Effective Arabic-English Cross-Language Information Retrieval via Machine-Readable Dictionaries and Machine Translation

Mohammed Aljlayl
Information Retrieval Laboratory
Illinois Institute of Technology
Chicago, IL 60616
aljlayl@ir.iit.edu

Ophir Frieder
Information Retrieval Laboratory
Illinois Institute of Technology
Chicago, IL 60616
ophir@ir.iit.edu

ABSTRACT
In Cross-Language Information Retrieval (CLIR), queries in one language retrieve relevant documents in other languages. Machine-Readable Dictionary (MRD) and Machine Translation (MT) are important resources for query translation in CLIR. We investigate MT and MRD to Arabic-English CLIR. The translation ambiguity associated with these resources is the key problem. We present three methods of query translation using a bilingual dictionary for Arabic-English CLIR. First, we present the Every-Match (EM) method. This method yields ambiguous translations since many extraneous terms are added to the original query. To disambiguate the query translation, we present the First-Match (FM) method that considers the first match in the dictionary as the candidate term. Finally, we present the Two-Phase (TP) method. We show that good retrieval effectiveness can be achieved without complex resources using the Two-Phase method for Arabic-English CLIR. We also empirically evaluate the effectiveness of the MT-based method using short, medium, and long queries from TREC. The effects of the query length on the quality of the MT-based CLIR are investigated.

Keywords
Information Retrieval, Cross-language, Two-Phase, MRD, MT

1. INTRODUCTION
With the rapid growth of the Internet, the World Wide Web (WWW) has become one of the most popular mediums for the dissemination of multilingual information. This ability to disseminate multilingual information has increased the need to automatically mediate across multiple languages, and in the case of the WWW, access to "foreign language" Web pages.

Our goal for Arabic Cross-Language Information Retrieval (CLIR) is to enable users to query in the Arabic language against an English collection. To achieve this goal, we investigate two techniques namely the use of Machine Readable Dictionaries (MRD) and Machine Translation (MT) based approaches.

In the MRD realm, we consider three possible techniques: the Every-Match (EM), the First-Match (FM), and the Two-Phase (TP) methods. The Every-Match method considers all the translations found in a bilingual dictionary. This leads to ambiguous translations because it introduces extraneous terms to the target query and yields relatively poor effectiveness. Another method is the First-Match method. Instead of considering all the target language equivalents in the bilingual dictionary, we use the first match in the bilingual dictionary as the candidate translation of the source query term. This approach takes advantage of the fact that dictionaries typically present the translations in the order of their common use. That is, the more common translations are listed first. The FM method ignores some of the less common translations of the source language, and thus, potentially improves the retrieval effectiveness. Finally, the Two-Phase method initially considers all the translations found in the bilingual dictionary as candidate terms then removes the translated candidate terms that do not return the original source query term. We found that the TP approach consistently outperforms the EM and FM methods. In addition to MRD, we also empirically evaluated the effectiveness of Arabic to English MT-based method.

Arabic, one of the six official languages of the United Nations (UN), is the mother tongue of 300 million people [13]. Unlike the Latin-based alphabets, the orientation of writing in Arabic is from right-to-left. The Arabic alphabet consists of 28 letters. As discussed in [24], the Arabic alphabet can be extended to ninety elements by additional shapes, marks, and vowels. Most Arabic words are morphologically derived from a list of roots. The root is the bare verb form; it can be triliteral, quadriliteral, or pentaliteral. Most of these roots are made up of three consonants.

Arabic words are classified into nouns (adjectives and adverbs), verbs, and particles. All verbs and some nouns are derived from a root. Arabic sentences are either verbal or nominal. Verbal sentences contain a verb before a nominative noun (the subject), and may contain complements. Nominal sentences begin with a subject followed by a noun, an adjective, a prepositional phrase, or an adverb. In formal writing, Arabic sentences are delimited by commas and periods as in English, for instance.

Arabic-English CLIR means the retrieval of documents based on queries formulated by a user in the Arabic language, and the documents are in the English language. In Section 2, we review the prior work in Arabic information retrieval and CLIR. The proposed dictionary-based methods for Arabic-English CLIR are presented in section 3. In Section 4, we discuss the effects of

using the MT-based approach to Arabic CLIR. We conclude our study in Section 5. All of our experimental findings use the NIST TREC data and relevance rankings available at the time of our experimentation.

2. PRIOR WORK

2.1 Arabic Information Retrieval

In the MICRO-AIR system [3], using only document titles, the authors compared three options for indexing: words, stems, and roots. Three similarity measures were used: the cosine measure, the Dice, and the Jaccard coefficient. A similar study was conducted by Abu-Salem, et al. [1], to improve the effectiveness of Arabic information retrieval by weighing a query term depending on the importance of the word, the stem, and the root of the query term in the collection. The weights were calculated using the standard *tf-idf* measures. The proposed method, called mixed-stemming, showed an improvement over the word indexing method using both the binary and *tf-idf* weighting schemes. Improvements over the stemming index approach were noted only in the case of binary weighting.

Hasnah [14] investigated full text processing, and passage retrieval for Arabic documents. Hasnah concluded that passage retrieval improves the retrieval precision. These were single language (Arabic) efforts only. No cross-lingual experiments were performed.

Beesley [7] described a morphological analyzer system of the modern Arabic standard words. An extensive resource of Arabic information retrieval and computational linguistics projects is found in [20].

In the most recent NIST Text Retrieval Conference (TREC-10), Arabic CLIR processing is introduced. However, at the time of the authoring of this paper, results from this conference are, as of yet, unknown.

2.2 Cross-Language Information Retrieval

In Cross-Language Information Retrieval (CLIR), either documents or queries are translated. There are three main approaches to CLIR: machine translation, comparable or parallel corpus, and machine-readable dictionary. Machine Translation (MT) systems seek to translate queries from one human language to another by using context. Disambiguation in machine translation systems is based on syntactic analysis. Usually, user queries are a sequence of words without proper syntactic structure [21]. Therefore, the performance of current machine translation systems in general language translations make MT less than satisfactory for CLIR [19].

In corpus-based methods, queries are translated on the basis of the terms that are extracted from parallel or comparable document collections. Dunning and Davis [12] suggested parallel and aligned corpus techniques. They used a Spanish-English parallel corpus and evolutionary programming for query translation [11]. Landauer and Littman [18] introduced another method for which no query translation is required. Their method is called Cross-Language Latent Semantic Indexing (CL-LSI), and requires a parallel corpus. Unlike parallel collection, comparable collections are aligned based on a similar theme [23].

Dictionary-based methods perform query translation by looking up terms on a bilingual dictionary and building a target language query by adding some or all of the translations. The practicality of dictionary-based translation is increasing due to the greater availability of machine-readable bilingual dictionaries. Moreover, the topic coverage of this technique is less limited than that of parallel corpus since a dictionary typically contains a wider variety of terms than a sample corpus [2].

Ballesteros and Croft [4] developed several methods using MRDs for Spanish-English CLIR. The first experiment was designed to test the effects of word-by-word (WBW) translation using the MRDs on retrieval performance. The average precision dropped 50-60%. The reason behind the low effectiveness is that many noise terms were added. To improve the effectiveness, they introduced the notion of pre-translation and post-translation methods. Ballesteros and Croft [5] also investigated the effect of phrasal translation in improving effectiveness. In their study, they investigated the role of phrases in query translation via local context analysis (LCA) [26] that uses global and local document analysis, and local feedback (LF). As an extension of [5], Ballesteros and Croft [6] proposed new methods to disambiguate the terms translation via MRD. Co-Occurrence statistics (CO) were used to resolve the ambiguity. They assumed that the correct translation of query terms should co-occur in target language documents and incorrect translation should tend not to co-occur. Pirkola [21] studied the effects of the query structure and setups in the dictionary-based method. Pirkola used a general dictionary and a domain specific (medical) dictionary.

3. DICTIONARY-BASED METHODS

The behaviors of certain techniques differ across languages, particularly languages from different origins, and our focus is strictly on Arabic-English processing. In spite language differences, adapting successful approaches from other languages to Arabic should be investigated. Thus, initially, we adopt some of the prior dictionary-based CLIR approaches, and then, we also develop an additional approach.

3.1 Every-Match Method

The Every-Match (EM) method studies the effects of simple word-by-word translation on Arabic-English retrieval performance by translating Arabic queries word-by-word via a MRD. Dictionary definitions often provide many senses for a single word. In this method, we retain every possible translation when more than one alternative is present, namely, we replace each term with every exact term match in the bilingual term list [5,19]. For example, query number 468 (*incandescent light bulb*) after translation into Arabic appears as (مصباح ضوئي وهـاج). In Table 1, we illustrate the EM method via an example.

The Arabic query words are translated by replacing them by their target English language equivalents. As shown in Table 1, the simple dictionary translation via MRD yields ambiguous translations. It is obvious that the number of word senses increases when the Arabic language word is translated to a target English language by all the equivalents.

Arabic Terms	EM Method
مصباح	lamp light burner
ضوئي	brightness light gleam glow illumination
وهـّـاج	glowing incandescent candescent candent ardent fervent white-hot red-hot blazing flaming radiant brilliant bright resplendent flamboyant glaring dazzling glittering glistening sparkling flashing

Table 1. Terms of the original Arabic query, and the result of the EM method

3.2 First-Match Method

In the First-Match (FM) method [4,,19], only the first match translation per query term is retained instead of using all of the listed translations. In Table 2, we illustrate an example of the Arabic query (مصباح ضوئي وهـاج) and the translations obtained using the First-Match method. As illustrated, in this case, the translations obtained by the FM method appear more precise than those obtained via the EM method.

Arabic Term	FM Method
مصباح	light
ضوئي	brightness
وهـّـاج	glowing

Table 2. Terms of the original Arabic query, and the result of the FM method

3.3 Two-Phase Method

To reduce the ambiguity of the every match method, but to loosen the inherent restrictions of the first match method, we introduce a method that uses some, but not all of the translations of a given Arabic term. The underlying assumption behind the Two-Phase (TP) method is that $f^{-1}(f(x)) = x$, namely, the translation of the translation of the term should yield the original term. If this condition holds, the translation is valid and does not introduce drift or noise.

Let *A* represent the original Arabic terms.

Let *E* represent the translated English terms of *A* using the Every-Match method.

Let *A'* represent the translated Arabic terms of *E* using the Every- Match method.

Then, the Two-Phase method can be implemented as follows:

Translate original Arabic terms A into English terms E using the Every-Match method via an Arabic-English dictionary.

Translate the English terms E to the Arabic terms A' using the Every-Match method via an English-Arabic dictionary.

Return the original Arabic terms A, and the translated Arabic terms A', to their infinitive form.

A candidate English term of E is one that it yields to its original Arabic term based on the comparison between A and A'.

In the rare case when the original terms do not yield a candidate translation term, the following modification is incorporated into the algorithm:

1. If an English term in E does not yield its original Arabic term in A, then:

Find the synonyms of the English term; translate them using the Every-Match method, each translated synonym that matches the original Arabic term A is selected as candidate translation.

2. If neither the English term nor its synonyms in E yield the original term, use the first match term in E as a candidate translation.

In Table 3, we illustrate an example of the original Arabic query (مصباح ضوئي وهـاج), as translated by the TP method.

Arabic Term	TP method
مصباح	lamp light
ضوئي	light
وهـّـاج	glowing incandescent candescent candent ardent fervent red-hot blazing flaming radiant flamboyant glaring flashing

Table 3. Terms of the original Arabic query, and the result of the TP method

As shown in Tables 1, 2 and 3, the TP method removes 13 terms from all possible translations found in the dictionary. The term *burner* results from the translation process of the original Arabic term (ضوئي) using the machine-readable dictionary. This term is a noise term since it is irrelevant to original query. Similarly, the terms "*brightness gleam glow illumination white-hot brilliant bright resplendent dazzling glittering glistening sparkling"* are filtered out reducing the extraneous terms.

3.4. Experimental Approach

We initially describe some of the Arabic complexities that impact the query term translation and then overview the resources used to conduct the experiments.

3.4.1 Pre-processing of Query Terms

Unlike English, in Arabic, nouns can be masculine or feminine and can be definite as in (المعلم) or indefinite as in (معلم). Adding the prefix (ال) makes the difference. Plurals in Arabic are of three kinds: masculine, feminine, and broken. The plural is formed via

suffixes or via pattern modification of the nouns. In the first case, the suffix ~uun for the accusative and genitive as in (معلمين) or ~oon for the nominative (معلمون) is appended to the masculine noun. While ~aat (معلمات) is appended to the plural feminine noun, and the letter "h" is attached to the end of the word to form singular feminine noun as in (معلمة). The dual is formed by adding (ان) or (ين) at the end of the noun as in (معلمان). In the third case, often referred to as broken plurals, the pattern of the singular noun is dramatically altered. The broken plurals can be recognized using patterns.

Another kind of suffixation is the personal pronouns. The personal pronoun can appear as an isolated form or as suffixes attached to the nouns, verbs, or prepositions. The suffixed pronouns can be verbal or nominal. The verbal suffixes express the nominative as in (كتبت) **,** (استعينوا), (أكرما) or the accusative as in (شاركك)**.** Certain suffixes are attached at the end of words to make them possessive pronouns. The letter (ي) is appended to the end of the word (بيت) to form "my house" as in English. For the plural, the letters (هم) are attached for the masculine nouns as in (بيتهم), and the letters (هن) for the feminine nouns as in (بيتهن). These are the most common modifications to the nouns and verbs.

Dictionaries do not store every form of regular words. Most of the dictionary entries are stored in singular and in indefinite form except the words that are usually used in the plural like (كماليات) which means "luxuries" in English. Verbs are stored in perfect form. Therefore, before matching the Arabic terms in the dictionary, some of the nouns must be returned to their singular form by removing all suffixes and prefixes. The procedure of removing the affixes is performed when the process of matching fails to find the source terms in the dictionary.

Arabic verbs have three aspects: perfect, imperfect, and imperative. Perfect forms refer to completed action as (كتب) in English "he wrote". Imperfect verbs refer to incomplete actions; it is commonly used for present or future forms as in (يكتب) "he writes" for singular form and (يكتبون) for masculine plural. For feminine, the word become (تكتب) or (تكتبين). Imperative verbs indicate an action that behaves as a command; so the speaker tells the listener to carry out an action as in (اكتب) for masculine and (اكتبي) for feminine. To conduct the Two-Phase method as described in Section 3.3, the verbs are returned to their infinitive form. The infinitive form is a noun that derived from the verbs without connected to the time. In our example, it becomes (كتابة), in English "Writing".

3.4.2 Experimental Environment

The Text Retrieval Conference (TREC) collections have three distinct parts: the documents, the topics, and the relevance judgments. To provide for direct comparison, we evaluated all three approaches using our search engine AIRE [9] on both the commonly used 2 GB subset of the TIPSTER collection and the 10 GB web data from TREC. For queries, we used a human translation of the TREC-7 (topics 351-400) and TREC-9 (topics 451-500) queries as our original Arabic queries. Since in practice most queries are only a few words long, we used the query titles representation of the 351-400 and 451-500 topics.

A native Arabic speaker manually translated the 100 queries from English into Arabic, and we used these translated versions as our original Arabic queries issued against the TREC English collection. The Arabic queries were translated back to English by means of dictionaries. This method is often used in dictionary-based CLIR studies [21]. To compare the effectiveness of the translated queries, we compare the results of the translated queries to the performance of the monolingual retrieval. The dictionary provides word and common phrases translations. Phrase based translations were used as appropriate.

Currently, TREC provides Arabic queries. However, such queries were unavailable at the time of our experimentation, and still, as of the time of the camera-ready deadline for this paper, they do not have associated relevance judgments. Hence, at present, we do not evaluate our approach using these standard queries, but intend to do so in the near future once relevance judgments become available.

We chose the Al-Mawrid Arabic-English and English-Arabic dictionary [10] in the translation process. Al-Mawrid is a bilingual dictionary with two sections: English-Arabic which has more than 100,000 entries and Arabic-English which has more than 67,000 entries; it is considered the most comprehensive and accurate bilingual dictionary. Al-Mawrid is the officially authorized dictionary by the United Nations (UN) as well as the most commonly used by academic institutions. It is specially designed for human understanding.

3.5 Results

Using the TREC data and queries described earlier, we evaluated our Arabic-English CLIR approaches. In all cases, the translated Arabic to English queries resulted in low retrieval accuracy (as measured by the average precision and recall) as compared with that of the original English queries. The results using the original and the translated queries for titles of TREC topics 351-400 and 451-500 are shown in Tables 4 and 5. As shown, for both data sets, the EM consistently performed the poorest while the TP method was consistently the best. Note that no relevance feedback was used in any of the runs.

	Average Precision	% Monolingual
Original	0.1737	
EM	0.0895	51.5%
FM	0.1197	68.9%
TP	0.1243	71.5%

Table 4. Average precision of queries 351-400

	Average Precision	% Monolingual
Original	0.1249	
EM	0.0566	45.3%
FM	0.0809	64.7%
TP	0.0862	69.0%

Table 5. Average precision of queries 451-500

In Tables 6 and 7, we demonstrate the effects on the precision-recall measure for the original and the three translation methods at 5, 10, 15, 20, and 30 top retrieved documents. Columns one through four correspond to the original queries, the Every-Match, the First-Match, and the Two-Phase methods, respectively.

Precision	Original	EM	F M	TP
at 5 Docs	0.4240	0.1920	0.2440	0.2560
at 10 Docs	0.3780	0.1840	0.2320	0.2460
at 15 Docs	0.3387	0.1680	0.2187	0.2240
at 20 Docs	0.3210	0.1610	0.2170	0.2190
at 30 Docs	0.2733	0.1447	0.1900	0.1967

Table 6. Precision at 30 documents retrieved of queries 351-400

As shown in Tables 6 and 7, again, the TP method outperforms the EM and the FM methods at 5, 10, 15, 20, and 30 top retrieved documents. A comparison of the retrieval performance of the three runs is shown in Figures 1 and 2. As shown, the TP approach outperforms all the other methods. At the higher-precision lower-recall levels (recall up to 0.3), the difference between the TP method and the other methods is even more noticeable.

Precision	Original	EM	FM	TP
at 5 Docs	0.1755	0.1061	0.1429	0.1667
at 10 Docs	0.1592	0.0918	0.1143	0.1437
at 15 Docs	0.1578	0.0816	0.1102	0.1194
at 20 Docs	0.1449	0.0735	0.1020	0.1042
at 30 Docs	0.1367	0.0728	0.0986	0.1000

Table 7. Precision at 30 documents retrieved of queries 451-500

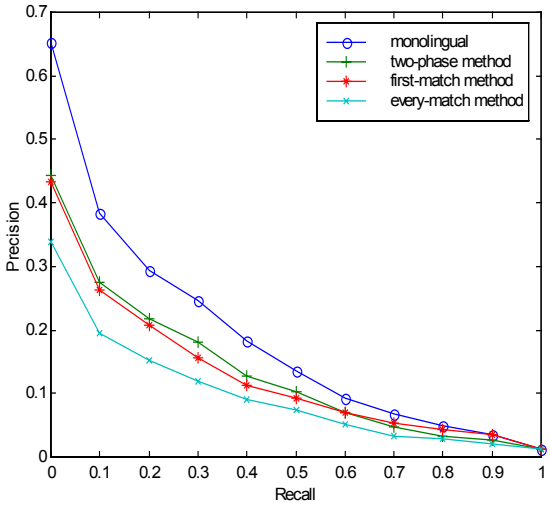

Figure 1. Average precision and recall of queries 351-400

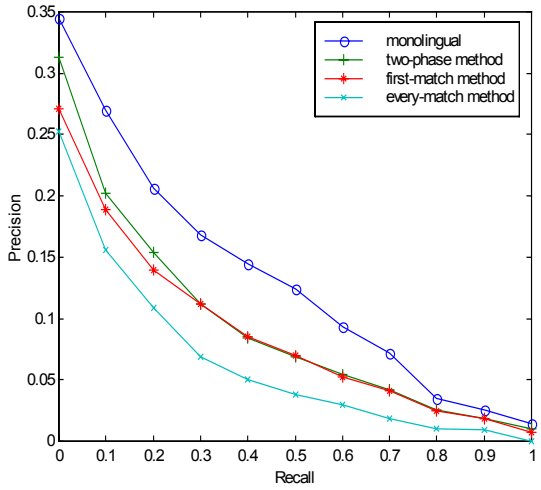

Figure 2. Average precision and recall of queries 451-500

	TP vs. EM	FM vs. EM	TP vs. FM
TREC-7	α= 0.01	α=0.01	α = 0.1404
TREC-9	α= 0.01	α= 0.01	α= 0.1090

Table 8. Statistical Significance Test

In Table 8, we summarize the statistical significant test interpretation of our experiments. The evaluation is conducted using the *paired t-test* [25]. The obtained α values demonstrate that the performance differences of the TP and FM methods over the EM method are significant at a 99% confidence interval for both the TREC-7 and TREC-9 datasets. Less significant are the performance differences between the TP and FM methods that are significant at an 86% (α = 0.1404) and 89% (α = 0.1090) confidence interval for the TREC-7 and TREC-9 datasets, respectively.

4. MACHINE TRANSLATION METHOD

Machine Translation systems can be defined as any computer-based process to transform a text from one language into another language without human intervention. The basic task of any machine translation system is to analyze the source text, including morphological, syntactic, and semantic analysis using special purpose lexicons, and target language generation. Therefore, a machine translation strategy for CLIR might allow the researchers to take advantage of the extensive research on machine translation and the availability of commercial products. There are two basic approaches to MT, translating the documents or the queries.

Many researchers criticize the MT-based CLIR approach. The reasons behind their criticisms mostly stem from the fact that the current translation quality of MT is poor. In particular, typical

search terms lack the context necessary for the MT system to correctly perform proper syntactic and semantic analysis of the source text. Another reason is that MT systems are expensive to develop, and their application degrades the retrieval efficiency (run time performance) due to the cost of the linguistic analysis.

Hull and Grefenstette [15] stated that current MT systems, in the setting of general language translation, are less than satisfactory for CLIR. A study by [22] compared the retrieval effectiveness of the French-English CLIR using SYSTRAN machine translation system with the effectiveness of their EMIR dictionary-based query translation. They determined that the EMIR was more effective than their MT-based query translation technique using SYSTRAN.

Other researchers, in contrast, showed that machine translation approaches could achieve reasonable effectiveness [17]. Jones, et al. [16], showed that full disambiguation by a MT system outperforms dictionary lookup methods that include several terms as candidates in the query. Also, participants in the TREC-8 CLIR track [8] concluded that MT-based CLIR is an effective strategy.

We are only in the initial phase of our Arabic Machine Translation research. Thus, our experiments aim only to provide insight rather than draw conclusions regarding the performance of the MT-based query translation approach on a large document collection. Further work is needed to better evaluate the MT approach. The MT system that we adapted for our experiments is a commercial product that is designed to assist humans by automatically translating full sentences, or even a paragraph. For higher accuracy, if the query terms are formulated as a phrase, we apply the MT system on the phrase as well.

4.1 Arabic-English MT System Descriptions

We used the ALKAFI Arabic to English machine translation system, a commercial system developed by CIMOS. ALKAFI is the first Arabic to English machine translation system. It produces translations in a wide range of subject areas like science and technology; commerce and banking; computers and the petroleum industry at about 60,000 words per hour. Usually, the Arabic text is not vocalized so ALKAFI can add vowels internally. But sometimes, the user must vocalize selectively some consonants to help ALKAFI with lexical and syntactic analysis. The vocalization is a very important step because the word sense depends on vocalization and on the place of the word in the sentence. It has a strong parser, a deep syntactic analysis and a selective semantic analysis to detect main verbs, phrasal verbs, and idioms. The system attempts to analyze words in context and then builds semantic links. The analysis process is ended by an internal representation of the sentence. The English text is generated by the transfer method according to the grammar rules of the English language.

4.2 Experimental Approach

The TREC-7 collection and topics are used as described in Section 3.4.2. The translation process is done by translating each part of the query topic separately to study the effectiveness of query translation using the ALKAFI system. For example, Table 9 shows the title of the original Arabic query terms that manually translated from query topic 384 and their translation using ALKAFI machine translation system. The title of the original English query consists of 3 terms. Actually, most of the titles of the query topics of 351-400 consist of just one or two terms. In Table 10, we show the description field (medium length) of the original Arabic query and the translated English query using the ALKAFI MT system.

Original Arabic Query	محطّة القمر الفضائيّة
Translated English Query	The spatial station of the moon

Table 9. Terms of the title field of the original Arabic query and the translation using MT system

Original Arabic Query	عرّف الوثائق التي تناقش بناء محطّة فضاء مع غايةاستعمار القمر؟
Translated English Query	Define the documentations which discuss the building of a space station with the purpose of the moon colonialism ?

Table 10. Description field of the original Arabic query and the translation using MT system

In Table 11, we provide an example of a long query, the narrative of the original Arabic query (384) is translated using the MT system. The titles of query topics 351-400 contain 137 terms with an average of 2-3 terms.

Original Arabic Query	الوثائق ذات العلاقة ستناقش الغرض من محطّة الفضاء ، المبادرة نحو استعمار القمر ، العوائق التي عاقت المشروع حتّى الآن ، الخطط الجارية حاليًا أو في مرحلة التّخطيط مثل المغامرة ، التّكلفة ، البلدان الذين تعهّدوا بالرجال ، الموارد ، التسهيلات ، و المال لإنجاز مثل هذا العمل.
Translated English Query	The documentations are related you will discuss the purpose from the station of the space, the initiative toward is the colonialism of the moon, the discouragements which she impeded the project until now, the current schemes are now or in planning stage the example of the adventure, the cost, the countries who you advocate by the men, the resources, the facilities, and the money a such carrying out has the work.

Table 11. Narrative field of the original Arabic query and the translation using MT system

4.4 Results of the MT-based Method

In Table 12, we summarize the average precision results for the TREC-7 collection and topics (351-400). The machine translation results are better than the Every-Match method (EM) in all runs. It yields to 61.3% of monolingual retrieval. As shown in Table 12, both the FM and TP methods outperform the machine translation approach. The reason behind the degraded effectiveness of the machine translation is that the used machine translation system is designed to perform best on well-formed sentences or at least any sequence of words that form a context. The titles of topics 351-400 are all three words or less. The effect of greater context is also

apparent in the performance of machine translation using the description field of query topics 351-400 as shown in Table 13. A comparison of the baseline, EM, FM, TP, and MT methods is represented in Figure 3 using the average precision and recall.

In Table 14, we illustrate the results of the top 5, 10, 15, and 30 documents retrieved. Actually, this Table gives more realistic performance of CLIR since it is unexpected for the foreign users to read many retrieved documents [4,15]. As shown, the MT-based technique outperforms the EM method, but again is lower than FM and TP methods. As shown in Table 13, the description field yields 64.6% and the narrative field yields 61.9% of the monolingual retrieval. According to these findings in Tables 12 and 13, we conclude that the MT system performs best once a context is determined. That is, adding more terms to the full context query does not help the machine translation to disambiguate the term pulse. Since our results are only preliminary, no additional conclusions are drawn.

	Average Precision	% Monolingual
Original	0.1737	
EM	0.0895	51.5%
FM	0.1197	68.9%
TP	0.1243	71.5%
MT-title	0.1066	61.3%

Table 12. Average precision of queries 351-400 of the four runs

	Original	MT method	% Monolingual
Description	0.1839	0.1189	64.6%
Narrative	0.1447	0.0897	61.9%

Table 13. Average precision of the description and narrative fields of query topics 351-400

Precision	Original	EM	FM	TP	MT-title
at 5 Docs	0.4240	0.1920	0.2440	0.2560	0.2080
at 10 Docs	0.3780	0.1840	0.2320	0.2460	0.1840
at 15 Docs	0.3387	0.1680	0.2187	0.2240	0.1827
at 20 Docs	0.3210	0.1610	0.2170	0.2190	0.1940
at 30 Docs	0.2733	0.1447	0.1900	0.1967	0.1693

Table 14. Comparison between MT-based method and dictionary-based methods measured by precision at 30 documents retrieved of queries 351-400.

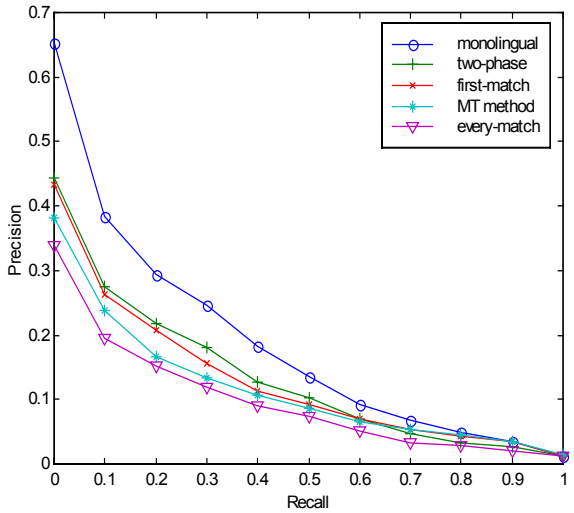

Figure. 3. Average precision and recall of queries 351-400 for the five runs

5. CONCLUSIONS

Our results demonstrate the potential of Arabic-English CLIR. Machine-readable dictionaries are cost effective as compared to the other methods such as parallel corpus, and Latent Semantic Indexing (LSI). The resources needed are readily available. The ambiguity introduced by the Every-Match (EM) method yields poor effectiveness; it achieved roughly half of the performance of the monolingual retrieval. A key factor affecting this is the transfer of too many senses that are inappropriate to the query.

It is common for a single word to have several translations, some with different senses. To reduce the number of extraneous terms, the First-Match (FM) technique was evaluated. This approach achieved 68.9% and 64.7% of the titles of TREC topics 351-400 and TREC topics 451-500, respectively. The drawback of this method is that many terms that are related to the original queries may be ignored. Therefore, we proposed a new method called the Two-Phase (TP) method. In the TP method, we ignore all the terms that do not retranslate to the original Arabic query word. This method achieved 71.5% and 69.0% of monolingual retrieval by using titles of TREC topics 351-400 and TREC topics 451-500, respectively. We found that our TP results were statistically significant at greater than a 99% and an 86% confidence interval over the EM and FM methods, respectively. We also conducted initial experiments with a commercial MT-based Arabic-English CLIR; we found its performance inferior to that of the FM and TP methods.

Our future work is to enhance the TP method by finding an appropriate weighting mechanism for each term in the query. For example, original English terms that return the original Arabic terms are assigned more weight than the synonyms of English terms that return the original Arabic terms. Another extension is to use term thresholds for the TP method. Instead of using all terms in the first phase, using term thresholds, only retaining the top terms. In this study, we have shown that eliminating unrelated terms by the TP method can significantly reduce the error associated with dictionary translation.

6. REFERENCES

[1] Abu-Salem, H., Al-Omari, M., and Evens, M. Stemming Methodologies over Individual Query Words for an Arabic Information Retrieval System. JASIS 50(6): 524-529, 1999.

[2] Adriani, M., and Croft, W. The Effectiveness of a Dictionary-Based Technique for Indonesian-English Cross-Language Text Retrieval. CLIR Technical Report IR-170, University of Massachusetts, Amherst, 1997.

[3] Al-Kharashi, I., and Evens, M. Comparing Words, Stems, and Roots as Index Terms in an Arabic Information Retrieval System. JASIS 45(8): 548-560, 1994.

[4] Ballesteros, L., and Croft, B. Dictionary Methods for Cross-Lingual Information Retrieval. 7th DEXA Conf. on Database and Expert Systems Applications. Pages 791-801, 1996.

[5] Ballesteros, L., and Croft, B. Phrasal Translation and Query Expansion Techniques for Cross-language Information Retrieval. SIGIR 1997, 84-91.

[6] Ballesteros, L., and Croft, B. Resolving Ambiguity for Cross-Language Retrieval. SIGIR 1998, 64-71

[7] Beesley, K Arabic Morphological Analysis on the Internet. Proceedings of the 6th International Conference and Exhibition on Multi-lingual Computing, Cambridge, 1998.

[8] Braschler, M., Peters, C. and Schuable, P. Cross-Language Information Retrieval (CLIR) Track Overview, TREC-8 Proceedings. 2000.

[9] Chowdhury, et. al, "AIRE in TREC-9", Proceedings of TREC-9, NIST, 2001.

[10] Dar El-Ilm Lilmalayin, http://www.malayin.com/

[11] Davis, M., and Dunning, T. Query Translation using Evolutionary Programming for Multilingual Information Retrieval. The 4th Evolutionary Programming Conf., 1995.

[12] Dunning, T. and Davis, M. Multi-lingual information retrieval. Technical Report MCCS-93-252. Computing Research Laboratory, New Mexico State University. 1993.

[13] Egyptian Demographic Center, www.frcu.eun.eg/www/homepage/cdc/cdc.htm.

[14] Hasnah, A. Full Text Processing and Retrieval: Weight Ranking, Text Structuring, and Passage Retrieval for Arabic Documents. C.S. Ph. D. Dissertation, Illinois Institute of Technology, 1996.

[15] Hull, D. and Grefenstette, G. Querying Across Languages: A Dictionary-based Approach to Multilingual Information Retrieval. In proceedings of the 19th Annual international ACM SIGIR 1996, Zurich, Switzerland, 49-57.

[16] Jones, G., Sakai, T., Collier, N., Kumano, K., and Sumita, K. A Comparison of Query Translation Methods for English-Japanese Cross-Language Information Retrieval. SIGIR 1999, 269-270.

[17] Kowk, K.L. English-Chinese Cross-Language Retrieval based on a Translation Package. Post-Conference Workshop on Machine Translation for Cross Language Information Retrieval at AAMT Machine Translation Summit VIII. 1999.

[18] Landauer, T. K., and Littman, M. L. Full Automatic Cross-Language Document Retrieval using Latent Semantic Indexing. The 6th Conf. of UW center for New OED and Text Research, pp. 31-38, 1990.

[19] Oard D. A Comparative Study of Query and Document Translation for Cross-language Information Retrieval. In Machine Translation and the Information Soup. 3rd Assoc. for Machine Transl. in the Americas Conf., 472-83, 1998.

[20] Oard, D., http://www.clis.umd.edu/dlrg/clir/arabic.html

[21] Pirkola, A. The Effects of Query Structure and Dictionary Setups in a Dictionary-based Cross-Language Information Retrieval. SIGIR 1998, Melbourne, Australia.

[22] Radwan, K., Fluhr, C. Textual Database Lexicon Used as a Filter to Resolve Semantic Ambiguity Application on Multilingual Information Retrieval. The 4th Symp. on Document Analysis and Information Retrieval, 121-136, 1995.

[23] Sheridan, P. and Ballerini, J.P. Experiments in Multilingual Information Retrieval using the SPIDER System. The 19th Annual International ACM SIGIR 1996, 58-65.

[24] Tayli, M., and Al-Salamah, A. Building Bilingual Microcomputer Systems. In Communications of the ACM, Vol. 33, No.5, Pages 495-505, 1990.

[25] Wonnacott, R., Wonnacott, T. Introductory Statistics, John Wiley & Sons, Fourth Edition,1990.

[26] Xu, J. and Croft, W. B. Query Expansion using Local and Global Document Analysis. The 19th Annual International ACM SIGIR 1996, Zurich, Switzerland, Pages 4-11.