Arabic Information Retrieval: A Relevancy Assessment Survey

Ahmad Hussein Ababneh *ahmad.ababneh@ hud.ac.uk*
University of Huddersfield
Huddersfield, UK

Joan Lu *j.lu@hud.ac.uk*
University of Huddersfield
Huddersfield, UK

Qiang Xu *q.xu2@hud.ac.uk*
University of Huddersfield
Huddersfield, UK

Abstract

The paper presents a research in Arabic Information Retrieval (IR). It surveys the impact of statistical and morphological analysis of Arabic text in improving Arabic IR relevancy. We investigated the contributions of Stemming, Indexing, Query Expansion, Text Summarization (TS), Text Translation, and Named Entity Recognition (NER) in enhancing the relevancy of Arabic IR. Our survey emphasizing on the quantitative relevancy measurements provided in the surveyed publications. The paper shows that the researchers achieved significant enhancements especially in building accurate stemmers, with accuracy reaches 97%, and in measuring the impact of different indexing strategies. Query expansion and Text Translation showed positive relevancy effect. However, other tasks such as NER and TS still need more research to realize their impact on Arabic IR.
Keywords: Arabic Language, Information Retrieval, Natural Language Processing (NLP), Relevancy Measurement.

1. Introduction

Nowadays, the Internet is considered the first source of information for researchers in different fields of studies. This repository of information contains millions of documents related to different kinds of knowledge. The information seekers in a particular field consume a lot of time and effort to determine the documents that are highly relevant for them. Therefore, Information Retrieval concentrates on obtaining valuable documents that reflect what the user needs.

With regard to the Arabic language, the IR researchers developed and experimented several techniques to enhance the relevancy of their IR systems. They performed syntactic and semantic text analysis processes to extract significant information from the text which in turn participated in increasing the relevancy of their systems. So in this paper, we collect, analyze, and abstract their main results to offer summary which can be used as a base in building Arabic IR system. For example from our survey, we found that the Statistical approaches for stemming Arabic words yielded the highest accuracy level and the semantic single word indexing scheme is more effective than the regular indexing strategies. Therefore, the researchers can use this information and avoid wasting time searching for the best stemming and indexing methods.

The performance and relevancy of the famous search engines such as Google, Yahoo, and others are being enhanced every day and we can use them to search for Arabic text. Gross in [30] compared the relevancy and performance of these search engines between the year 2004 and 2014. Gross found a significant improve especially in Google's results but he found also

serious problems that cannot be solved in those search engines because of the special features of the Arabic language such as the prefixed and suffixed queries.
To achieve our aim we investigated and analyzed the quantitative IR relevancy measurements used by Arabic IR researchers. We summarize in tables and graphs the values of Precision (P), Recall (R), Average Precision (AP: the average of all precision obtained values when each relevant document is retrieved), and Mean_AP (which is the average of APs), if the system's relevancy calculations performed on one query then AP will equal MAP).

2. Linguistic Issues and Challenges

The Arabic language is Semitic language consists of 28 alphabets written from right to left. It is usually categorized as Classical, Modern, and Colloquial Arabic. **Classical Arabic** is the language of Al Quran and is used in Arabic literature. The **Modern Arabic language** (MAL) is similar to Classical Arabic but it uses simple vocabularies which makes it easy to understand. MAL is used in educational institutes, official documents, and media. Finally, **Colloquial or Dialectal** Arabic is a form of Arabic used in every day talking. The processing of Arabic language is a real challenge due to the existence of special features that complicated the process of Arabic script.

The Arabic Language is an inflectional and derivational language with a large number of words constructed from the same root. For example, 48 variations that can be obtained from the Arabic root "درس" means “learn”. In addition, the Arabic language has special kind of affixes called infixes in the middle of some words. For example, "دارس " has" ا " as infix and "دروس" has ""و as infix.
Arabic language **lacks the presence of short vowels** and instead it uses special characters called diacritics which are usually omitted by the writers of MAL because they assume their meanings can be captured from the context. Also, it **lacks the use of Capitalization** which increases the complexity of identifying proper nouns, adjectives derived from proper nouns, nouns, acronyms, and titles. Besides that, Arabic **allows the composition of sentences without subject.** Accordingly, the sentence ("eat food" اكل الطعام) is complete sentence even though it does not contain an explicit subject.
The **Diglossia** which means two or more variations of the same language is used but in different situations appears clearly in Arabic societies. Arab people daily use three forms of Arabic language, they use classical Arabic to recite their prayer five times a day or to read Quran, they use Modern Standard Arabic in their classrooms, Newspapers, Media, and they speak to each other in their local dialect.
Providing a reliable Arabic corpus is critical in measuring the results of new NLP or IR systems and to compare the obtained results with previous experiments. Goweder in [29] presented important work to establish a corpus from newspaper domain. Darwaish and Magdy in [22] listed some the available document collections together with their relevancy judgment used by IR researchers. In 2001 Arabic language starting new age when it included in the TREC (Text Retrieval Conferences) cross-lingual track sponsored by NIST (National Institute of Standards and Technology) which offers uniformed datasets and relevancy judgments.

3. Arabic IR Enhancement Survey

Numerous techniques have been developed to improve the relevancy of Arabic IR systems. In one hand, some researchers attempted to enhance the IR system by improving the IR systems' components such as: 1) Building high accurate stemmer suits the special features of Arabic language [15], [28], [40], [43]. 2) Improving the index structure in a way that reduces the time required to obtain the relevant documents [19], [39]. 3) Enhancing the user query by inserting new semantically related terms to the query terms. [1], [2],[16],[34], [47],[55],[58]. On the other hand, some of the surveyed researchers attempt to benefit from Natural Language processing tasks. For instance, 1) Text Summarization can reduce the document size which results in

reducing the index size and accelerating the retrieval process. 2) Named Entity Recognition (NER) can play a significant role in specifying the part of speech for each word and determining the entities appeared in the documents being searched. 3) Machine Translation (MT) and Machine Readable Dictionary (MRD) may use to translate the query and initiate the search over anther languages.
Next subsections will be structured as follows, Stemming impact: which surveys the important stemmers appeared for Arabic and their influence on Arabic IR, Indexing impact: in which we discusses the influence of different indexing strategies used for Arabic, Text Summarization Impact: in this subsection we cannot find real and integrated research completely employed TS on Arabic IR but we address it to carriage other researchers to give more attention to TS and how IR systems benefit from text summarization, Translation Impact :this subsection summarizes the impact of MT and MRD on Arabic CLIR, NER and POS Impact: in which we investigated how NER and POS were utilized on Arabic IR.

Stemming Impact

It is the process of stripping derivational and inflectional affixes of a word. This ensures that all variants of the word will be treated in a similar way by NLP or IR systems.
For the Arabic language, we can abstract four strategies used in literature to stem Arabic words: 1) *Rule-based Affix removal*: affixes stored in database dictionary. The dictionary holds Arabic language affixes such as "ال","ون","ها". Simple control statements remove those affixes out from the word. This strategy used in [10], [14], [40], [43]. 2) *Statistical based Affix removal*: In which statistical calculations determine the most important part of the processed word. This strategy used in [15], [32], [37]. 3) *Pattern matching strategy* which is mainly a morphological pattern matching process to extract three, four, five, or even six Arabic roots [10], [11], [18], [28], and [40]. 4) *Dictionary based stemming* in which the stems of words are stored in a lexicon. This strategy used in [7]. Mainly, Arabic IR researchers mixed more than one strategy to enhance the accuracy level. (Accuracy level equals correct stems divided by number of inputted words).

Table 23. Stemming Accuracy Obtained from the Surveyed Publications

Ref #	Year	Data collection	Accuracy Level	Main Finding
[40]	2001	50,000 words	70.89%	Rule-based strategy with pattern matching
[11]	2015	6081 words	75.03%	Rule-based strategy with pattern matching
[10]	2013	6000 words	75.89%	Improvement of Khoja stemmer
[15]	2005	2000 words	80%	Statistical: Successor variety stemming
[50]	2014	1000 documents	83.9%	Pattern matching strategy
[37]	2010	---	89.6%	Statistical based
[13]	2014	6225 words	92%	Statistical based
[14]	2007	1000 word	95%	Rule-based Affix removal strategy
[27]	2005	----	95%	Pattern matching strategy
[28]	2009	15180 words	95%	Rule-based strategy with pattern matching
[18]	2014	59548 words	96%	Pattern matching strategy
[24]	2002	9606 words	96%	Statistical based
[7]	2011	----	96.29%	Dictionary based with Rule based removal
[45]	2003	28449 words	97%	Statistical based

The important matter in this context is to measure the impact of stemming On Arabic IR. In [23] three stemmers were tested, Al-Stem[19], UMass, and Modified UMass stemmers. The authors used TREC2001&2002data and the obtained MAP values were 0.32, 0.32, and 0.33 respectively. Light 10 stemmer in [43], showed significant improvement and achieved 0.413 AP. Bessou and Touahria in [18] obtained 57% AP and 69% AR when they tested the effective of their stemming on Arabic IR. They achieved 15% improvement in AP and 28% in AR comparing with no stemming IR system.

[19] Available online

Aljlayl and Frieder in [8] used AP to compare the relevancy of three Arabic IR system, the first one called surface-based retrieval in which no stemming where performed, the second one is root-based retrieval in which they used aggressive algorithm to extract the words' roots, and the final system used light stemmer to remove the prefixes and suffixes of the words. They showed that the AP improved by 43.2% using root based stemming and by 71.3% using light stemming (surface-base stemming AP=0.2517, root-base stemming AP = 0.3604, and light stemming based AP=0.4312).
In [20] Chen and Gey studied the impact of light stemming and Machine translation based stemming on Arabic IR. The authors used precision and recall relevancy measures to show that the light stemmer outperformed the MT-based stemmer when those stemmers were used as stemming tool (IR system with MT-based stemmer the R =0.83 and P=0.33, IR system with light stemmer the R=0.84 and P=0.37).

Indexing Impact

The index is a data structure that represents the documents' contents as a set of weighted terms. The index is inspected against the user query to find which terms inside the index best match the user query. Therefore, a well-established index is necessary to optimize the speed and performance [48].
The Index can be a *regular index* with a single word (SW) or multi-words (MW) entries or *semantic* index relates the semantically related terms. The single word entry in the index might be a complete word, stem (the word without prefixes and suffixes), or root (represents the base or dictionary form of the word without prefixes, suffixes, and infixes). Phrase indexing was also investigated and the basic idea is phrases semantically characterize document contents more effectively than single word terms. Semantic indexing relays on word senses disambiguation and tries to find the correct sense of the word among different senses. Researchers have examined different strategies to identify suitable indexing strategies to improve the Precision measure.

Table 24. Relevancy Calculations (references addresses Indexing Impact)

Ref	Index type, Main Finding, and IR Measurement
[3]	AP= 39%(key terms index), AP = 54%(semantic, without disambiguation words resolve) ,AP=60% (semantic, with disambiguation words resolve).
[4]	Building semantic index inst. MAP = 39%(key-terms index) ,MAP = 60%(semantic index)
[6]	SW terms index. MAP=54% (Roots), MAP=42% (Stems) , MAP=29% (Full word)
[8]	SW terms, Light stemming AP = 43% (Stems), AP = 36% (Roots),AP = 25% (Full word)
[12]	SW terms index, MAP=66% (Roots), MAP=58 % (Stems), MAP=39% (Full word)
[18]	SW terms index, pattern matching stemming algorithm, AP=57%(stem),AP=43(Full word)
[19]	AP using MW is greater than the AP using SWT by 5.8%. (AP value = 31.9)
[25]	Use KP-Miner system to extract MW index entries. (AP=0.19,AR=0.43)
[26]	Use statistical measures with the linguistic knowledge to extract MW index entries. P=65%,R=40%
[35]	SW stem terms index, VSM as IR matching Model. AP = 66%(Stem), AR = 80%(Stem)
[36]	SW terms index. AP = 28.4 (Root), AP=26.8 (Stem) , AP=19.8 (Full word)
[39]	Nouns as index terms. The AP/AR declined by less than 1% and the index size shrinks by 45%.
[46]	SW terms, experimented n-gram as a source of indexing. R = 41.3% (Stem), P= 32.81%.(Stem)
[48]	SW term index, VSM as IR matching Model. AP=64%(Root), AR=46%(Stem)

Since 1975, Salton began his investigation about the importance of indexing in the field of IR. At the beginning, the indexing process was performed in manual bases which are mainly expensive, time-consuming, inaccurate, and harder to maintain and update [54]. Salton emphasized on term frequency as a base to select index terms. He neglected the most and least frequent terms and he indexed midrange terms. For the Arabic Language, the authors of [36] experimented the use of term frequency to select index terms for Arabic language and they indexed the terms that have frequencies greater than one and less than 261. They built an Information Retrieval system with manual and automatic index and conducted a series of experiments using full word, stem, and root as index terms. The results showed that the automatic indexing is comparable to manual indexing. In addition, it is cheaper and faster. The main finding, the IR relevancy measurements calculations, and the type of indexing of some of the surveyed publications were summarized in table 2

Query Expansion (QE) Impact

QE is an IR technique used to improve the relevancy of IR systems through the amendment of the query key terms. Normally, the user query is short contains few words and sometimes the ignorance of the subject being searched makes the user uses inadequate key terms. This technique assumes that the user query is the problem in retrieving irrelevant documents and the relevancy can be improved by either adding new words to the query or by replacing the query words with new words.

Two main methods have been employed to expand users' queries; the first one uses linguistic knowledge to add synonyms to the terms used in the query and drive new semantically related terms to the query terms. The second method uses user feedback to enhance the query. Another categorization of query expansion methods is Global or Local. Global methods perform the expansion by hiring external resources such as lexicon thesaurus or stemming algorithm and Local methods use the data collected from the first run of the IR system to expand the query. The surveyed publications that measure the impact of QE appears in table 3.

Table 25. Relevancy Calculations (references addresses Query Expansion Impact)

Ref	Expansion type and Relevancy measurements Values
[1]	Without QE: R=70%, P=21%, and MAP=37%. Automatic QE (Global, thesaurus): R=91% ,P=6%, MAP= 25%. Interactive QE (local, Relevance feedback): R=82%, P=15%, MAP=41%.
[2]	Without expansion: R = 34%, P = 6%. With Local expansion : R=74%, P=12%
[16]	Baseline retrieval: R = 47%, MAP = 16%, F_Measure = 24. QE using Semantic WordNet and Semantic Similarity: R= 50%, MAP= 29%, F_Measure= 37%
[34]	Local, thesaurus based, With expansion: MAP = 56.6, Without expansion: MAP = 55.5
[47]	Before expansion: R = 31%, P = 33% After expansion (Local, thesaurus): R = 74%, P = 81%.
[55]	Before expansion: R=84%, After expansion(Local, statistical base) : R=91%
[58]	Without expansion: MAP=34%, With expansion(Global, thesaurus based): MAP=48 %

Automatic Text Summarization (ATS) Impact

ATS is the computer ability of to simulate the human beings skills in drawing the main ideas or the key sentences from a particular text. The summaries represent important information found in any article. If the summary of the document contains sufficient information it can be employed in place of the full document itself in the IR systems. We should satisfy the equation that keeps the relevancy of the IR system reasonable and at the same time reduces the retrieval time

No actual work was found for the employment of ATS in Arabic IR and we reviewed some experiments applied for the English language to encourage Arabic Language researchers to address this field. Ronald et al in [52] obtained high precision measuring rate when they used domain-independent automatic summary (extractive summary) based on tf-idf sentence selection as index source. They compared the results obtained from their extractive summary with anther simple summary whose sentences are selected from the first few sentences in the original document (called Lead summary) and also with full-text indexing. In [53] Sakai and Sparck-Jones employed the same idea of using the summary for indexing but they studied the impact of Generic summaries with or without Pseudo-relevance feedback. Sakai's results in [53] were in line with the results obtained by [52] in which a summary-only is very effective and roughly give the same results as using full-text for precision-oriented searching and the relevance feedback used in this research will also add more value and combines the new indexing approach with query expansion methods. However, the using of Lead Summary as experimented by Wasson in [57] decreases the acceptability rate for special News documents such as Lists, news briefs, and transcripts. Another employment of ATS in IR -especially in

Geographical Information Retrieval subfield GIR - done by Perea-Ortega et al in [49].They utilized two kind of summaries, General summary based on word frequency and noun phrases , and Geographic summary which gave more attention to the sentences that contain Geographical entities. If we compare these values with the MAP value obtained in the same experiment using full text (no summaries) we can accept the Geographical summary as a source of the index. Accurately, the IR system used in [49] based on Geographical summary achieved 95% of the results obtained based on full text and at the same time reduces the index size by 40%.
In [49], [52], [53], [57] the authors used Compression Rates (CR) of fixed size. Firstly, the CR equals the summary length over the document length and these parameters could be expressed in terms of a number of words or sentences or as a percentage. Ranald in [52] generated summaries containing 60, 150, and 250 words whereas; in [49] the compression rates were 20%, 40%, 60%, and 80% of the original text. Indeed, the use of fixed sized CR may have a negative impact and it seems unfeasible especially for IR tasks. In more details, the richness of information defers from one document to another and a certain document may need 30% CR to capture its entire salient information and anther one needs 50%. The IR relevancy measurements of the publications surveyed in this section summarized in table 4.

Table 26. Relevancy Calculations (references addresses ATS Impact)

Ref	IR Relevancy Measure and value
[49]	MAP= 0.31 using General-summary index, MAP= 0.36 using Geographical summary
[52]	AP=0.37 full index, AP=0.45 extraction-summary index,AP=0.47 lead-summary index
[53]	AP = 0.249 full text, AP = 0.231 Abstract summary, AP=0.233 Extract summary
[57]	P = 0.86 ,R = 0.23 ,F-Measure = 0.33

Automatic Translation Impact

At first, Machine Translation (MT) is the ability of the computer to translate written or spoken text from one language to another on sentences level. The IR strategy that uses translation facilities is the Cross-Language Information Retrieval (CLIR). CLIR is a retrieval strategy returns relevant documents written in several languages. In CLIR the system normally translates the query and tries to find relevant documents in original and target languages.
In this section we concentrate on the following points: 1) the use of MT or Machine Readable Dictionary (MRD) as a translation tool. 2) The number of candidate translations generated for each query. And 3) the employment of proper noun transliteration in CLIR.
The queries translate to another language using either MT system or Machine Readable Dictionary (MRD). Aljilay in [9] held a comparison between two Arabic CLIR systems, the first one used ALKAFI MT systems and the second one uses Arabic to English dictionary called Al-Mawrid. His results came in line with Radwans Results in [51] and they both stated that the retrieval system that uses bilingual dictionary with the reasonable number of translations for each query term outperforms a retrieval system that uses MT system to translate the query. Darwish in [23] experimented the use of two Arabic MT systems Tarjim and AL-Misbar and Salmone Arabic to English dictionary in IR system. In [44] two bilingual dictionaries were used (UMass dictionary & TREC standard probabilistic dictionary).
Abu Salam in [6] used Sakher MT system to generate a translation matrix for each query. One, two, or three candidate translation(s) for each query term were generated. His experiment showed that the retrieval performance of CLIR using MT system outperformed the monolingual retrieval especially when he used the word as index term not stem or root. Abo Salam results are consistent with Hull and Grefenstette [38] who showed that word –by- word translation degraded the retrieval performance by 40 to 60%.
A simple CLIR system can use Bilingual dictionary to translate the query terms without use complete MT system. The problem with this kind of dictionary is the out of vocabulary (OOV) words which are mainly proper nouns for persons, places, and others that are missing in such dictionary. In [17] Bellaachia and Amor-Tijani used English to Arabic dictionary to do the necessary translation and every OOV word was transliterated using statistical techniques

followed by n-gram string matching. Prior to Bellaachia and Amor-Tijani in [17], Larkey et al in [42] conducted two levels of experiments in this context. Firstly, they found that the MAP jumped from 0.14 in the case of no name transliteration to 0.33 in the case of name transliteration. Secondly, they found that as the number of transliterations for the query proper nouns increased the MAP increased. Abdul Jaleel and Larkey in [5] experimented a simple technique for statistical n-gram transliteration called selected n-gram. They tested the effectiveness of this technique on Arabic IR. The authors tested the mapping of each name and word to 20 transliterations. The IR relevancy measurements of the publications surveyed in this section were summarized in table 5

Table 27. Relevancy Calculations (references addresses Automatic Translation Impact)

Ref	IR Relevancy Measure and values
[5]	AP=0.15 baseline, AP=0.20 with names transliteration, AP=0.21 with words transliterations
[6]	AP= 0.45(Two candidate translations), AP= 0.42(Three candidate translations)
[9]	Ap= 0.17 (MT as translation tool), Ap= 0.20 (bilingual dictionary translation)
[17]	MAP= 0.46 (without transliteration), MAP= 0.72 (with transliteration)
[23]	MAP= 0.33 (two Arabic MT system -Tarjim , AL-Misbar- and one Arabic to English dictionary -Salmone- employed in IR system
[42]	MAP=0.14 without transliteration, MAP=0.19 with one transliteration MAP=0.25 with five transliterations, MAP=0.30 with 20 transliterations
[44]	MAP= 0.3996 UMass dictionary & TREC standard probabilistic dictionary

NER & POS Impact

Named Entity Recognition (NER): It is an information extraction task aims to identify the entities mentioned in a given text document. The output of NER systems is mainly a list of proper nouns of the entities mentioned in the text.

Part of Speech Tagging (POS) is considered as an essential tool for any robust NLP application. It's a morphological analysis process in which we determine the lexical form for each word in the text (Noun, Adjective, Verb ...).

Firstly we want to explain that this section combined Named Entity Recognition (which is NLP task) with POS tagging (which is a morphological analysis process) because both of them have been employed in IR systems in similar manner. Mainly, POS taggers and NER systems are used in IR to measure the effect of nouns in the IR systems whether they are common nouns or proper nouns. Unfortunately, NER task has not taken that much of attention from Arabic IR researchers and a little work can be found in this area.

For English, in [31] the authors addressed the existence of Named entities in the user's queries and through their analysis, 71% of queries contained Named Entities. In 1997, Thomson in [56] investigated the effect of NER systems in IR systems. They claimed that effective NER system will support the IR processes and improve their outcomes. The Information retrieval system developed by him benefited from the NER system and obtained 83.9 average precision comparing with baseline IR system which yielded 74.8. Another experiment in English language and other languages (Dutch and French) reported in [41], in which four off-the-shelf name entity recognition tools were used to extract proper names.

The output of NER system is a set of proper names. Abdur Chowdhury in [21] experienced the use of nouns index instead of full-text index in Information retrieval. In [39] the authors adopted the hypothesis that nouns are the most important part in any text document and could be used as IR discriminator. They analyzed their corpus and found that 55% of the words are nouns. Their experiments showed that the use of nouns as index term obviate the need for using anther part of speech types.

Table 28. Relevancy Calculations (references addresses NER Impact)

Ref	IR Relevancy Measure and values	Language
[21]	MAP=0.13 (full text index) , MAP = 0.121 (proper noun index)	Arabic
[39]	MAP=0.285 (full text index) , MAP = 0.277 (proper noun index)	Arabic

[41]	P= 0.97 (with NER system), R= 0.68 (with NER system),	English, Dutch, French
[56]	AP= 0.839 (with NER system), AP = 0.748 (without NER system)	

4. Results and Analysis

In this survey, we concentrate on investigating the relevancy calculations computed in 58 publications that handled the problem of Arabic IR.

Stemming dominated large space of interest in Arabic IR research. Mainly, three approaches were used in finding the stem of Arabic words, statistical approach, rule-based approach, and pattern matching approach. Table 1 shows that the three approach succeed to obtain 95% level of accuracy. But to be more accurate we computed the Average accuracy achieved by each approach using our references as the sample and the results are presented in figure 1. The positive Effect of stemming appears clearly in all surveyed publications that studied the effect of stemming on Arabic IR. Both stem and root based retrieval returned more relevant documents comparing with full-word based retrieval. For example, in [5], [8], [12], [36] the AP was boosted with percentages of improvements 25%, 18%, 27%, 8% respectively.

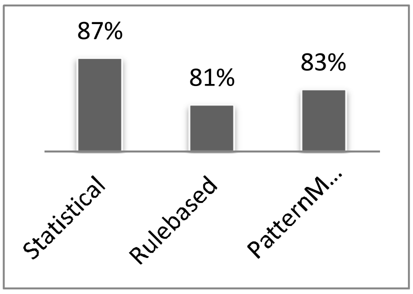

Fig. 1. The Average Accuracy for different stemming approaches

The **index** structure and the data to be indexed investigated in Arabic IR literature and we reviewed the most recent ones. SW-full index, WS-stem index, SW-root index, SW-ngarm index, MW index, and Semantic index are employed either manual or automatic. Table 2 summarizes the main finding in 14 research publications which investigated the impact of different indexing strategies on Arabic IR. Single-term indices dominated a large number of those researches with AP swung between 20% and 90%. On the other hand, Semantic index surpassed the regular key terms index. In general, the main findings related to indexing strategies can be summarized in the following points:

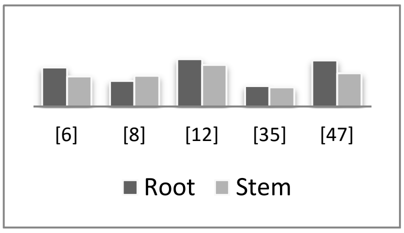

Fig. 2. Roots or stems as SW index entries

1. In [6], [12], [36], [48] the use of root as index term gave the best relevancy results comparing with stem or full word term index. See figure 2
2. In our survey, we note that the choice between MW indexing and SW indexing tends toward the latter. In SW index: the AP values swung between 0.198 in [36] and 0.90 in [33]. In MW index: the maximum value of AP appeared in our survey for Arabic IR systems with phrases indexing was 0.37 found in [19].
3. The choice between key terms index and semantic index tends toward the latter. The improvement of MAP values using semantic index over key term index reached 15% in [3] and 21% in [4].

Named Entity Recognition: we investigated the NER impact on Arabic IR system and we found modest effort spent in this NLP task. In general, the output of the surveyed NER systems showed a positive impact. The noun generated by NER system used as index terms and did not degrade the IR system relevancy and improve the efficiency. For example, in [38] and [21], the authors showed that the use of nouns as index terms Obviated the use of other types of part of speech. See Table 7.

Machine Language translation: In our survey we discussed the impact of translation on Arabic IR, From [6], [9], [17], [23], [44], [51] we can conclude that the use of word by word bilingual dictionary to translate the user query will degrade the retrieval system. Therefore, the researchers improved their results by expanding the query by taking several translations for each query term. For example in [6] the authors took 2 or 3 term translations and in [17] they expanded the query by taking different proper nouns transliterations. On the other hand, the use of Machine translation systems does not give the expected enhancement on CLIR system.

For example, in [9] the AP with MT (0.17) was less than the AP of the monolingual retrieval (0.27). Normally, using MT to translate the user query will yield single translation because the MT generates one translation for each query term. As stated by the authors of [59], the output of MT is a literal mapping and this ignores the availability of multiple expressions in the target language.

Table 7. Relevancy Calculations (references addresses Noun as index entry Impact)

REF	Full-text	Nouns only	Difference	Language
[21]	MAP=0.13	MAP=0.121	0.009	English
[39]	MAP=0.285	MAP=0.277	0.008	Arabic

Query Expansion: the researchers in Arabic information retrieval investigated the problem of expansion from different perspectives: 1) Linguistic analysis with relevance feedback expansion. 2) Automatic and Interactive feedback expansion. 3) Local and global thesaurus based expansion. 4) Local linguistic or statistical analysis expansion. And 5) Local semantic analysis based on statistical calculation or global semantic analysis based on Arabic WordNet. Our survey showed a positive impact of QE and this appears clearly in all publications examined. The recall was improved and the percentage of improvement in MAP lies between 1% in [34] and 14% in [58] which gives strong indication that the query expansion recalled more relevant documents and discarded some of the irrelevant ones. Another important note related to precision measure which showed enhanced [2], [16], [47], [58] or stable trend [1], [34]. See figure 3

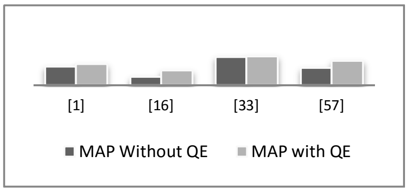

Fig. 3. the Impact of QE from the surveyed publications

5. Conclusion

Our research pinpoints the important efforts appeared in last few years to enhance the Arabic IR systems. **Stemming** deeply investigated in our survey showing that Statistical methods suited the Arabic language. Furthermore, the **indexing** process played an important role in any IR system. For the Arabic language, single-term index, key-phrase index, and semantic index were reviewed. The survey depicts that the use of Arabic roots or stems as index term gives the most significant relevancy results. Moreover, we tend to support semantic indexing over regular key terms indexing. **Text Translation** has a significant impact on Arabic IR with bilingual dictionary translation surpassed the Machine Translation systems. **ATS** also shows a positive improvement in IR system in term of retrieval speed and efficiency. Finally, local, Global, and Semantic **QE** with relevance feedback strongly support Arabic IR systems due to the richness of Arabic Language vocabularies.

Reference

1. Abbache, F., Barigou, F. Z., Belkredim, G. Belalem: The Use of Arabic WordNet in Arabic Information Retrieval. International Journal of Information Retrieval Research. 4 (3) , 54-65 (2014)
2. Abdelali, A., Cowie, J., and Soliman, H. S.: Improving Query Precision Using Semantic Expansion. Information Processing and Management. 43(3), 705-716 (2006)
3. Abderrahim, M., Dib, M., Abderrahim, M. E., Chikh, M. A.: Semantic Indexing of Arabic Texts For Information Retrieval System. International Journal of Speech Technology. Springer US. pp 1-8 (2015)
4. Abderrahim, M., Abderrahim, M. E., Chikh, M. A.: Using Arabic WordNet for Semantic Indexation in Information Retrieval System. IJCSI International Journal of Computer Science Issues. 10 (10) (2013)
5. Abduljaleel, N., Larkey, L. S.: Statistical Transliteration for English-Arabic Cross Language Information Retrieval. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management, New Orleans (2003)

6. Abu-Salem, H., Chan, P. K.: English-Arabic Cross-Language Information Retrieval Based on Parallel Documents. International Journal of Computer Processing of Oriental Languages. 19 (1), 21–37 (2006)
7. Alhanini, Y., Aziz, M. A.: the Enhancement of Arabic Stemming by Using Light Stemming and Dictionary-Based Stemming. Journal of Software Engineering and Applications. 4 (9), 522-526 (2011)
8. Aljlayl, M., Frieder, O.: On Arabic Search: Improving the Retrieval Effectiveness via a Light Stemming Approach. In: Proceedings of the ACM Eleventh International Conference on Information and Knowledge Management, pp. 340-347, Virginia (2002)
9. Aljlayl, M., Frieder, O.: Effective Arabic-English Cross-Language Information Retrieval via Machine-Readable Dictionaries and Machine Translation. In: Proceedings of the Tenth International Conference on Information and Knowledge Management. pp. 340-347, pp. 295-302, Atlanta (2001)
10. Al-Kabi, M.: Towards Improving Khoja Rule-Based Arabic Stemmer. In: Proceedings of the 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). pp 1-6, Amman (2013)
11. Al-Kabi, M., Kazakzeh, S. A., Abu Ata, B. M., Al-Rababah, S. A., Alsmadi, I. M.: A Novel Root Based Arabic Stemmer. Journal of King Saud University - Computer and Information Sciences. 27 (2), 94-103 (2015)
12. Al-Kharashi, I. A., Evens, M. E.: Comparing Words, Stems, and Roots as Index Terms in an Arabic Information Retrieval System. Journal of the American Society for Information Science (Wiley Periodicals Inc). 45 (8), Start Page:548 (1994)
13. AL-Omari, A., AbuAta, B.: Arabic Light Stemmer (ARS). Journal of Engineering Science and Technology. 9 (6), 702-716 (2014)
14. Al-Shalabi, R., Kanaan, G.,Ghwanmeh, S.: Stemmer Algorithm for Arabic Words Based on Excessive Letter Locations. In:Proceedings of the 4th International Conference on Innovations in Information Technology. UAE: IEEE, pp. 456-460 (2007)
15. ALShalabi, R., Kanaan, G., Hilat, I., Ababneh, A., AlZubi, A.: Experiments with the Successor Variety Algorithm Using the Cutoff and Entropy Methods. Information Technology Journal. 4 (1), pp. 55-62 (2005)
16. Atwan, J., Mohd, M., Rashaideh, H., Kanaan, G.: Semantically Enhanced Pseudo Relevance Feedback for Arabic Information Retrieval. Journal of Information Science. ISSN 0165-5515, (2015)
17. Bellaachia A., Amor-Tijani, G.: Proper Nouns in English–Arabic Cross Language Information Retrieval. Journal of the American Society for Information Science and Technology. 59 (12), (2008).
18. Bessou, S., Touahria, M.: an Accuracy-Enhanced Stemming Algorithm for Arabic Information Retrieval. Neural Network World. 24 (2), 117-128 (2014)
19. Boulaknadel, S., daille, B., Driss, A.:Multi-word term Indexing For Arabic Document Retrieval. Computers and Communications. ISCC 2008. IEEE Symposium on. pp. 869 - 873, IEEE, Marrakech (2008)
20. Chen A., Gey, F.: Building an Arabic Stemmer for Information Retrieval. University of California at Berkeley. CA 94720-4600, USA (2002)
21. Chowdhury, A., Mccabe, M. C.: Improving Information Retrieval Systems Using Part of Speech Tagging. Digital Repository at the University of Maryland (DRUM). http://hdl.handle.net/1903/5958, (1998)
22. Darwish, K., Magdy, W.: Arabic Information Retrieval. Now Publishers Inc.7 (4), 239-342 (2014)
23. Darwish, K., Oard, D. W.: CLIR Experiments at Maryland for TREC-2002: Evidence Combination for Arabic-English Retrieval. In TREC. 2002. Gaithersburg (2002)

24. Darwish. K.: Building a Shallow Morphological Analyzer in One Day. In: Proceedings of the ACL-2002 Workshop on Computational Approaches to Semitic Languages .(2002)
25. ElBeltagy, S., Rafea, A.: KP-Miner: A Keyphrase Extraction System for English and Arabic Documents. Information Systems (Elsevier Ltd). 34 (1), 132-144 (2009)
26. Elshishtawy, T., Al-sammak, A.:Arabic Keyphrase Extraction using Linguistic knowledge and Machine Learning Techniques. In: Proceedings of the Second International Conference on Arabic Language Resources and Tools. CArabic IRo, (2009)
27. Ghawanmeh, S., Al-Shalabi, R. , Kanaan, G., Khanfar, K., Rabab'ah, S.: An Algorithm for Extracting the Root for the Arabic Language, Proceedings of the 5th International Business Information Management Association Conference (IBIMA), CArabic IRo, Egypt (2005)
28. Ghwanmeh, S., Rabab'ah, S., Al-Shalabi, R.,Kanaan., G.: Enhanced Algorithm for Extracting the Root of Arabic Words. Computer Graphics, Imaging and Visualization, CGIV '09. Sixth International Conference on. pp. 388 - 391, Tianjin:IEEE (2009)
29. Goweder, A., De Roeck, A.: Assessment of a Significant Arabic Corpus, Presented at the Arabic NLP Workshop at ACL/EACL 2001. Toulouse (2001).
30. Gross, A., M.: Information Retrieval in Arabic: an Evaluation of Three Multilingual Search Engines on their Capabilities in Dealing with Arabic Search Queries. International Journal of Information Technology and Business Management. 40 (1). 1 – 21 (2015).
31. Guo, J., Xu, G., Cheng, X., Li, H.: Named Entity Recognition in Query. In: Proceedings of the 32nd international ACM SIGIR conference on Research and development in Information Retrieval, Boston, MA, USA, (2009)
32. Hafer, M. A., Stephen, F. W.: Word Segmentation by Letter Successor Varieties, Information Storage and Retrieval. 10(11–12), pp. 371–385 (1974)
33. Hammouda, F., Almarimi, A.: Heuristic Lemmatization for Arabic Texts Indexation and Classification. Journal of computer science. 6 (6), pp. 660-665 (2010)
34. Hanandeh, E.: Building an Automatic Thesaurus to Enhance Information Retrieval, IJCSI International Journal of Computer Science Issues. 10 (3), (2013)
35. Harrag, F., Aboubekeur, H-C. , El-Qawasmeh, E.: Vector Space Model or Arabic Information Retrieval — Application to "Hadith" Indexing, Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference IEEE. pp. 107-112, Ostrava (2008)
36. Hmeidi, I., Kanaan, G., Martha, E.: Design and Implementation of Automatic Indexing for Information Retrieval with Arabic Documents. Journal of the American Society for Information Science. 48 (10), 867–881 (1997)
37. Hmeidi, I., Alshalabi, R., Al-Taani, A., Najadat, H., Al-Hazimah, S. : A Novel Approach to the Extraction of Roots from Arabic Words Using Bigrams. Journal of the American Society for Information Science. 61 (3), 583-591 (2010)
38. Hull, D., Grefenstette, G. : Querying Across Languages: A Dictionary-Based Approach to Multilingual Information Retrieval. In: Proceedings of the 19th annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp.49–57, New York (1996)
39. Kanaan, G., Al-Shalabi, R.,Sawalha, M.: Improving Arabic Information Retrieval Systems Using Part of Speech Tagging, Information Technology Journal (Asian Network for Scientific Information). 4 (1), 32-37(2005)
40. Khoja, S.: Khoja Stemmer (2001), http://zeus.cs.pacificu.edu/ shereen/research.htm, Accessed January, 2016.
41. Kumar, N., De Beer, J., Vanthienen, J., Moens, M.F: Evaluation of Information Retrieval and Text Mining Tools on Automatic Named Entity Extraction. Intelligence and Security Informatics, Springer Berlin Heidelberg. 666-667 (2006)

42. Larkey, L. S., Abduljaleel , N. ,Connell, M.: What's in a Name?: Proper Names in Arabic Cross Language Information Retrieval, Technical report, CIIR Technical Report, IR-278 (2003)
43. Larkey, L. S., Ballesteros, L., Margaret, E., Light Stemming for Arabic Information Retrieval. In: Arabic Computational Morphology Text, Speech and Language Technology. pp. 221-243, Springer, Netherlands, (2007)
44. Larkey, L. S., Allan, J., Connell, M. E., Bolivar, A. , Wade, C.: UMass at TREC 2002: Cross Language and Novelty Tracks. In: The Eleventh Text Retrieval Conference (TREC 2002). Gaithersburg: NIST, (2003)
45. Lee, Y., Paining, K. , Roukos, S. , Emam, O. , Hassan, H.:Language model based Arabic word segmentation. In: the Proceedings of the 41st annual Meeting of the Association for Computational Linguistics. Sapporo, Japan. pp. 399-406. (2003)
46. Mahmoud, R. , Sanan, M., Improving Arabic information retrieval system using N-gram method. WSEAS Transactions on Computers. 10 (4), 125-133 (2011)
47. Mallat, S. , Zouaghi, A. , Hkiri, E. , Zrigui, M.: Method of Lexical Enrichment in Information Retrieval System in Arabic. International Journal of Information Retrieval Research. 3 (4), 35-51 (2013)
48. Mansour, N., Haraty, R.A., Daher, W., Houri, M.: An Auto-Indexing Method for Arabic Text. Information Processing and Management. 44 (4), 1538 - 1545 (2008)
49. Perea-Ortega, M. J.: Application of Text Summarization Techniques to the Geographical Information Retrieval Task. Expert Systems with Applications (Elsevier B.V). 40 (8), 2966 – 2974 (2013)
50. Qussai, Y., Hmeidi, I.: Extracting the Roots of Arabic Words without Removing Affixes. Journal of Information Science, Chartered Institute of Library and Information Professionals. 40 (3) 376 – 385 (2014)
51. Radwan, K., Fluhr, C.: Textual Database Lexicon Used as a Filter to Resolve Semantic Ambiguity Application on Multilingual Information Retrieval. The 4'h Symp. On Document Analysis and Information Retrieval. pp. 12l-l36, LasVegas (1995)
52. Ronald, B., Karl, M., Rau Lis, F.: Automatic Condensation of Electronic Publications by Sentence Selection, Information Processing & Management. 31(5), 675–685 (1995) .
53. Sakai, T., Sparck-Jones, K.: Generic Summaries for Indexing in Information Retrieval, ACM Special Interest Group on Information Retrieval. pp. 190 – 198 (2001).
54. Salton, G. , McGill, M. J.: Index construction, Introduction to Modern Information Retrieval. Online Ed, McGraw-Hil, New York (1983)
55. Shaalan, K. , Al-Sheikh, S. , Oroumchian, F. : Query Expansion Based-on Similarity of Terms for Improving Arabic Information Retrieval. In: Z. Shi, D. Leake & S. Vadera (eds), Intelligent Information Processing VI:Proceedings of 7th IFIP TC12 International Conference, pp. 167-176, Springer, Heidelberg (2012)
56. Thompson, P., Dozier, C .C.: Name Searching and Information Retrieval. In: Proceeding of 2nd conference of Empirical Methods in Natural Language Processing EMNLP 199. pp. 134-140, Rhode Island (1997)
57. Wasson, M. : Using Leading Text for News Summaries: Evaluation Results and Implications for Commercial Summarization Applications. In: Proceedings of the 17th international conference on Computational linguistics. Association for Computational Linguistics. pp. 1364-1368, Stroudsburg (1998)
58. Wedyan, M. , Alhadidi, B., Alrabea, A.: The Effect of Using a Thesaurus in Arabic Information Retrieval System. IJCSI International Journal of Computer Science Issues. 9(6), (2012).

59. Zhou, D., Truran, M., Brailsford , T. , Wade, V., Ashman, H.: Translation Techniques in Cross-Language Information Retrieval. ACM Computing Surveys (CSUR). 45 (1), 1-44 (2012)