirable either to improve recall or to refine results in topics with many documents Agapie et. al (2013). According to Gupta & Bendersky (2015), effective handling of verbose queries has become a critical factor for adoption of information retrieval techniques. Over the past decade, the information retrieval community has deeply explored the problem of transforming natural language verbose queries using operations like reduction, weighting, expansion, reformulation and segmentation into more effective structural representations.

In comparison, short queries usually lack sufficient information regarding the user intent (Wu et al., 2014) in comparison with longer queries which would provide more information in the form of context (Kraft et al., 2006). According to Agapie et. al (2013), although longer queries can produce better results for information seeking tasks, people tend to type short queries. Some studies confirm users’ persistence in using short queries (Jansen & Spink, 2006; Spink et al., 2001).Overwhelmed by the task of sifting through a massive volume of returned search results, users frequently examine just the first page (top 10 results) and continue with a new, reformulated, yet still short query,

hoping to find more relevant results (Taksa & Spink, 2007). To deal with this limitation of short keyword queries, several query expansion techniques have been proposed and investigated in the literature. Given an arbitrary short query, the goal of query expansion is to find and include additional related and suitably-weighted terms to the original query to produce a more effective version (Kumaran & Allan, 2009).

In this paper, we examine the effect of using queries of different lengths on the performance of Arabic Information retrieval as measured by recall and precision. For this purpose, query length has been defined as the number of terms used in a query. The rest of this paper is organized as follows: section 2 presents the related research works with special focus on the effect query length on IR performance, section 3 presents the experimental setup, section 4 presents the analysis and discussion of the research results, and finally section 5 provides some concluding remarks.

2. RELATED WORK

2.1 Query Expansion

Many query expansion approaches have been attempted. Keyword-type approaches are the predominant type. Four basic types of keyword approaches have been reported. (a) Human-in-the-loop approach which uses multiple-query searches that are built manually (Carpineto & Romano 2001). (b) Thesaurus-based approach which is utilized for automatic query expansion (Zohar et al., 2013; Imran & Sharan, 2009; Voorhees, 1994). (c) Pseudo-relevance or blind feedback approach (Chinnakotla et al. 2010; Abderrahim, n.d.). (d) Relevance feedback approach using human-in-the-loop which expands queries using words from the user identified top retrieved documents (Jin et al., n.d.).

Walker (2001) divided the query expansion technique into three types as follows: (1) human and computer generated thesauri, (2) relevance feedback, and (3) automatic query expansion, The weaknesses and strength of each type were highlighted and examined. The conclusion showed that automatically expanded queries via pseudo-relevance feedback and computationally generated thesauri addressed the needs of users, but have not improved effectiveness of search engines beyond that encountered in relevance feedback.

Liebeskind et al.(2013) described three ways for building the thesaurus automatically, including statistical co-occurrence analysis, the concept space approach, and the representation of terms as Bayesian networks. The researcher described how the three approaches work, but did not show how these approaches affect retrieval performance.

As of Arabic, several techniques for expanding queries in Arabic have been reported in the literature. Mahgoub et al. (2014) introduced a query expansion approach using an ontology built from Wikipedia pages in addition to other thesaurus to improve search accuracy for Arabic language. This approach outperformed the traditional keyword based approach in terms of both F-score and NDCG measures.

Shaalan et al., (2012) introduced what they described as Expectation Maximization (EM) algorithm based on term similarity. They employed the EM algorithm in the process of selecting relevant terms for enlarging the query and removing the non-related terms. The main finding of this research was increasing recall while keeping the precision at the same level by this method.

Thesaurus-based expansion was reported by Khafajeh & Yousef (2013). The aim was to design and build an automatic Arabic thesaurus using Local Context Analysis technique that can be used in any special field or domain to improve the expansion process and to get more relevance documents for the user's query. According to the authors, this technique can be used in any special field or domain to improve the expansion process and to get more relevant documents for the user's query. Results of this study were compared with the classical information retrieval system.

2.2 Query Length

Lv and Zhai, (2011) presented how to adapt TF normalization to the global distribution of individual terms in the whole corpus. But, since their results were not conclusive, Lv (2015) reexamined later the issue of TF normalization with respect to query length. As he states, although many studies have attempted to improve retrieval models from various perspectives, they mainly focused on either document properties (e.g., term frequency and document length) or the estimation of term statistics from the corpus (e.g., IDF and background model), while little attention has been paid to the query

length. He adds, query length has generally been regarded as a query-specific constant that does not affect document ranking. He based his research on the assumption that query length actually interacts with term frequency (TF) normalization, which is a key component of all effective retrieval models. Specifically, the longer the query is, the smaller the TF decay speed should be. In other words, the longer the query is, the less penalization the repeated term occurrences should receive. A document that mismatches a query term from a longer query should not be penalized as much as that of another document that mismatches the query term from a shorter query. Using a formal query length constraint and a query-length aware heuristic, Lv's results provided evidence that query length affects TF normalization.

The interaction between query length and document length normalization has also been investigated by many researchers. Chung et al. (2006) have incorporated the query length into the pivoted document length normalization in the space vector model, but they did not reach any conclusive results. Cummins et al. (2012) examined a similar heuristic, and formalized a constraint to describe the heuristic, but their method only performed comparably to the baseline retrieval models. Moreover, both works focused only on document length normalization, but did not pay attention to the effect of query length.

Kumaran et al., (2009) presented a technique to reduce long queries to more effective shorter ones that lack extraneous terms. Their work was motivated by the observation that perfectly reducing long TREC description queries can lead to an average improvement of 30% in mean average precision. The researchers' approach involved transforming the reduction problem into a problem of learning to rank all sub-sets of the original query (sub-queries) based on their predicted quality, and select the top sub-query. They used different measurements of query quality described in the literature as features to represent sub-queries, and train a classifier. Replacing the main long query with the top-ranked sub-query chosen by the ranking classifier could result in a statistically significant average improvement of 8%. Analysis of the results revealed some interesting properties of long queries such as the dependence of the utility of query reduction on the quality of the original long query. By showing significant improvement in performance using easily computed features, they paved the way for

easy adoption of query reduction.

Belkin et al., (2003) tested the effectiveness and usability of increasing the length of initial search query length using a relatively simple interface query elicitation technique and examined the relationship between query length and search effectiveness in web-based interactive IR. The results indicated that this technique was at least as usable as the baseline technique. Longer queries, irrespective of query elicitation mode, are significantly associated with increased searcher satisfaction with search results, and longer initial queries lead to equivalent "objective" results. The study concluded that longer searcher queries result in increased search effectiveness in general, indicating that more words from the searcher describing the person's information problem results in better interactive IR performance.

Kumaran et al., (2007) developed a suite of automatic techniques to re-write queries and study their characteristics. They showed that rewriting the query to a version that comprises a small subset of appropriate terms from the original query greatly improves effectiveness of almost 50% on some difficult TREC queries and their associated collections. They showed that the shortcomings of automatic methods can be ameliorated by some simple user interaction, and report results that are on average 25% better than the baseline. The results clearly showed that shorter reformulations of long queries could greatly impact performance. The researchers believe that their technique has great potential to be used in an adaptive information retrieval environment, where the user starts off with a more general information need and a looser notion of relevance. The initial query can then be made longer to express a most focused information need.

Wu et al., (2014) addressed the issue of improving search relevance for short queries in community question answering. The researchers mentioned that existing methods mainly focus on long and syntactically structured queries. And when an input query is short, the task becomes challenging, due to a lack of information regarding user intent. They tested different types of user intent from various sources for short queries. With these intent signals, they proposed a new intent-based language model. The model takes advantage of both state-of-the-art relevance models and the extra intent information mined from multiple sources. The evaluation results showed that, with user intent

prediction, this proposed model can significantly improve state-of-the-art relevance models on question retrieval for short queries.

Phan et al. (2007) hypothesized that the degree of specificity of an IR request might correspond to the length of a search query. The results showed a strong correlation between decreasing query length and increasing broadness or generality of the IR request. They found that an average cross-over point of query specificity from broad to narrow is about 3 words. The experiments which carried out partially validate this theory (for broadness), and indicate a cross-over of 3 words for queries being 2-3 times more frequently about narrow than broad information needs. These results have implications for search engines in responding to queries of differing lengths.

Mu et al., (2012) studied how the specificity of a query term and query length influence the effectiveness of medical information retrieval. Experiment demonstrated that term expansion on longer and more specific queries based on the Unified Medical Language System outperformed shorter and less specific queries; and expansion using the string index method achieved better performance than using the word index method. In addition case analysis revealed that for more specific queries, when using the string index method, the positive factor of adding useful related terms outweighed the negative factor of over expanding irrelevant terms.

3. RESEARCH METHODOLOGY

3.1 Corpus

The absence of standard comprehensive corpus is considered one of the main challenges that face most researchers in the field of Arabic information retrieval. As the case with many other research studies, this study was based on a manually collected dataset. The dataset consisted of one thousand documents collected from a number of books on herbal medication. Each document is devoted for a given medical herb and includes a description of what diseases it can be for. A special feature of this dataset is that authors of these books included indexes providing a list of diseases and the herbs that can be used of treatment. This gave the study a reliable took for expert judgment when evaluating retrieval results. The documents differ in length ; they range from one page to about six pages.

3.2 Text Preprocessing and Indexing

Document preprocessing involves a number of tasks that aim to facilitate the information retrieval process. These tasks include: tokenization, removing diacritics, removing non-textual words and special symbols, eliminating stop words using the list prepared by Khoja (2001), and letter normalization. This was applied to both corpus documents and to the set of queries used for evaluating retrieval performance. To construct the index, tokens were stemmed using the light stemmer described by Mustafa (2012). The same procedure was also applied to each query.

3.3 Queries Set

A set of 90 queries was used for evaluating the effect of query length on Arabic information retrieval. The queries were categorized into three classes, each having 30 queries: short queries consisting of one or two words, medium-size queries consisting of three to five words, and longer queries consisting of more than five words. To provide a basis for comparison, each query was represented in three forms: short, medium, and long. This means that we have basically 30 queries represented in three versions in accordance with the objective of this study which is investigating the effect of query length on the performance of Arabic information retrieval. This means also that the same set of relevant documents for a short query is the same for the corresponding two versions: medium and longer. Table 1 presents the three query sets used in the study

Table (1): Query Categorization Based on Length

Category	# Terms	Example	# Queries
Short	1-2	عسر الهضم	30
Medium	3-5	اعراض عسر الهضم	30
Long	> 5	أعراض عسر الهضم بعد تناول أي وجبة سواء في الإفطار أو الغداء	30

3.4 Performance Evaluation

To achieve the objectives of this study, a simple retrieval system was developed using Visual Studio 2010 and SQL MS SQL Server. Based on the categorization scheme described above, each query was executed and its results were collected and compared with the set of relevant documents as determined by human judgment. The well-known three retrieval performance measures were used to determine the effect of query length: average recall, average precision, and interpolated F-measure.

4. RESULTS AND DISCUSSION

Table 2 presents the results of executing the set of queries as described above. The results indicate the general inverse relationship pattern known for recall and precision. While recall is relatively high for all three types of queries ranging between 0.75 and 0.84, precision is considerably low ranging between about 0.48 and 0.55.

However, as we consider the average recall results as presented in Figure 1, we notice that as the query gets shorter, we get more relevant documents from the dataset. On the other hand, we notice the same trend when we view the results of precision as shown in Figure 2. This means that the ratio of relevant documents in the set of documents retrieved is expected to increase in the manner as recall. Since F-measure is a tradeoff between recall and precision, we note that this measure gives higher values for shorter queries.

Table (2): Results of Average Recall, Precision, and F-Measure

Experiment Type	Recall	Precision	F-Measure
Exp. 1: Using Short queries	84.06	55.48	65.19
Exp. 2: Using medium queries	80.92	52.65	61.81
Exp. 3: Using long queries	76.42	48.10	56.69

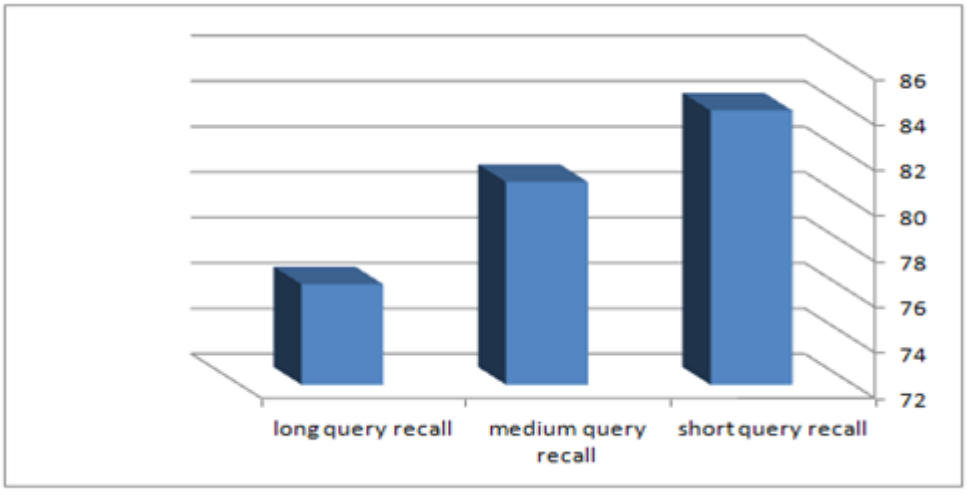

Figure (1): Average Recall Ratios (over 90 Queries)

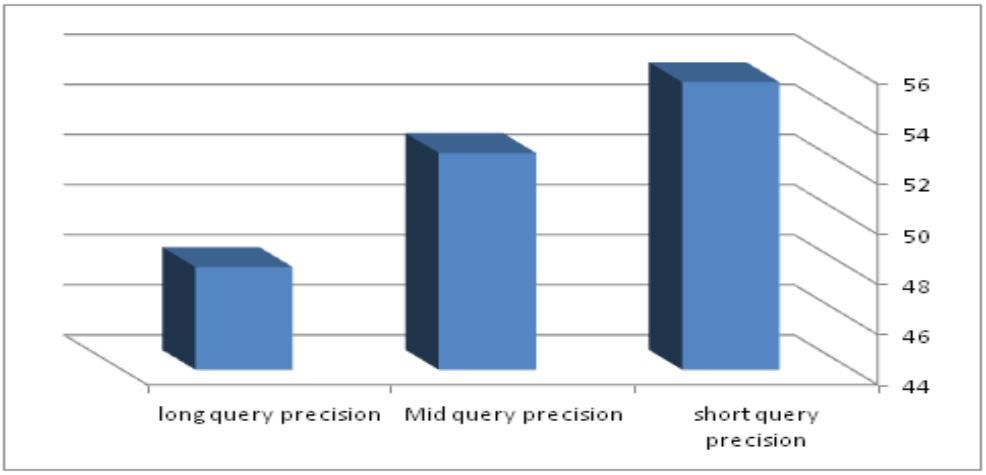

Figure (2): Average Precision Ratios (over 90 Queries)

There are three assumptions concerning query length in the literature. The first assumption says that short keyword queries usually lack sufficient information regarding the user intent (Wu et al., 2014) which might affect retrieval performance negatively in terms of recall values in comparison with longer queries. It has been reported that longer queries are more effective at retrieving useful documents than short keyword queries (Belkin et al., 2002). That is why researchers introduced the idea of query expansion using different techniques.

The second assumes that long descriptive queries usually involve some extraneous terms that do not add to the semantic content of a given query (Kumaran et al., 2009), which might introduce a negative effect on retrieval performance. While the first assumption stands behind the idea of query expansion, the second has lead to the idea of query reduction. Some researchers have attempted to make long queries shorter on the assumption that longer queries usually bring with them extraneous terms that lead to lower precision values (Kumaran et al., 2009; Agapie et. al 2013).

Yet, a third assumption says that query length is generally regarded as a query-specific constant that does not affect document ranking (Lv, 2015). Given these assumptions, the first observation about the results of this study is that they do not seem to support any of these assumptions. It is apparent that query length has an effect on the retrieval performance. Short queries have exhibited better recall and precision values than longer queries. The second observation is that the difference between the performance values of the three categories of query length is relatively not significant. We are talking about a difference of ≈5% improvement which should not be regarded a high value. One might say that the kind of queries and the dataset used might impact the kind of results obtained.

5. CONCLUSION

The purpose of this study was to investigate the impact of query length on recall and precision of Arabic information retrieval. The results indicate that both precision and recall improve when we use shorter queries. This general result should be considered in view of the dataset used and the type of queries used. Hence, it might be appropriate to suggest that the issue of query length be reinvestigated using other datasets and maybe with a different scheme of length categorization.

References

Abderrahim, Mohammed E. (n.d.). Concept Based vs. Pseudo Relevance Feedback Performance Evaluation for Information Retrieval System. Available at: https://arxiv.org/ftp/arxiv/papers/1403/1403.4362.pdf Retrieved 17/July//2016.

Agapie, Elena; Golovchinsky, Gene; Qvarfordt, Pernilla (2013). Leading People to Longer Queries. CHI 2013, April 27–May 2, 2013, Paris, France.

Belkin, N.J., Kelly, D., Kim, G., Kim, J.Y., Lee, H.J., Muresan, G., Tang, M.C., Yuan, X.J. and Cool, C., 2003. Query length in interactive information retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 205-212

Belkin, N.J., Cool, C., Kelly, D., Kim, G., Kim, J.-Y., Lee, H.-J., Muresan, G., Tang, M.-C., Yuan, X.-J. (2002) Query Length in Interactive Information Retrieval. In Proc. SIGIR 2003 (Toronto, Ont, Canada). ACM Press.

Carpineto, C., De Mori, R., Romano, G., and Bigi, B. (2001). An information theoretic approach to automatic query expansion. ACM Trans. Info. Syst. pp. 19, 1, 1–27

Chung, T.L., Luk, R.W.P., Wong, K.F., Kwok, K.L. and Lee, D.L., 2006. Adapting pivoted document-length normalization for query size: Experiments in chinese and english. *ACM Transactions on Asian Language Information Processing (TALIP)*, *5*(3), pp.245-263.

Chinnakotla, M.K.; Raman, Karthik; Bhattacharyya, Pushpak (2010). Multilingual Pseudo-Relevance Feedback: Performance Study of Assisting Languages. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 11-16 July 2010. c 2010 Association for Computational Linguistics, pp. 1346–1356,

Cummins, R. and O'Riordan, C., 2012. A constraint to automatically regulate document-length normalisation. In *Proceedings of the 21st ACM international conference on Information and knowledge management* , pp. 2443-2446.

Darwish, K. and Magdy, W., 2014. Arabic information retrieval. Now Publishers.

Farghaly, A. and Shaalan, K., 2009. Arabic natural language processing: Challenges and solutions. ACM Transactions on Asian Language Information Processing (TALIP), 8(4), p.14

Gupta, M. and Bendersky, M., 2015. Information Retrieval with Verbose Queries. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* pp. 1121-1124.

Hiemstra, D., 2009. Information retrieval models. *Information Retrieval: searching in the 21st Century*, pp.2-19.

Imran , Hazra & Sharan, Aditi (2009). Thesaurus and Query Expansion. International Journal of Computer science & Information Technology (IJCSIT), 1(2): 89-97.

Jansen, B.J. and Spink, A.(2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs," Information Processing and Management, Vol. 42(1), pp. 248–263

Jin, Xiangyu; French, James; Michel, Jonathan (n.d). Query Formulation for Information Retrieval by Intelligence Analysts. Retrieved: July,19, 2016. Available at: http://www.cs.virginia.edu/~xj3a/research/publications/TR05.pdf

Khafajeh, H., Yousef, N. and Kanaan, G. (2010). Automatic query expansion for Arabic text retrieval based on association and similarity thesaurus. In *Proceedings he European, Mediterranean & Middle Eastern Conference on Information Systems (EMCIS), Abu Dhabi, UAE*.

Khafajeh, Hayel and Yousef, nidal (2013). Evaluation of Different Query Expansion Techniques by using Different Similarity Measures in Arabic Documents. IJCSI International Journal of Computer Science Issues, 10 (4), 160-166.

Khoja, S. (2001). APT: Arabic Part of Speech Tagger. Proceedings of the Student Workshop at NAACL, pp. 20-25.

Kumaran, G. and Allan, J. (2007). A Case For Shorter Queries, and Helping Users Create Them. In *HLT-NAACL* , pp. 220-227.

Kumaran, G. and Carvalho, V.R. (2009). Reducing long queries using query quality predictors. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 564-571.

Liebeskind, Chaya Dagan, Ido; Schler, Jonathan (2013).Semi-automatic Construction of Cross-period Thesaurus. Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pp. 29–35, Sofia, Bulgaria, August 8 2013. c 2013 Association for Computational Linguistics

Lv, Y. and Zhai, C. (2011). When documents are very long, BM25 fails!. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* pp. 1103-1104.

Lv, Y. (2015). A Study of Query Length Heuristics in Information Retrieval. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1747-1750.

Mahgoub, Ashraf Y.; Rashwan, Mohsen A.; Raafat, Hazem; Zahran, Mohamed A.; Fayek, Magda B. (2014). Semantic Query Expansion for Arabic Information Retrieval. Association for Computational Linguistics, October 25, 2014, Doha, Qatar. c 2014 pp. 87-92.

Moukdad, H. (2004). Lost in cyberspace: How do search engines handle Arabic queries. In the 32nd Annual Conference of the Canadian Association for Information Science , pp. 1-7.

Mu, X. and Lu, K. (2012). Improving UMLS Metathesaurus Query Expansion Based on the Query Specificity and Length.

Mustafa, S.H. (2012). Word stemming for Arabic information retrieval: The case for simple light stemming. Abhath Al-Yarmouk: Science & Engineering Series, 21(1), p.2012.

Phan, Nina; Bailey, Peter; Wilkinson, Ross (2007). Understanding the Relationship of Information Need Specificity to Search Query Length. SIGIR'07 Proceedings, July 23–27, 2007, Amsterdam, The Netherlands. ACM 978-1-59593-597-7/07/0007, pp. 709-710.

Raman, S., Chaurasiya, V.K. and Venkatesan, S. (2012). Performance comparison of various information retrieval models used in search engines. In Communication,

Information & Computing Technology (ICCICT), 2012 International Conference on , pp. 1-4.

Reiner Kraft, Chi Chao Chang, Farzin Maghoul, and Ravi Kumar. 2006. Searching with context. In 15th International Conference Proceedings, pp. 477–486.

Saini, B., Singh, V. and Kumar, S. (2014). Information retrieval models and searching methodologies: Survey. Information Retrieval, 1(2), p. 20.

Shaalan, K., Al-Sheikh, S. and Oroumchian, F. (2012). Query expansion based-on similarity of terms for improving Arabic information retrieval. Springer Berlin Heidelberg. InIntelligent Information Processing VI, pp. 167-176.

Spink, A., Wolfram, D., Jansen, B.J., and Saracevic, T. (2001). The public and their queries," Journal of the American Society for Information Science and Technology, Vol. 52(3), pp. 226–234

Taksa, Isak and Spink, Amanda H. (2007). Evaluating Usability of a Long Query Meta Search Engine. In Sprague, S., Eds. Proceedings 40th Annual Hawaii International Conference on System Sciences, 2007. (HICSS 2007), Hawaii, USA., pp. 1-9

Voorhees. E.M. (1994). "Query expansion using lexical-semantic relations", In Proceedings of the 17th ACM-SIGIR Conference, pp. 61-69.

Walker, D. (2001). Query expansion using thesauri: Previous approaches and possible new directions. University of California, Los Angeles.

Wu, H., Wu, W., Zhou, M., Chen, E., Duan, L. and Shum, H.Y. (2014). Improving search relevance for short queries in community question answering. In Proceedings of the 7th ACM international conference on Web search and data mining, pp. 43-52.

Zohar, H., Liebeskind, C., Schler, J. and Dagan, I. (2013). Automatic thesaurus construction for cross generation corpus. Journal on Computing and Cultural Heritage (JOCCH), 6(1).