On Bidirectional English-Arabic Search

M. Aljlayl, O. Frieder, & D. Grossman
Information Retrieval Laboratory
Illinois Institute of Technology
{aljlayl, frieder, grossman}@ir.iit.edu

Abstract

In Cross-Language Information Retrieval (CLIR), queries in one language retrieve relevant documents in other languages. Machine-Readable Dictionaries (MRD) and Machine Translation (MT) systems are important resources for query translation in CLIR. We investigate the use of MT systems and MRD to Arabic-English and English-Arabic 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 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 Arabic-English MT approach using short, medium, and long queries of TREC7 and TREC9 topics and collections. The effects of the query length to the quality of the MT-based CLIR are investigated. English-Arabic CLIR is evaluated via MRD and English-Arabic MT. The query expansion via post-translation approach is used to de-emphasize the extraneous terms introduced by the MRD and MT for English-Arabic CLIR.

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 Web pages. Automatic mediation of access to foreign Web pages is becoming an increasingly important problem. Therefore, the importance of CLIR is noticeable. 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 contrast, English-Arabic CLIR is the retrieval of Arabic documents based on queries in the English language.

In a dictionary-based approach, translation is performed by looking up the terms in the bilingual dictionary and forming a target query by considering one or more than one translation per query term. To achieve this goal, we investigate three approaches

associated with Machine-Readable Dictionaries (MRD) for Arabic-English CLIR. The Every-Match method considers all the translations found in a bilingual dictionary. This leads to ambiguous translation 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. The First-Match method ignores some of the less common translations of the source language, and thus, potentially improves the retrieval effectiveness. The third method is the Two-Phase method. This method considers all the translations found in the bilingual dictionary as candidate terms then it removes the translated candidate terms that do not return its original source query term. We also empirically evaluate the effectiveness of the Arabic-English MT-based approach using short, medium, and long queries of TREC topics and collections. The effects of the query length to the quality of the MT-based CLIR are likewise investigated.

English-Arabic CLIR is evaluated via MRD and English-Arabic MT system. The post-translation expansion approach is used to de-emphasize the extraneous terms introduced by the MRD and MT. We found that query expansion after translation via post-translation approach yields significant improvement on the performance of the English-Arabic CLIR.

Arabic, one of the six official languages of the United Nations, is the mother tongue of 300 million people (Egyptian Demographic Center, 2000). 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 (Tayli and Al-Salamah, 1990), 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. In formal writing, Arabic sentences are delimited by commas and periods as in English, for instance.

In Section 2, we review the prior work in CLIR and also specifically on Arabic information retrieval. The proposed dictionary-based methods for Arabic-English and English-Arabic CLIR are presented in section 3. The effects of the MT-based approach on Arabic-English and English-Arabic CLIR are investigated in Section 4. We conclude our study in Section 5.

2. Prior work

We begin with an overview of prior work done in Arabic information retrieval. We continue with a review of other prior CLIR efforts because some of this prior work can be easily adapted to Arabic language processing, and in fact, part of our work includes this adaptation.

2.1 Arabic Information Retrieval

In the MICRO-AIR system (Al-Kharashi and Evens, 1994), 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. The result of these experiments showed that using roots as index terms was more efficient than using words or stems. A similar study was conducted by (Abu-Salem, et al., 1999). The authors attempted 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, which is 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 (1996) investigated full text processing, and passage retrieval for Arabic documents. Hasnah concluded that passage retrieval improves the retrieval precision. Beesley (1998) described a morphological analyzer system of the modern Arabic standard words. These were Arabic monolingual retrieval efforts only. No cross-lingual

experiments were performed. Recent Arabic monolingual and English-Arabic CLIR resources are found in the TREC web site (TREC, 2001).

2.2 Cross-Language Information Retrieval (CLIR)

The rapid growth of the Internet has created worldwide multilingual document collections. Accordingly, IR research has begun to pay attention to CLIR systems. In CLIR, either documents or queries are translated. The research has focused on the accuracy of query translation since document translation is computationally expensive (Hull and Grefenstette, 1996).

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 few words without proper syntactic structure (Pirkola, 1998). Therefore, the performance of current machine translation systems in general language translations make MT less than satisfactory for CLIR (Radwan and Fluhr, 1995; Hull and Grefenstette, 1996). Another study by Oard (1998), however, did confirm that machine translation does yield reasonable effectiveness in the case of long queries.

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 (1993) suggested parallel and aligned corpus techniques. They used a Spanish-English parallel corpus and evolutionary programming for query translation (Davis and Dunning, 1995). Landauer and Littman (1990) 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 (Sheridan and Ballerini, 1996).

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. Dictionary-based translation is very practical with the increasing availability of machine-readable bilingual dictionaries (MRD). 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 parallel corpus (Adriani and Croft, 1997). Ballesteros and Croft (1996) developed several methods using MRD for Spanish-English CLIR. The first experiment was designed to test the effects of word-by word translation using the MRD on retrieval performance. Each query word was replaced by the corresponding word or words in the dictionary. 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 (1997) 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) (Xu and Croft, 1996) that uses global and local document analysis, and local feedback (LF). They concluded that combining pre and post translation expansion is more effective and improves precision and recall. As an extension of (Ballesteros and Croft, 1997), Ballesteros and Croft (1998) proposed new methods to disambiguate the terms translation via MRD. A Co-Occurrence statistics (CO) method was used to resolve the ambiguity. They assumed that the correct translation of query terms co-occur in target language documents and incorrect translation tend not to co-occur. A combined approach of pre and post translation yielded better effectiveness.

Pirkola (1998) studied the effects of the query structure and setups in a dictionary-based approach. The effectiveness of the English queries against English documents was compared to the performance using translated Finnish queries. Pirkola used a general dictionary and a domain specific (medical) dictionary. Hull and Grefenstette (1996) performed experiments at Xerox to build a multilingual IR system to understand the factors that drive effectiveness. The percentages of the original English queries are: automatic word-based dictionary 59%, manual word-based dictionary 68%.

Our initial efforts in Arabic-English CLIR are described in (Aljlayl and Frieder, 2001). Here, we extend the results presented in there, and also address the reverse problem of English-Arabic CLIR. Given the shear number of previous efforts using MRD and MT approaches, one tends to believe in their practicality. Furthermore, the topic coverage is

wider than that of parallel corpus. The effectiveness of these methods depends on the ability to choose the right term from many possible terms.

3. Dictionary-based Approaches

Unlike others, our efforts target the Arabic language. We adapt some of the prior dictionary-based CLIR approaches, particularly those of Ballesteros and Croft, to the Arabic language as well as develop an additional approach for Arabic-English CLIR. It is common for a single word to have several translations, some with very different senses. Removing the noise terms increases the retrieval performance; so taking this into account, we designed and implemented three dictionary-based query translation methods for Arabic-English CLIR and one approach for English-Arabic CLIR.

3.1 Every-Match Method

The Every Match (EM) method is designed to study the effects of simple word-by-word translation on retrieval performance and to determine the factors that produce these effects. The Arabic queries are translated 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 in the term list in the MRD. We replace each term with every exact term match in the bilingual term list (Oard, 1998). For example, query number 468 (*incandescent light bulb*) after translation into Arabic appears as (مصباح ضوئي وهـــاج). Now we apply the EM method to this query. 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. The average query length after translation via EM method is 10 and 28 terms for TREC-7 and TREC-9, respectively.

Original Arabic Terms	Every Match Method
مصباح	light lamp burner
ضوئـي	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 Every-Match (EM)

3.2 First-Match Method

In the First-Match (FM) method (Oard 1998; Ballesteros and Croft, 1997; Ballesteros and Croft, 1998), 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 FM method. As illustrated, in this case, the translations obtained by the FM method appear more precise than those obtained via the EM approach. The retained terms via applying the FM method are subset of the retained terms via EM method.

Original Arabic Term	First Match Method
مصباح	light
ضوئـي	brightness
وهـّـاج	glowing

Table 2. Terms of the original Arabic query, and the result of the First-Match

For English-Arabic CLIR, Table 3 illustrates an example of the Arabic terms retained by the FM method.

Original English query	First Match Method
Information	اعلام
Technology	التكنلوجيا
Arab	العربي
World	الدنيا

Table 3. English query terms and their translation using FM method

3.3 Two-Phase Method

To reduce the ambiguity of the EM method, but to loosen the inherent restrictions of the FM method, we introduce a method for Arabic-English CLIR that uses some, but not all of the translations of a given Arabic term. The underlying assumption behind the Two-Phase 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 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 a machine-readable dictionary.

Translate the English terms E to the Arabic terms A' using the Every-Match method via English-Arabic machine-readable 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

The Two-Phase method can likewise be illustrated as shown in Figure 1. Circled terms are assumed to be the most appropriate translation of the original Arabic terms. For term a_2 in original Arabic terms, the English translation is e_3 and e_4. Neither e_3 nor e_4 yield the Arabic term a_2 in the original Arabic terms set. To overcome this situation, the approach is to find the synonyms of the term that do not yield the original translation

after the second phase. For example, e_3, does not yield to the original Arabic term a_2, then the synonyms of e_3 are translated to Arabic, every synonym that yields the original Arabic term is chosen as candidate for translation. The first match approach is applied when no synonyms yield to the original Arabic terms.

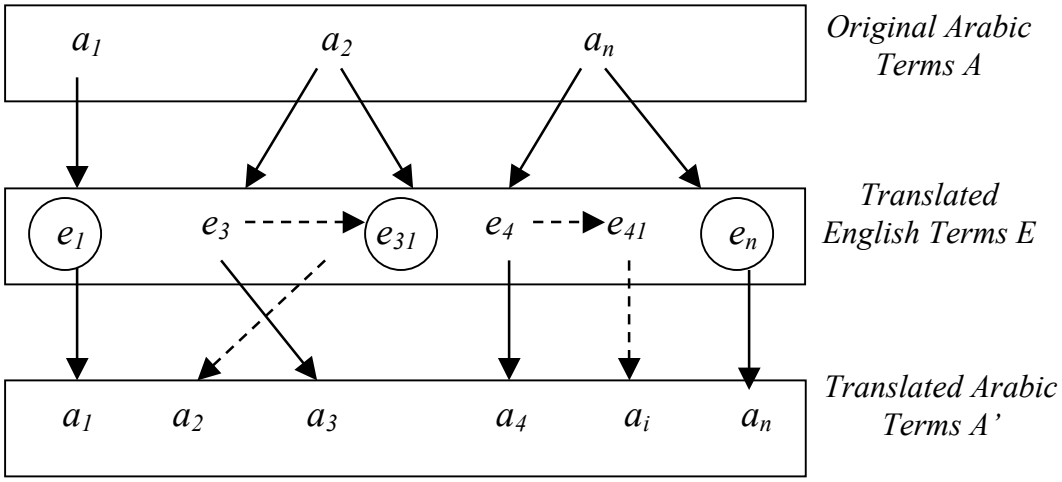

Figure 1. The Process of the Two-Phase Approach

In Table 4, we illustrate an example of the original Arabic query (مصباح ضوئي وهــاج), as translated by the Two-Phase method.

Original Arabic Term	Two-Phase method
مصباح	lamp light
ضوئـي	light
وهـــّاج	glowing incandescent candescent candent ardent fervent red-hot blazing flaming radiant flamboyant glaring flashing

Table 4. Terms of the original Arabic query, and Two-Phase (TP) technique

As shown in Tables 1 and 4, the Two-Phase 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. The retained terms via applying the TP method are subset of the translated terms via EM method. There are overlaps

between term translations via FM and TP methods. In most cases the first match in the dictionary is retained in the translation process of the TP method. The average query length after translation using TP method is 6 and 12 terms for TREC7 and TREC9, respectively.

3.4 Experimental Approach

Some of the Arabic complexities that impact the query term translation are described in Section 3.4.1. In Section 3.4.2, we describe the resources that we used to conduct the experiments.

3.4.1 Pre-processing of the Source Arabic Terms

Unlike the English language, in the Arabic language, nouns can be masculine or feminine. The nouns can be definite as in (المعلم) or in indefinite as in (معلم). Adding the prefix (ال) makes the difference. Plurals in Arabic are three kinds: the masculine plural, the feminine plural, and the broken plural. The plural is formed via suffixes or via pattern modification of the nouns. In the first case, the suffix ~een for the accusative (معلمين) and genitive 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 (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. We can recognize these plurals from the patterns. There are 27 patterns for most of the broken nouns.

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. Certain suffixes are attached at the end of words to make them possessive pronouns. The attached can be one letter, for example (بيتي) when the letter "ي" is attached to the end of the word (بيت) to form "my house" in English. For plural, two letters are attached to the end of the word, for the masculine, the letters "هم" are attached (بيتهم), and the letters "هن" for the feminine nouns (بيتهن) . These are the most common modifications to Arabic words.

Dictionaries do not store every form of regular words. Most of the dictionary entries are stored in singular form except the words that are usually used in the plural like (كماليات) which means "luxuries" in English. The 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.

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". This infinitive form is implemented as a base of comparison in the Two-Phase method.

Similarly, for English-Arabic CLIR, the source English queries are normalized to match them in the dictionary. For example, the terms "p*erforming*" and "*performance*" are normalized to "*perform*".

3.4.2 Experimental Environment

For Arabic-English CLIR, we evaluated all the three dictionary-based approaches using our search engine AIRE (Chowdhury, et al, 2000) on both the commonly used 2 GB subset of the TIPSTER (Disks 4 and 5) collection and the 10 GB web data from TREC. Each of these collections contains over 500,000 documents. To obtain a set of standard test queries, we manually translated the English TREC topics into Arabic TREC topics. The Text Retrieval Conference (TREC) has three distinct parts to the collections used in TREC: the documents, the topics, and the relevance judgments. For queries, we used a human translation of the TREC7 (topics 351-400) and TREC9 (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 approach is often used in dictionary-based CLIR studies (Pirkola,

1998). To compare the effectiveness of the translated queries, our Arabic-English CLIR system compares the results of the translated queries to the performance of the monolingual information retrieval. The dictionary provides words and some phrases as keyword entries. Phrase based translations were used as appropriate. In the translation process, we start to match the phrases in the query to the phrases in the dictionary, if match then the result is retained. If not, then word-by-word translation basis is performed by applying the proposed dictionary-based methods.

For English-Arabic CLIR, we used the Arabic collection that consists of 383,872 documents provided by Linguistic Data Consortium (LDC). TREC provided 25 topics in three parallel languages; Arabic, English and French. Our focus is on English-Arabic CLIR, so we used the English queries as our source queries against the Arabic collection. The titles of the TREC Arabic topics are used.

We chose the Al-Mawrid Arabic-English and English-Arabic Dictionary (aDawliah Universal Electronics) 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 67000 entries; it is considered the most comprehensive and accurate Arabic-English bilingual dictionary. Al-Mawrid is the official dictionary used by the United Nations (UN) as well as most academic institutions. It is specially designed for human understanding. We converted a portion of Al-Mawrid to a transfer dictionary suitable for information retrieval. The dictionary includes word-based and some multi-word expression as a keyword entry. The process of extracting the term lists from the dictionary involved the removal of a large amount of excess information, such as examples and descriptions.

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 5 and 6. As

shown, for both data sets, the Every-Match consistently performed the poorest while the Two-Phase Method was consistently the best. Note that no relevance feedback was used in any of the runs.

	Original	Every Match	First Match	Two Phase
Average Precision	0.1737	0.0895	0.1197	0.1243
% Monolingual		51.5%	68.9%	71.5%

Table 5. Average precision of queries 351-400 of the four runs

	Original	Every Match	First Match	Two Phase
Average Precision	0.1249	0.0566	0.0809	0.0862
% Monolingual		45.3%	64.7%	69.0%

Table 6. Average precision of queries 451-500 of the four runs

In Table 7, we summarize the statistical significant test interpretation of our experiments. The evaluation is conducted using the *paired t-test* (Wonnacott, R. and Wonnacott, T, 1990). 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% ($\alpha = 0.1404$) and 89% ($\alpha = 0.1090$) confidence interval for the TREC-7 and TREC-9 datasets, respectively.

	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 7. Statistical Significance Test

In Tables 8 and 9, 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. Column one corresponds to the original queries. Column two shows the Every-Match method. Column three shows the First-Match method. The last column shows the Two-Phase method.

Precision	Original	Every Match	First Match	Two-Phase
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 8. Precision at 30 retrieved documents of queries 351-400

Precision	Original	Every Match	First Match	Two-Phase
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 9. Precision at 30 retrieved documents queries 451-500

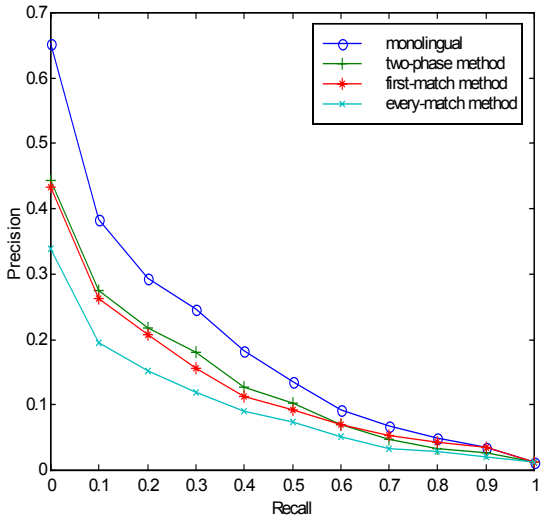

Figure 2. Average precision and recall of queries 351-400

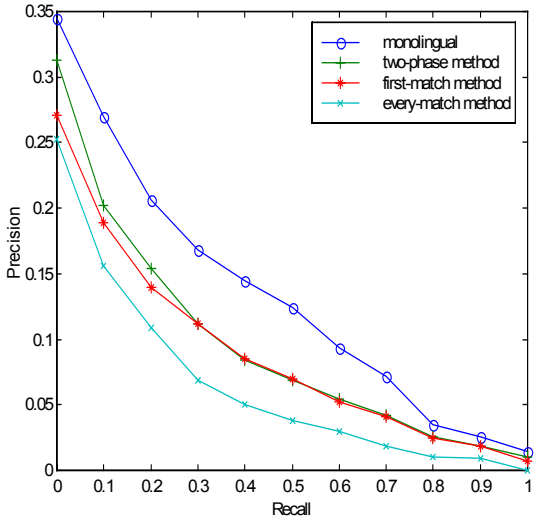

Figure 3. Average precision and recall of queries 451-500

As shown in Tables 8 and 9, again, the Two-Phase method outperforms the Every-Match and the First-Match 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 2 and 3. As shown, the Two-Phase approach outperforms all the other methods. At the higher precision-lower recall levels (recall up to 0.3), the difference between the Two-Phase method and the other methods is even more noticeable. Since it is unrealistic to expect the user to read many retrieved documents that are expressed in a language other than the user's native language, the higher precision region is of greater interest. Therefore, higher precision results obtained by the Two-Phase method are even more significant. As measured in average precision, the First-Match method improves the effectiveness over the Every-Match method by 33.7% and 42.9% for TREC topic 351-400 and TREC topics 451-500, respectively. The Two-Phase method outperforms the First-Match method by 3.8% and 6.5% for TREC topics 351-400 and TREC topics 451-500, respectively.

Table 10 summarizes the results of using the FM method for English-Arabic CLIR. The FM method achieved 60.2% of the monolingual run for English-Arabic CLIR.

Applying the two-phase method in English-Arabic direction requires finding a generalized form of English terms for more chances of matching. In fact, such characteristics of Arabic words make the two-phase more practical since words in Arabic are primarily based on three letter roots. The three letter roots dramatically simplifies the formation of word classes making it practical to use the two-phase method.

	Average Precision	%Monolingual
Monolingual	0.3709	
MRD	0.2236	60.2

Table 10. Average precision of original and cross-language runs using the FM method

The queries were expanded after the translation via post-translation technique (Ballestoreos and Croft, 1996). The post-translation technique expands the query via Pseudo-Relevant Feedback (PRF). The top 15 from the top 10 documents, which are assumed relevant, are added to the translated Arabic terms. In Table 11, the MRD refers to using the First-Match method, MRD+Post refers to using the First-Match method and post-translation expansion.

	MRD	MRD+Post
Average Precision	0.2236	0.2605
%Improvement		16.5
Statistical test		P=0.05

Table 11. Average precision of the FM method before and after query expansion

As described in Table 11, feedback after translation improved the effectiveness by 16.5% over the FM method without post-translation expansion. The differences between the MRD on title and post-translation expansion of the translated Arabic queries are statistically significant at 95% confidence level.

Since it is not realistic for the foreign users to read many retrieved documents, we demonstrate the effects on the precision-recall measure for the MRD and the MRD+post

translation approaches at lower levels of recall, up to 1000 documents retrieved. In Table 12, column one corresponds to MRD without expansion. As shown in column three, the MRD augmented with query expansion via PRF outperforms the MRD without expansion.

Precision	MRD	MRD+Post
at 5 Docs	0.3840	0.4480
at 10 Docs	0.3160	0.4040
at 15 Docs	0.3013	0.3733
at 20 Docs	0.2960	0.3540
at 30 Docs	0.2933	0.3293
at 100 Docs	0.2268	0.2712
at 200 Docs	0.1752	0.2076
at 500 Docs	0.1173	0.1333
at 1000 Docs	0.0811	0.0863

Table 12. Precision at 1000 retrieved documents of MRD and MRD+post runs

4. MT-based approach

We explore the retrieval effectiveness of Machine Translation (MT) systems for Arabic-English and English-Arabic Cross-Language Information Retrieval (CLIR), as well as what factors affect performance, and to what extent. As mentioned in Section 2.2, one of the approaches being tested for CLIR makes use of existing machine translation systems to provide automatic translation of the queries or documents, from one language to another. The basic task of any machine translation system is to analyze the source text, including morphological, syntactic, and semantic analysis using bilingual dictionaries or 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. The drawbacks of the document translation approach, as compared to translating the queries, are the extensive processing required to translate very large amount of data, and in the case of multiple query languages, the need to duplicate the documents in all of the query languages. In the case of translating the queries, Oard (Oard, 1998) discussed the

technique and concluded that it is less costly than translating the documents. This provides an obvious approach to query translation.

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. A study by (Radwan and Fluhr, 1995) 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. Jones, et al. (1999), showed that full disambiguation by a MT system outperforms dictionary lookup methods that include several terms as candidates in the query. Also, many participants in the TREC-8 CLIR track (Braschler et al., 1999) concluded that MT-based CLIR is an effective strategy. Another advantage of using MT systems for CLIR is that if L_1-L_2 MT and L_2-L_3 MT systems are available, it is possible to construct a L_1-L_3 CLIR system without developing a L_1-L_3 MT system, where *L1, L2*, and *L3* are three different languages (Kwok, 1999).

Our experiments provide insight into the performance of the MT-based query translation approach on a large document collection described in Section 3.4.2. The machine translation systems that we adapted for our experiments are commercial products that are designed to assist humans by automatically translating full sentences, or even a paragraph. For higher accuracy, if the query terms are formulated as phrases, we can apply MT systems as well. However, experience shows that users typically prefer to give isolated words, or at best, short phrases to an information retrieval system. Therefore, we are considering short queries directed at the titles of TREC-7, TREC-9 and Arabic TREC-10 topics to experiment with this situation.

4.1 Experimental Approach

In Arabic-English CLIR, presently, no benchmark data are available for Arabic-English CLIR. To provide a means to compare our efforts with future Arabic-English CLIR efforts, we used readily available English benchmark document collections and provide our Arabic queries, a translation of the National Institute of Science and Technology, Text Retrieval Conference (TREC) queries on our web site at www.ir.iit.edu. We used these 100 translated versions as our original Arabic queries issued against the TREC English collection.The Arabic queries were translated back to English using the ALKAFI MT system. Indexing is done using the Porter and K-stem algorithms after eliminating the stop-words. Similarly, querying is done after stemming and eliminating the stop-words of the translated target English queries. The ALKAFI Arabic-English MT system is a commercial system developed by CIMOS Corporation and it is the first Arabic to English machine translation system.

Usually, the Arabic text is not vocalized; so ALKAFI can add vowels internally. But sometimes, the user must vocalize some consonants to help ALKAFI at lexical and syntactic analysis. Vocalization is crucial step since word sense depends on vocalization and on word position in context. The system attempts to analyze words in context and then builds semantic relations. Then, the English text is generated by a transfer method according to English language grammar rules. ALKAFI uses five dictionaries:

The TREC queries (or topics in the TREC vernacular) consist of three fields: title, description, and narrative. The title is considered short; it consists of one, two or three concept terms. In Table 13, we illustrate an example of the original Arabic title and its translation. The description field is of medium length; it consists of one or two sentences. In Table 14, we provide an example of the description field and its translation. The longest part is the narrative field; in Table 15, we show an example of the narrative field and its translation using the ALKAFI MT system. To measure the effectiveness of an MT system for CLIR, we experimented using all three-query types to determine the effects of query length (short, medium, and long) on the performance of the MT-based method for CLIR.

Arabic query	Translated English Query
مُعَدَّات التشفير التصديريّة	The export equipments of the encryption

Table 13. The title of the original Arabic and the translated English query using the ALKAFI MT system

Arabic query	Translated English Query
عرِّف الوثائق التي تَناقش اهتمامات الولايات المتّحدة بشأن المصدَّر من مُعَدَّات التشفير.	Define the documentations which the debate of the United States concerns regarding exported from equipments of the encryption.

Table 14. The description of the original Arabic and the translated English query using the ALKAFI MT system

Arabic query	Translated English Query
الوثائق التي تذكر اسم الشركة أو المجموعة التي تُنتج مُعَدَّات التشفير فقط ، و لكن لا تذكر بالمصدَّر و/ أو الاستثمار التجاريّ لمُعَدَّات التشفير ليست ذات صلة. الوثائق التي تشير إلى الوصول الحكوميّ لنُظم التشفير لغايات ضدّ نشاطات الاستخبارات أو نشاطات مكافحة الجريمة ، تكون ذات صلة.	The documentations which she remembers the name of the company or the group which produces encryption equipments of the encryption only , but not you remember by exported and / or the commercial investment equipments of the encryption have no relevancy. The documentations which she points out the governmental arrival for organisms of the encryption the purposes of a briskness opposite have the secret services or the briskness of the crime struggle , you are a relevancy.

Table 15. The narrative of the original Arabic and the translated English query using the ALKAFI MT system

For English-Arabic CLIR, we conducted the experiments using Al-Mutarjim Al-Arabey English to Arabic commercial system (ATA Software Technology). The titles of the source Arabic queries are translated to English by the Al-Mutarjim Al-Arabey MT system. The average length of the titles of Arabic TREC topics is 6.2 words. The minimum speed of translation is 1000 words per minute on a system with just the basic hardware requirements. The translation result of query's title AR23 using Al-Mutarjim Al-Arabey system is shown in Table 16.

Original English query	**MT translation**
Information technology and the Arab World	تقنية المعلومات والعالم العربي

Table 16. English query terms and their translation using Al-Mutarjim Al-Arabey MT system

4.2 Results

We use three performance measures. The first uses the recall-precision scores at 11 standard points. In CLIR systems, given the expenses of the translation, a user is most likely to be interested in only the top few retrieved Web pages. Thus, we provide measures for the top *n* retrieved documents. We also provide the overall average of precision of each run. We evaluate the effects of the MT system in Arabic-English CLIR. As described earlier, we used both the TREC-7 and TREC-9 topics and TREC-9 collections. For TREC-7, as shown in Table 17, the machine translation achieved 61.8%, 64.7%, and 60.2% for title, description, and narrative fields, respectively. The 11-point average recall-precision for TREC-7 topics is shown in Figures 4, 5, and 6 for the title, description, and narrative fields, respectively. As shown, the MT-based approach on description is more effective than title and narrative. In each figure, we also illustrate the "ideal" system score, which is represented by the monolingual query. At the higher precision-lower recall levels, the difference is even more noticeable. The degraded effectiveness of the machine translation on title is that the ALKAFI machine translation system is designed to perform best on well-formed sentences or at least on a sequence of words that form a context. However, the titles of topics 351-400 are all three words or less; thus, no substantive context is formed.

For the narrative run results shown in Figure 6, the MT system is unable to preserve its accuracy when extra, potentially noise, terms are presented in the source query. The greater the number of source query terms, except for, of course, keywords or words of high query disambiguation content, the greater is the performance degradation of a CLIR system. These additional, potentially noise, terms do not provide a strong basis of the source query. The ALKAFI MT system, however, is still capable of maintaining 60.2% of the monolingual retrieval. At the higher precision-lower recall levels, the narrative run is more effective than the title. At the higher recall level (up to 0.8), the title run is more effective than the narrative run. As measured by average precision, there is a slight difference between the narrative and the title runs. It is not surprising that the narrative run is strictly worse in accuracy then the descriptive run since the MT system achieves its best performance on the fewest sequence of words that still provides a full context.

	Original	MT method	% Monolingual
Title	0.1733	0.1071	61.8
Description	0.1838	0.1190	64.7
Narrative	0.1522	0.0917	60.2

Table 17. Average precision of the title, description and narrative fields of queries topics 351-400

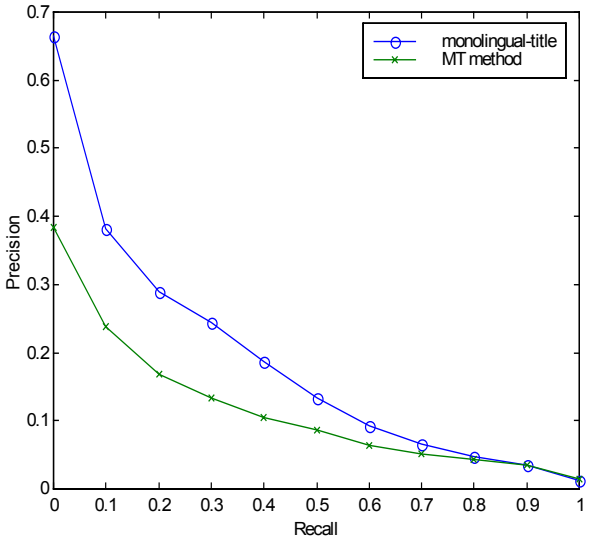

Figure 4. Average precision and recall of the titles of the original Arabic query of topics 351-400 and MT method

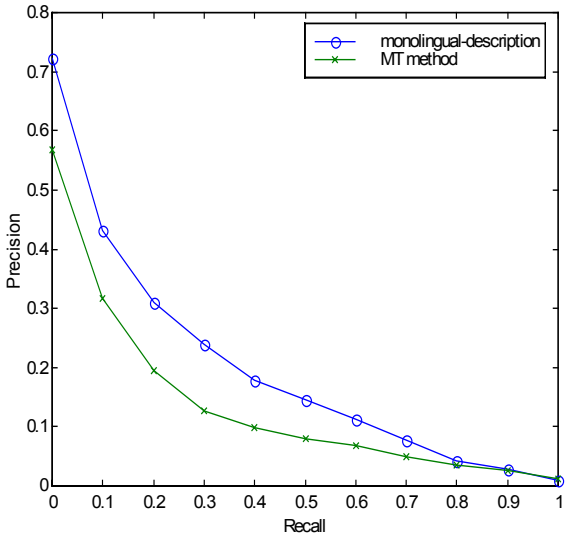

Figure 5. Average precision and recall of the descriptions of the original Arabic query of topics 351-400 and MT method

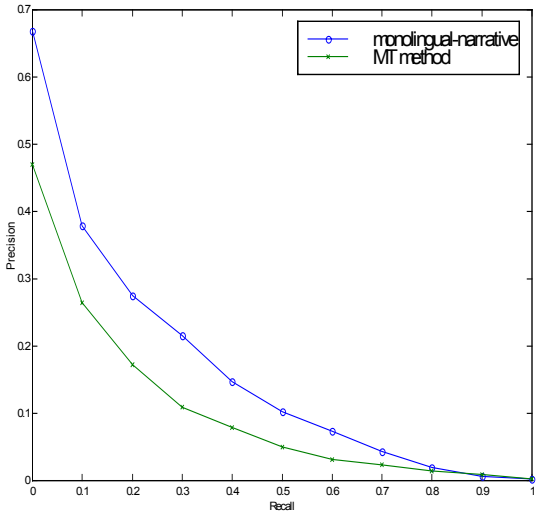

Figure 6. Average precision and recall of the narratives of the original Arabic query of topics 351-400 and MT method

In Table 18, we illustrate the results up to 1000 documents retrieved for TREC-7 queries

351-400. As shown, the description run consistently outperforms both the title and the narrative runs.

Precision	Original Title	MT Title	Original Description	MT description	Original Narrative	MT Narrative
at 5 Docs	0.4240	0.2200	0.4880	0.3560	0.4360	0.3000
at 10 Docs	0.3800	0.1960	0.4160	0.2920	0.3780	0.2620
at 15 Docs	0.3413	0.1907	0.3787	0.2573	0.3347	0.2413
at 20 Docs	0.3170	0.1940	0.3420	0.2340	0.3130	0.2180
at 30 Docs	0.2700	0.1667	0.3020	0.2040	0.2700	0.1793
at 100 Docs	0.1746	0.1162	0.1780	0.1206	0.1656	0.1066
at 200 Docs	0.1226	0.0825	0.1245	0.0833	0.1124	0.0742
at 500 Docs	0.0731	0.0497	0.0721	0.0494	0.0624	0.0432
at 1000 Docs	0.0459	0.0305	0.0455	0.0315	0.0389	0.0270

Table 18. Precision at 1000 retrieved documents of topics 351-400

In Table 19, we illustrate the average precision of TREC-9 topics. Our CLIR approach using the ALKAFI MT system achieves 58.4%, 57.1%, and 53.4% for title, description, and narrative fields, respectively. The 11-point average recall-precision for TREC-9 topics is shown in Figures 7, 8, and 9 for the title, description, and narrative fields, respectively. Again, the "ideal" monolingual run is likewise illustrated in each figure.

	Original	MT method	% Monolingual
Title	0.1305	0.0763	58.4
Description	0.1857	0.1061	57.1
Narrative	0.1678	0.0897	53.4

Table 19. Average precision of the title, description and narrative fields of queries topics 451-500

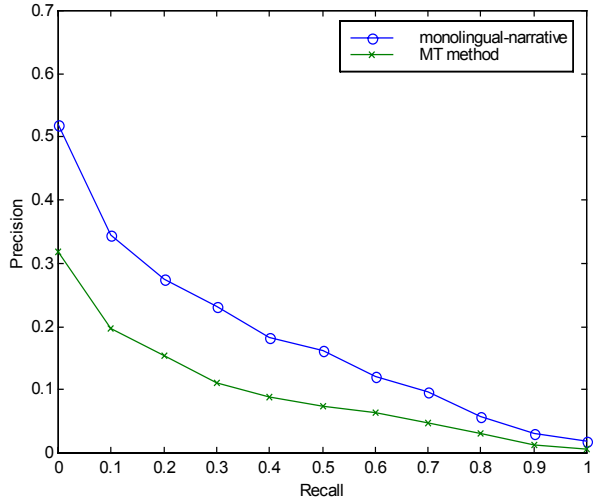

Figure 7. Average precision and recall of the titles of the original Arabic query of topics 451-500 and MT method

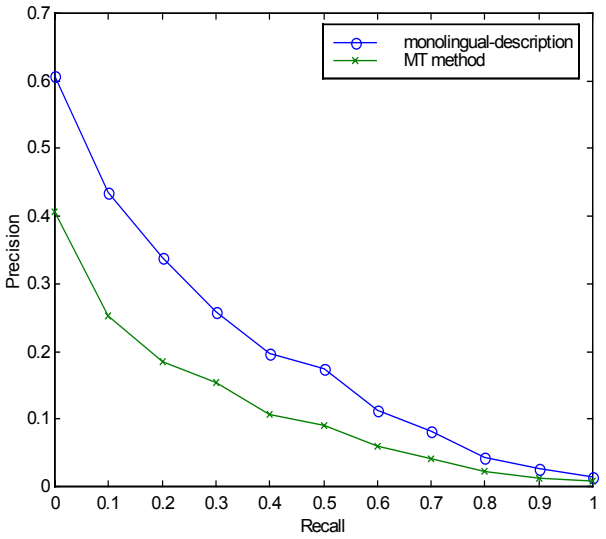

Figure 8. Average precision and recall of the descriptions of the original Arabic query of topics 451-500 and MT method

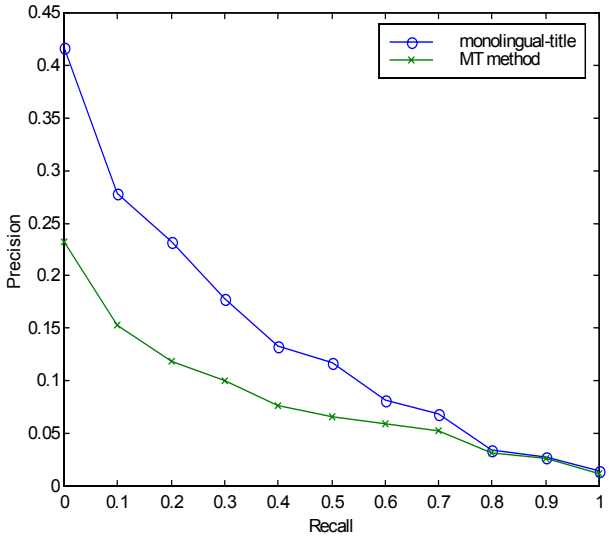

Figure 9. Average precision and recall of the narratives of the original Arabic query of topics 451-500 and MT method

Precision	Original Title	MT Title	Original Description	MT description	Original Narrative	MT Narrative
at 5 Docs	0.2227	0.1222	0.3560	0.2360	0.2600	0.1880
at 10 Docs	0.1886	0.1111	0.2740	0.1980	0.2460	0.1580
at 15 Docs	0.1712	0.1037	0.2627	0.1800	0.2200	0.1373
at 20 Docs	0.1545	0.0944	0.2330	0.1690	0.1990	0.1240
at 30 Docs	0.1348	0.0917	0.2167	0.1447	0.1700	0.1053
at 100 Docs	0.0834	0.0633	0.1260	0.0898	0.1064	0.0642
at 200 Docs	0.0581	0.0457	0.0894	0.0664	0.0713	0.0433
at 500 Docs	0.0316	0.0276	0.0514	0.0373	0.0384	0.0253
at 1000 Docs	0.0184	0.0165	0.0314	0.0235	0.0229	0.0156

Table 20. Precision at 1000 retrieved documents of topics 451-500

In Tables 20, we illustrate the results up to 1000 documents retrieved for TREC-9. As shown, again, the description run consistently outperforms both the title and narrative runs. However, as shown in Table 20, the percentage of degradation of the title run from the "ideal" monolingual title run is less that that of the descriptive run. This result is seemingly inconsistent with the results obtained for the machine translation on titles run for queries 351-500 as presented in Table 17. The reason behind this seeming contradiction in accuracy performance is that the titles of query 451-500 are actually quite long. The average title query length for queries 351-400 is 2.72 word per query while the average length for queries 451-500 is 3.46 words. This 27% difference in

query length was sufficient to provide our MT system with the possibility to form a proper context for many more queries in the TREC-9 query set as compared to the TREC-7 query set. This is especially so considering that for the TREC-9 query set had 16 queries with 4 or more words as compared to the only 6 queries of similar length in the TREC-7 query set. For example, the title of the query number 482 is:

أين يُمكِن أن أجِدْ معدّلات النّموّ لشجرة الصنوبر ؟

The translated query using ALKAFI MT system is:

"Where is he possible that I find the rates of the growth for the tree of the pine? "
This query provides a full context to make the ALKAFI machine translation produces the most accurate translation. Adding more contexts to that query does not help the MT system to provide better translation accuracy.

Finally, for completeness, we provide a brief overview of efficiency results. In Table 21, we summarize the efficiency (run time performance) of the ALKAFI MT system to translate the titles, descriptions and narratives fields of topics TREC-7 and TREC-9.

	Title	Description	Narrative
TREC-7	6	18	51
TREC-9	7	17	40

Table 21. The total running time of the MT system measured in seconds

In Table 22, we summarize the efficiency (running time performance) of AIRE search engine to run the translated titles, descriptions and narratives fields of topics TREC-7 and TREC-9.

	Title	Description	Narrative
TREC-7	445.65	1972.25	6143.79
TREC-9	3752.00	12630.42	28708.65

Table 22. The total running time of the queries measured in seconds

The narrative fields as described in Tables 18 and 20, which represent the long queries, are not effective compared to the description fields, which represents the medium length queries. According to theses findings, the fewer terms provided in the original query that form a context to obtain unambiguous representation, the better running time as well as the better retrieval effectiveness. As presented in Tables 21 and 22, the total running time for the description and narrative runs of TREC-7 is 6194.79 and 1990.25 seconds, respectively. The running time of the narrative is 211% of the running time of the description. In fact, the difference of the running time degrades the performance of our CLIR system without any improvement on the effectiveness. These findings are consistent with TREC-9 topics and collection as presented in Tables 21 and 22.

The description runs perform 340% much more time compared to title runs of TREC-7 dataset. Accordingly, the achieved performance of the description run is more effective than the title run. Thus, choosing few terms that form a full context achieves better accuracy at the expense of efficiency, a trade-off whose merits are application dependant. Similar findings exist for the TREC-9 queries.

As shown in Table 23, the MT system achieved 70.2% of the monolingual retrieval. The MT system is capable to preserve its accuracy since most of the titles of the Arabic topics are quite long to form a context.

	Average Precision	%Monolingual
Monolingual	0.3709	
MT	0.2605	70.2

Table 23. Average precision of original and cross-language runs using MT approach

The queries were expanded using post-translation technique. The 15 terms from the top 10 documents are added to the translated Arabic terms.

	MT	MT+Post
Average Precision	0.2605	0.3012
%Improvement		15.6
Statistical test		P=0.02

Table 24. Average precision of the MT approach before and after query expansion via PRF

The post-translation expansion technique improved the performance by 15.6%, the difference between the MT and MT+post is statistically significant at 98% confidence level. Table 25, describes the runs at lower level of recalls, up to 1000 retrieved documents. As sown, the MT with query expansion after translation consistently outperforms the MT approach without query expansion.

Precision	MT	MT+Post
at 5 Docs	0.5040	0.5280
at 10 Docs	0.4840	0.5120
at 15 Docs	0.4293	0.4933
at 20 Docs	0.4020	0.4780
at 30 Docs	0.3880	0.4280
at 100 Docs	0.2992	0.3244
at 200 Docs	0.2354	0.2596
at 500 Docs	0.1505	0.1609
at 1000 Docs	0.0947	0.0984

Table 25. Precision at 1000 documents retrieved of MT and MT+post runs

5. Conclusions

Our results demonstrate the potential Arabic-English and English-Arabic CLIR. Automatic dictionary translation is 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. The factor affecting this is the transfer of too many senses that are inappropriate to the source 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 for Arabic-English and English-Arabic. For Arabic-English CLIR, this approach achieved 68.9% and 64.7% of the titles of English only 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 for Arabic-English CLIR; it is called the Two-Phase method.

In the Two-Phase method, we ignore all the terms that do not retranslate to the original Arabic query term. 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. The Two-Phase method yields a 38% and 52% improvement over the Every-Match (EM) method of TREC topics 351-400 and TREC topics 451-500, respectively. It also yields a 4% and 7% improvement over the First-Match (FM) method 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% confidence interval over the EM for both TREC-7 and TREC-9. It achieved 86% and 89% over FM method for TREC-7 and TREC-9, respectively. In this study, we showed that eliminating unrelated terms by the Two-Phase method can significantly reduce the ambiguity associated with dictionary translation. 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.

We also evaluated the MT-based Arabic-English CLIR; we found that the query length affects the performance of the MT system. The evaluation was conducted by using the ALKAFI system and two standard TREC collections and topics. To explore the effects of the context to the quality of translation, we experimented with various query lengths.

We studied the effects of using Al-Mutarjim Al-Arabey MT system and MRD for English-Arabic CLIR. The post-translation approach was used. We found that the query expansion after translation via PRF is consistently more effective for both MT and MRD approaches.

The experimental results indicate that the less source terms that are needed to form a context, the better is the retrieval accuracy and efficiency. However, the problem of semantics is perennial due to the complexities of the Arabic grammar. Without some level of semantic representation, MT systems are unable to achieve high quality translation, because they cannot differentiate between cases that are lexically and syntactically ambiguous. Accordingly, a well-formed source query makes the MT system able to provide its best accuracy.

A possible extension to our work is to expand the original source query using PRF for Arabic-English CLIR to emphasize the context of the source query and finding term

threshold for the TP method. Another extension is to apply the Two-Phase method by using parallel corpus or a combination of MRD and parallel corpus.

6. References

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

aDawliah Universal Electronics (1999), http://www.adawliah.com.sa/.

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

Aljlayl, M., and Frieder, O. (2001).Effective Arabic-English Cross-Language Information Retrieval via Machine Readable Dictionaries and Machine Translation. *ACM CIKM*, pp. 295-302.

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

ATA Software Technology Ltd., http://www.atasoft.com

Ballesteros, L., and Croft, B. (1996). Dictionary Methods for Cross-Lingual Information Retrieval. In the *Proceedings of the 7th International DEXA Conference on Database and Expert Systems Applications*. pp. 791-801.

Ballesteros, L., and Croft, B. (1997). Phrasal Translation and Query Expansion Techniques for Cross-language Information Retrieval. *SIGIR 1997*, pp. 84-91.

Ballesteros, L., and Croft, B.(1998). Resolving Ambiguity for Cross-Language Retrieval. *SIGIR*. pp. 64-71

Braschler, M., Peters, C. and Schuable, P (1999). Cross-Language Information Retrieval (CLIR) Track Overview, *TREC-8 Proceedings*.

Chowdhury, A., Beitzel, S. ,Jensen, E. , Sai-lee, M. , Grossman, D. , Frieder, O. , McCabe C. , Holmes, D. (2000)."IIT TREC-9 - Entity Based Feedback with Fusion", Proceedings of TREC-9, NIST, pp. 241-248.

Davis, M., and Dunning, T. (1995). Query Translation using Evolutionary Programming for Multilingual Information Retrieval. In *Proceeding of the Fourth Annual Conference on Evolutionary Programming*.

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

Egyptian Demographic Center, (2000).
http://www.frcu.eun.eg/www/homepage/cdc/cdc.htm.

Hasnah, A. (1996). Full Text Processing and Retrieval: Weight Ranking, Text Structuring, and Passage Retrieval for Arabic Documents. Ph. D. Dissertation, Computer Science Department, Illinois Institute of Technology, Chicago, IL.

Hull, D. and Grefenstette, G. (1996). Querying Across Languages. A Dictionary-based Approach to Multilingual Information Retrieval. In *proceedings of SIGIR*, pp. 49-57.

Jones, G., Sakai, T., Collier, N., Kumano, K., Sumita, K.(1999). A Comparison of Query Translation Methods for English-Japanese Cross-Language Information Retrieval. *SIGIR,* pp. 269-270.

Kwok, K.L. (1999). 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*.

Landauer, T. K., and Littman, M. L. (1990). Full Automatic Cross-Language Document Retrieval using Latent Semantic Indexing. In *Proceedings of the 6th Conference of UW center for New OED and Text Research*, pp. 31-38.

Oard D. (1998). A Comparative Study of Query and Document Translation for Cross-Language Information Retrieval. In *Machine Translation and the Information Soup. Third Conference of the Association for Machine Translation in the Americas*, pp. 472-83.

Pirkola, A. (1998). The Effects of Query Structure and Dictionary Setups in a Dictionary-Based Cross-Language Information Retrieval. *ACM SIGIR*. pp. 55-63.

Radwan, K., Fluhr, C. (1995). Textual Database Lexicon used as a Filter to Resolve Semantic Ambiguity Application on Multilingual Information Retrieval. In *Fourth Annual Symposium on Document Analysis and Information Retrieval*, pp. 121-136.

Sheridan, P. and Ballerini, J.P.(1996). Experiments in Multilingual Information Retrieval using the SPIDER System. In *Proceedings of ACM SIGIR*. pp. 58-65.

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

TREC, (2001). http://trec.nist.gov/act_part/tracks.html

Wonnacott, R., Wonnacott, T. (1990). Introductory Statistics, John Wiley & Sons, Fourth Edition.

Xu, J. and Croft, W. B. (1996). Query Expansion using Local and Global Document Analysis. In *Proceedings of ACM SIGIR*, pp. 4-11.