An Efficient Ranking Module for an Arabic Search Engine

Esraa Abd Elraouf **Nagwa Lotfy Badr** **Mohamed Fahmy Tolba**

Faculty of Computers and Information Sciences Ain Shams Unversity

Summary
Searching online provides you with a wealth of information, but not all of it will be useful or of the highest quality. Search engines are distributed programs that dive into the World Wide Web to find relevant information for a given search query. Their fundamental components are: the crawlers, the indexer module, the collection analysis module, the query engine, and the ranking module. Many of today's search engines use a traditional text process to retrieve pages related to a user's query. Traditional text processing is done to find all documents using the query terms, or related to the query terms by semantic meaning. With the massive size of the web, this result in thousands of retrieved pages may or may not related to the query. The main function of the ranking module is to sort the search results by relevance or importance using information retrieval (IR) algorithms. There were two kinds of methods in information retrieval, based on content and based on hyper-link. The quantity of computation in systems based on content was very large and the precision in systems based on hyper-link only was not ideal. It was necessary to develop a technique combining the advantages of two systems [3]. Many web users are interested in Arabic web browsing whether the reason is academic or commercial… suffer to find their search and request over the Arabic search engine etc. As the existing web search engines are designed to perform English web searches. They don't generate morphological variations of Arabic words but they just match the word as it is. Therefore their results contain only the pages that exactly match the user query. They also don't consider the different meanings of a word so search results contain unrelated pages to user query. In this research, we focus on implementing an enhanced ranking algorithm by combining both the page content and the Hyper-Link with the focus on Arabic search engines by taking into account the stem and the context of the Arabic word by combining both the count of words related to query in the page and the count of words related to query in outlinks pages of that page to calculate its rank, using external database having the morphological meanings of the most Arabic words. Then sort the pages according its rank. If there is more than one meaning to an input query word in case the user does a query in using only one word, the user may choose the meaning he/she wishes to search for. The search results will largely contain the inflected forms of the word that belong to that meaning. This helps reduce the redundancy that results from morphological search only.

Keywords:
Morphological Variations, Information Retrieval, Ranking, Arabic Search Engine

1. Introduction

Recent studies show that search engines play an increasingly important role in people's surfing of the web; when a user wants to look up information from the web, the user often goes to his favorite search engine, issues keyword queries, and clicks on the returned pages. Given the sheer quantity of information available on the web, the widespread use of search engines is not surprising. Because of the Web's size, and the fact that users typically only enter one or two keywords, result sets are usually very large [4]. The ranking module therefore has the task of sorting the results such that results near the top are the most likely ones to be what the user is looking for.

The Query Engine collects search terms from the user and retrieves pages that are likely to be relevant. There are two main reasons why traditional information Retrieval (IR) techniques may not be effective enough in ranking query results. Firstly, the Web is very large, with great variation in the amount, quality and the type of information present in Web pages. Thus, many pages that contain the search terms may be of poor quality or not relevant. Secondly, many Web pages are not sufficiently self-descriptive, so the IR techniques that examine the contents of a page alone may not work well. An often cited example to illustrate this issue is the search for "search engines" [5]. The homepages of most of the principal search engines do not contain the text "search engine". Moreover, Web pages are frequently manipulated by adding misleading terms so they are ranked higher by a search engine (spamming). Thus, techniques that base their decisions on the content of pages alone are easy to manipulate.

The link structure of the Web contains important implied information, and can help in filtering or ranking Web pages. In particular, a link from page A to page B can be considered a recommendation of page B by the author of A. Some new algorithms have been proposed that exploiting this link structure is not only for keyword searching, but also other tasks like building a Yahoo-like hierarchy, or identifying communities on the Web. The qualitative performance of these algorithms is generally better than the IR algorithms since they make use of more information than just the contents of the pages. While it is indeed possible to influence the link structure of the Web locally, it is quite hard to do so at a global level. So link

Manuscript received February 5, 2010
Manuscript revised February 20, 2010

analysis algorithms that work at a global level are relatively robust against spamming [4]. One of the successful and well-published link-based ranking system is PageRank, the ranking system used by the Google search engine. Actually, for pages related to a query, an IR (Information Retrieval) score is combined with a PR (PageRank) score to determine an overall score, which is then used to rank the retrieved pages [6]. These combination of the IR and the PR algorithms proofed it successful and achieving high quality of the best search but on the English –based search only, however the Arabic search still not covered.

Arabic has a unique beauty, poetry and logic. Yet, it is not possible to make easy and meaningful searches in Arabic through standard internet search engines – until now!

Traditional search engines used for English, for example, use wildcards to search for words. This may be appropriate for English, but it is deficient in searching through Arabic text. The reason stems from the complexity of Arabic morphology. While an English word typically has five or six inflections - changes in the form of the word giving it extra meaning - Arabic words can have up to 10,000 inflections. The morphology of a language is a basic component in the way search engines operate. In order to conduct proper online search on Arabic text, it is thus necessary to process the text according to Arabic language's own morphology.

If a user were to search for the word "اجتمـاع" (meeting), the search results would contain various inflected forms, such as "اجتمـع" ([he] met), "يجتمعـــون" ([they] meet), etc. Traditional engines that do not have morphological search might, at most, retrieve words that have "اجتمـاع" (a meeting) as part of the word, such as "اجتمـاعهم" ([their] meeting), "واجتمـاع" ([and] a meeting).

The main goal of this research is to build our own Arabic Search Engine based on ranking system. Because the foreign ones don't support Arabic language well. They don't generate morphological variations of Arabic words but they just match the word as it is. Their results therefore contain only the pages that exactly match the user query. They also don't consider the different meanings of a word so search results contain unrelated pages to user query. Therefore our main aims to develop an efficient ranking algorithm for an Arabic search engine by combining both the count of words related to query in the page and the count of words related to query in its outlinks pages using external database having the morphological meanings of the most Arabic words to produce more relevant pages in search results.

If there is more than one meaning to an input query word, the user may choose the meaning he wishes to search for. The search results will largely contain the inflected forms of the word that belong to that meaning. This helps reduce the redundancy that results from morphological search only.

We implemented our idea on a Prototype by applying our proposed ranking technique for ranking WebPages. We assume an Arabic query word and search in the two famous search engines result then using some crawler we get its outlinks pages then compute the rank of each page result using external database having the morphological meanings of the most Arabic words. We sort the output of WebPages with our new ranking technique.

This paper is organized as follows. In section 2, we present the research architecture.Section3, we discuss our prototype. Section 4 we evaluate our technique comparing with the most better search engines .Section 5 conclude our paper with some predictions about the future of Web IR.

2. Our Proposed Architecture

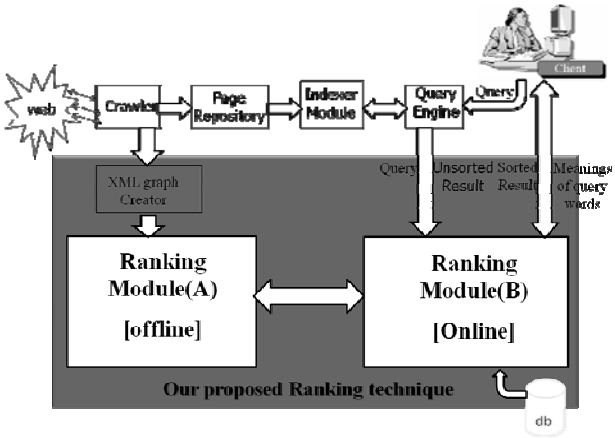

Figure 1. Our proposed Architecture

In our proposed Architecture (Figure1) for the Arabic Search Engine, the Crawler downloading the web pages. It browses the web using a starting set of URLs to be fetched to the crawler. The web pages that are fetched are then sent to Page Repository. While crawling the web the XML graph creator create xml file that have each downloaded page stored with its ID number that assigned to it using hash key function and its URL address and for each page save pages that outlinks from it as a child pages. The Indexer module reads the repository and generates a lookup table with all the URLs that point to pages that contain a given word (the text index). When a user enters a query into a search engine (typically by using keywords), the Query engine examines its index and provides a listing of best-matching web pages according to its criteria, usually with a short summary containing the document's title and sometimes parts of the text. While there may be millions of web pages that include a particular word or

phrase, some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to provide the "best" results first. The ranking module sorts the results such that results near the top are the most likely ones to be what the user is looking for. How a search engine decides which pages are the best matches, and what order the results should be shown in, varies widely from one engine to another. The methods also change over time as Internet usage changes and new techniques evolve. There were two kinds of methods in information retrieval, based on both content and based on hyper-link. Because that the quantity of computation in systems based on content was very large and the precision in systems based on hyper-link only was not ideal. In this research, we focus on implementing an enhanced ranking algorithm by combining both the Content Information and the Hyper-Link with the focus on Arabic search engines by taking into account the stem and the context of the Arabic word. This combination of Link and Content Information proofed it successful [3] and achieving high quality of the best search but on the English –based search only, however the Arabic search still not covered.

In our proposed ranking technique, we have two modules. Ranking module (A) that is working offline and Ranking module (B) that is working online. Section 2.1 and 2.2 will discuss both modules with details.

2.1 Ranking Module (A)

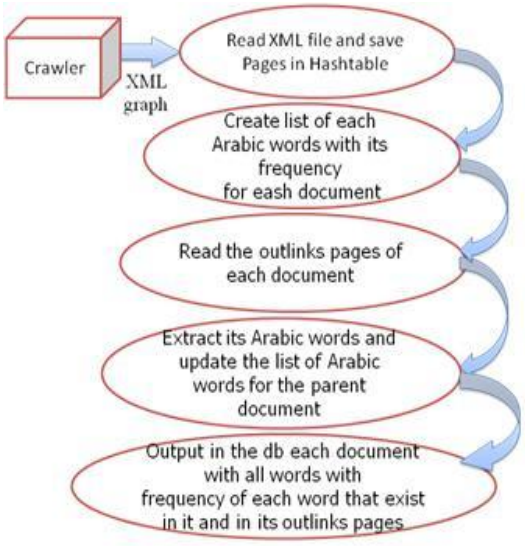

Figure 2. Steps of Ranking Module (A)

Figure 2 Illustrate how Ranking Module (A) works. The Ranking Module (A) works offline.

- It receives the XML file graph from the crawler after XML graph creator creates it. The XML file contains each downloaded page stored with its ID number that assigned to it using hash key function and its URL address and for each page save pages that outlinks from it as a child pages.
- Read the XML file and save parent documents in hash table as an object making its ID number is the key.
- Dividing the document into words and storing all words in the string array.
- Eliminating the stop words-stop words represent nothing for comparing and retrieving information-eliminating also English litters, numbers, and list of non Arabic litters.
- Create hash table for words and save objects that have each word with its frequency and its key is word name.
- Read the outlinks pages of each document.
- Dividing each into words and eliminating also all stop words and all non-Arabic litters.
- Then update hash table of words with those Arabic words, if the word already exist update its frequency, else add new word with a new frequency.
- Now each document in hash table have hash table of all Arabic words with its frequency that exist in it and in it's outlinks documents.
- Then output in database each document with all Arabic words and its frequency in it and in its outlinks pages.

2.2 Ranking Module (B)

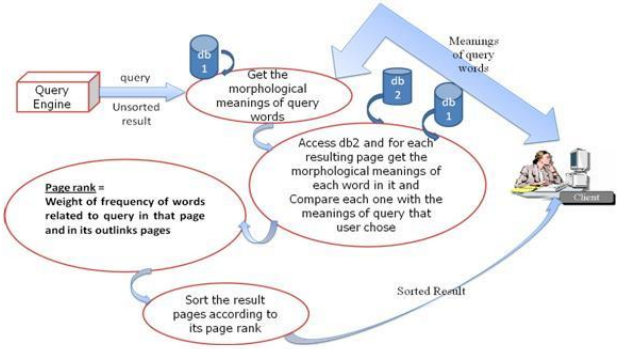

Figure 3. Steps of Ranking Module (B)

Figure 3 Illustrate how Ranking Module (B) work. The Ranking Module (B) works online.

- First, Get the query and unsorted result from the query engine to rank it using our proposed ranking technique.
- Get the morphological meaning of query words using db1 that have the morphological meanings of the most Arabic words and if there is more than one different meaning to an input query word, the user may choose the meaning he/she wishes to search for. The search results will largely contain the inflected forms of the word that belong to that meaning. This helps reduce the redundancy that results from morphological search only.

- Save in the list the query meanings that user chose.
- Access the db2 that output from module (A) and have all web pages downloaded from crawler with all Arabic words with its frequency that exist in it and in its outlinks pages.
- For each resulting page access its Arabic words and save in list of words and compare the meaning of each word with the list of query meanings, for each word in the list of words follow this steps:
  - step1: Get the morphological meanings of the first word and save it in the list of meanings.
  - step2: Compare this list of meanings of word with the list of query meanings.
  - step3:If the word has similar meaning to one of the query meanings, weight the word according to its meanings number from 0-1 (ex. If it have one meaning then its weight equal one and its weight decreasing when its meanings increasing) then add to page rank the word frequency multiplied by its weight.
- The total page rank for each resulting page equal the weight of frequency for each word related to query which exist in the page and in its outlinks pages.
- Then sort the resulting pages according to its rank and show them to the user.

3. The Experiment

We propose a new method to rank the Arabic web sites which are based on morphological meaning of Arabic word combined with the link structure of the web. The suggested study using external database of the morphological meanings of Arabic words for the query word Islamic "إســـلاميات". We retrieve the relevant documents for the suggested query. It chooses two famous search engines to retrieve the relevant documents for the suggested query word. The retrieved documents for each search engine are ranked by using our proposed method which is based on the combination between Page content and link taking into account the stem and the context of the Arabic word. The performance of our proposed method, plus the rank results of each search engines were compared to the number of users ranking which aims to determine the best one. The two selected search engines are Google search engine, and Yahoo search engine.

3.1 Pre-process steps

We did some pre-processing steps in order to retrieve data to use it in our proposed method. These are the pre-processing steps:

-We use The Arabic dictionary which is a comprehensive dictionary of contemporary Arabic (Modern Standard Arabic). That includes up-to-date words used in the various media. This dictionary is based upon the published dictionary "A dictionary of the Contemporary Arabic Language", by Prof. Ahmed Mokhtar Omar[13], the late renowned Arabic lexicographer.

• Write any query word such as Islamic "إســـلاميات" in order to retrieve the relevant Arabic documents of each query using Yahoo, and Google.

• Pick up the first thirty documents, which are retrieved by each search engine, and then the retrieved documents are saved as text documents.

-using existing distributed crawlers retrieve for each set of documents its outlinks and save in xml files.

The following steps (i.e enclosed between the BEGIN&END) are proposed in our research which are based on combination between the morphological analysis and the hyperlink structure for ranking Arabic documents for some Arabic domains: Begin

-Reading xml file which represent each document with it's outlinks pages, then dividing the document and it's outlinks documents into words. After that storing all words in the string array which is called split.

-Eliminating the stop words. Stop words represent nothing for comparing and retrieving information.

-Eliminating also English litters, numbers, and list of non Arabic litters.

-Now for each document save all Arabic words with its frequency that exist in it and in it's outlinks documents in the database.

-Show to the user the morphological meaning of query word and let him choose the meaning he need (ex: the query word"إســـلاميات" its morphological meanings are "الإخـلاص" , "الإســلام" , "التســـــليم" or "الخــذلان").

- For each resulting page access the database that save all documents with its Arabic words and get its Arabic words then save them in list of words and compare the meaning of each word with the query meaning that user chose as following steps:

-step1: Get the morphological meanings of the first word and save it in the list of meanings.

-step2: Compare this list of meanings of word with query meaning that user chose.

-step3: If the word has similar meaning to the query meaning, weight the word according to its meanings number from 0-1 (ex. If it have one meaning then its weight equal one and its weight decreasing when its meanings increasing) then add to page rank the word frequency multiplied by its weight.

- The total page rank for each resulting page equal the weight of frequency for each word related to query which exist in the page and in its outlinks pages.
- Then sort the resulting pages according to its rank and display the result to the user.

End

4. Performance results

We implement a software using Visual studio .Net software. The purpose of the system is to write code for our proposed method which is based on the combination between the morphological meanings of Arabic words in the page content and in its outlinks pages, then to apply it on real data to see its performance when it is compared with others. We consider two different search engines to compare between their results with our proposed method for ranking. These search engines are Google search engine, and Yahoo search engine.

An important point that we should mention is the reason for the selection of the first twenty documents. That is in our study we want to notice the effectiveness of our method to other ranking method, which are used by search engines. So, we chose twenty documents for each search engines, because the number of considered documents do not affect on the algorithm performance, and do not effect at the algorithm results. The working at 20, 30, 100, 1000 documents… etc, it will be same approximately.

4.1 Yahoo search engine results

Table 1 in the next page shows the performance results for Yahoo search engine. The table shows the performance results of comparison between our proposed method, and five interested users in the query word. It also shows comparison between Yahoo search engine and ranking of the five users. The table was divided for six parts: the first part represents the document ranking position. The second part of the table shows the users ranking for the first twenty documents which were retrieved by Yahoo as relevant documents for the query word Islamic. The first column in the second part shows the user 1 ranking for these twenty documents. The second column shows the user 2 ranking for these thirty documents, and the third column shows the user 3 ranking for these twenty documents and so on. The third part of the table shows the Yahoo search engine ranking for the same documents. Yahoo ranking for these thirty documents is the order of these documents as they retrieved by Yahoo. This means that document 1 is the first document in Yahoo ranking, document 2 is the second… document 20 is the twenty. The fourth part of the table shows our proposed method ranking for the same twenty documents. Our proposed method ranks the documents according to the number of words related to query word in the document and in its outlinks documents. Each number in the first part, the second part, and the third part of the table represents the document name. For example number 1 in the third row represents document 1, also number 14 in the same row represents document 14, and so on.

The fifth part of table shows the distance between each document position yahoo search engine ranking, and its correct position in user 1,user 2….user 5 ranking. For example document 11 in the fourth row is ranked by user 1 in the position 4 while it is ranked by Yahoo search engine in the position 11. So, the difference of the ranking position for document 11 between Yahoo ranking and user1 ranking is 11 - 4 = 7 positions.

The last part shows the distance between each document position in our proposed method ranking, and its position in user 1, user 2….user 5 ranking. For example, document 11 is ranked by user 1 in the position 4 (row 4), while it is ranked by our proposed ranking in the position 5 (row 5). So, the difference of the ranking position for document 11 between our proposed ranking and user 1 ranking is 5 - 4 = 1 positions. Also document 11 ranked by user 3 in position 5. So, the difference of the ranking position for document 11 between our proposed ranking and user3 ranking is 5– 5 = 0 positions.

The average ranking error represents the distance of the document from its position in the expert ranking on average. The average ranking error for Yahoo search engine according to user 1 is taken by calculating the average for the values in user 1 of the part 4 of the table. This means that in Yahoo ranking every document is away from its correct position in user 1 ranking 5.2 positions on average.

While the average ranking error for our proposed rankings according to user 1 are taken by calculating the average for the values in column 1 of the part 5 of the table. This means that in our proposed method for ranking every document is away from its correct position in user 1 ranking 1.6 positions on average.

The same calculations for user1 are repeated for user 2, user 3...user5. Therefore, according to user 2, user 3, user4, and user5 the average ranking errors for Yahoo search engine are 4.3, 5, 7.5, and 5 respectively.

Also, the average ranking errors for our proposed method where is compared with Yahoo are 1.8, 0.9, 1.2, and 0.8 respectively.

The average error for Yahoo search engine ranking compared with user 1 ranking, user 2 ranking, …and user 5 ranking is equal to (5.2+4.3+5+7.5+5) / 5 =5.4. The average error for our proposed method is equal (1.6+1.8+0.9+1.2+0.8) / 5 = 1.26. Therefore, by dividing the average ranking error for Yahoo search engine, on the average ranking error for our proposed method, we can conclude that our proposed ranking is better than Yahoo search engine ranking by (5.4 / 1.26) = 4.3 times.

Figure 1 shows two curves, one of the two curves represents the difference between each document position in Yahoo ranking, and the position of each document position in the five users ranking on average. And another curve represents the difference between each document

position in our proposed ranking, and the position of each document position in the five users ranking. For example, the difference between document 10 position in Yahoo ranking, and its position in the five users in average is (8 + 9 + 9+8+8) / 5 = 8.4. While the difference between document 10 position in our proposed ranking, and its correct position in the five users ranking on average is (1 + 0 + 0+1+1) / 5 = 0.6.

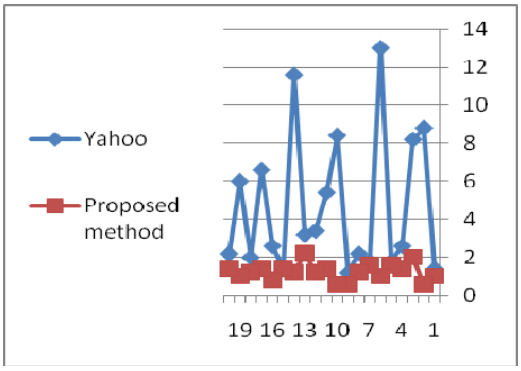

Figure 4: Yahoo search engine results and our proposed method results compared with the five interested users

As shown in figure 4 using our proposed method for ranking gives better results than Yahoo search engine according to the ranking of the five users. For example as shown in the figure document 6 in Yahoo ranking is far from its correct position on average according to its correct position in the five users ranking (13) positions. While document 6 in our proposed ranking is far from its correct position according to the five users (1) position.

4.2 Google search engine results

The previous test has been repeated, however, this time with Google search engine. According to user 1, the results show that the average ranking error Google search engine is 4.1. But the average ranking error for our proposed method compared with Google is 1.6. According to user 2, user 3, user 4, and user 5 , the average ranking error for Google are 5, 4.8, 4.9 and 5.3.
Also, the average ranking errors for our proposed method compared with Google are 2.3, 1.3, 1.3, 1 respectively.
The results show that our proposed method is better than Google search engine ranking 3.2 times.

The following figure shows two curves, one of them represents the distance between each document position in Google ranking, and the position of each document position in the five users ranking.
And the other curve represents the distance between each document position in our proposed ranking, and the position of each document position in the five users ranking

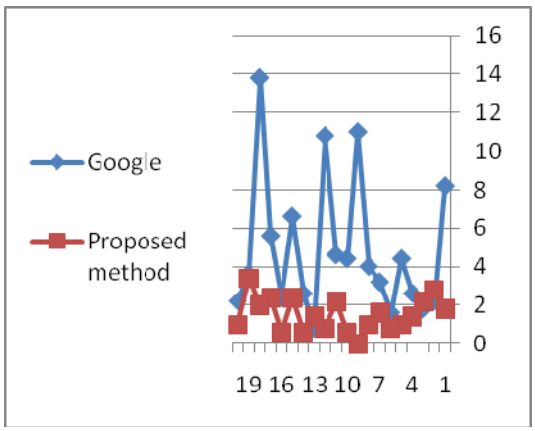

Figure 5: Google search engine results and our proposed method results compared with the five interested users

As shown in figure 5 using our proposed method for ranking give better results than Google search engine according to the ranking of the five users. For example as shown in the figure document 12 in
Google ranking is away from its correct position on average according to its correct position in the five users ranking (10.2) positions. While document 12 in our proposed ranking is away from its correct position according to the five users (0.8) positions.

Table 1: Yahoo search engine results compared with our proposed method results according to five interested users

POSITION	User Ranking 1	2	3	4	5	Ranking according to yahoo search engine	Ranking according to our proposed method	Average Error for yahoo search engine according to interest users 1	2	3	4	5	Average Error the Proposed ranking according to interest users 1	2	3	4	5
1	14	10	10	1	14	1	10	2	1	1	0	3	0	1	1	2	1
2	10	1	1	10	10	2	14	10	8	9	8	9	1	1	0	1	0
3	1	14	8	14	7	3	1	9	5	9	8	10	2	5	1	2	0
4	11	11	14	7	1	4	7	2	1	4	2	4	2	3	0	2	0
5	8	4	11	5	5	5	11	3	4	2	0	0	2	3	1	1	1
6	4	7	7	4	11	6	5	12	13	14	14	12	2	1	0	0	2
7	7	8	5	8	8	7	8	0	1	1	3	4	3	2	2	0	1
8	5	3	4	9	4	8	4	3	1	5	1	1	2	0	4	0	0
9	17	5	17	11	17	9	17	1	2	1	1	1	0	1	0	2	0
10	9	2	9	2	9	10	9	8	9	9	8	8	1	0	0	1	1
11	3	9	2	3	2	11	2	7	7	6	2	5	1	1	0	4	1
12	2	19	3	17	15	12	19	3	3	2	4	5	2	2	3	1	0
13	19	17	19	19	3	13	3	4	7	2	1	2	3	6	1	0	1
14	20	18	12	13	19	14	13	13	11	10	11	13	1	1	2	1	1
15	12	12	13	15	13	15	15	1	1	2	0	3	1	1	2	0	3
16	15	15	18	12	18	16	18	4	1	3	2	3	1	2	0	1	0
17	13	16	15	18	12	17	12	8	4	8	5	8	0	4	0	3	0
18	6	20	20	16	6	18	20	1	4	2	1	2	3	2	0	1	0
19	18	6	16	20	16	19	16	6	7	6	6	5	1	0	1	1	2
20	16	13	6	6	20	20	6	6	2	2	1	0	4	0	0	1	2
The Average Error								5.2	4.3	5	7.5	5	1.6	1.8	0.9	1.2	0.8

5. Conclusion

In this research, we focus on implementing an enhanced ranking algorithm by combining both the page content and the Hyper-Link with the focus on Arabic search engines by taking into account the stem and the context of the Arabic word. We implement ramking module based on the combination of both the count of words related to query in the page and the count of words related to query in outlinks pages of that page to calculate its rank, using external database having the morphological meanings of the most Arabic words. Then sort the pages according its rank. If there is more than one meaning to an input query word in case the user do a query in using only one word, the user may choose the meaning he/she wishes to search for. The search results will largely contain the inflected forms of the word that belong to that meaning. This helps reduce the redundancy that results from morphological search only.

The results show that the average ranking error for five different interest users for Google search engine is 4.82. But the ranking for the same documents using our proposed method has 1.5 average error ranking. Also the results show that the average ranking error for the same test using Yahoo search engine is 5.4. But the ranking using same documents using our proposed method has 1.26 average error ranking.

The results show that our proposed method for ranking better than Google search engine ranking 3.2 times, also better than Yahoo search engine ranking 4.3 times.

Finally we show that our proposed method rank is better than the ranking using any engine not considered the morphological meaning aspect. And we show that our proposed method is effective way for ranking for the retrieved documents for the user query.

References

[1] Improving User's Web Search Experience by Interactive Re-ranking and Zooming Interfaces. Jianquan Liu, Hanxiong Chen, Kazutaka Furuse, and Nobuo Ohbo. Jaban : Journal of Digital Information Management, February 2009.

[2] Extracting Related Words from Anchor Text Clusters by Focusing on the Page Designer's Intention. Jianquan Liu, Hanxiong Chen, Kazutaka Furuse, and Nobuo Ohbo,. Linz, Austria : The 20th International Conference on Database and Expert Systems Applications (DEXA2009), August 31-September 4 2009.

[3] Hao-ming Wang, Ye Guo and Bo-qin Feng. Optimizing Personalized Retrieval System Based on Web Ranking. China : Springer Berlin / Heidelberg, 2006. pp. 629-640. Vol. Volume 3967/2006. 978-3-540-34166-6.

[4] Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke and Sriram Raghavan. Searching the Web. s.l. : ACM Transactions on Internet Technology, 2001.

[5] Kleinberg, Jon. Authoritative sources in a hyperlinked environment. New York, NY, USA : ACM, 1999. Vol. Volume 46. 0004-5411.

[6] Hao-ming Wang and Ye Guo. CLBCRA-Approach for Combination of Content-Based and Link-Based Ranking in Web Search. Department of Computer Science, Xi'an University of Finance & Economics : 23–34, 2007.

[7] Zakaryia Qawaqneh, Eyas El-Qawasmeh and Ahmad Kayed. New Method for Ranking Arabic Web Sites Using Ontology Concepts. s.l. : IEEE, 2007.

[8] Junghoo Cho, Hector Garcia-Molina, Taher Haveliwala, Wang Lam, Andreas Paepcke, Sriram Raghavan, Gary Wesley. Stanford WebBase Components and Applications. s.l. : ACM Transactions on Internet Technology, 2006.

[9] Junghoo Cho, Sourashis Roy, and Robert E. Adams. Page Quality: In Search of an Unbiased Web Ranking. 2005.

[10] Monica Bianchini, Marco Gori, and Franco Scarselli. Inside PageRank. s.l. : ACM Transactions on Internet Technology, 2005.

[11] Amy N. Langville and Carl D. Meyer. Deeper Inside PageRank. 2004.

[12] Amy N. Langville and Carl D. Meyer. A Survey of Eigenvector Methods of Web Information Retrieval. s.l. : To appear in The SIAM Review, 2004.

[13] Mahgoub, Dr. Hossam. http://www.alkhawarizmy.com/En/ksearch.html. http://www.alkhawarizmy.com. [Online] alkhawarizmy, 2006.

[14] Measuring semantic similarity between words using web search engines. Bollegala, D., Matsuo, Y., & Ishizuka, M. s.l. : 16th international conference on world wide web, 2007.

[15] An editable browser for reranking web search results. Yamamoto, T., Nakamura, S., & Tanaka, K. s.l. : The rd international special workshop on databases for next-generation researchers, 2007(a).

[16] Rerank-byexample: Efficient browsing of web search results. Yamamoto, T., Nakamura, S., & Tanaka, K. s.l. : 18th international conference on database and expert systems applications, 2007(b).

Esraa Abd Elraouf received the B.S. degrees in Computer Science in 2004 .She now assistant lecture, Faculty of computer and information science, Ain shams university, Egypt.

Nagwa Badr received the B.S. degrees in Computer Science in 1996, and PhD 2000-2003, in Software engineering and distributed system, Liverpool John moorse university, U.K. Postdoctoral in Glasgow university , U.K 2003-2006. She now assistant Professor, Faculty of computer and information science, Ain shams university, Egypt.

Prof. Dr. Mohamed Tolba was the X-Vice President of Ain Shams University in 2004-2006. Dr. Tolba was the first Dean of the Faculty of Computer and Information Sciences, Ain Shams University. He has More than 130 publications in the fields of AI, Image Processing and Scientific Computing.