What's in a Name?: Proper Names in Arabic Cross Language Information Retrieval

Leah S. Larkey, Nasreen AbdulJaleel, Margaret Connell
Center for Intelligent Information Retrieval
Department of Computer Science
University of Massachusetts
Amherst, MA 01003

{larkey, nasreen, connell}@cs.umass.edu

ABSTRACT

Proper names are problematic for cross language information retrieval. Standard bilingual dictionaries typically have poor coverage of proper names. On the other hand, IR tasks involving news corpora, like TDT and TREC cross language IR, have proper names at their core. In this study, we demonstrate the importance of proper names in one such task, the TREC 2002 (Arabic-English) cross language track, by showing that performance degrades a tremendous amount when the bilingual lexicons do not have proper names. We then examine several different sources of proper name translations from English to Arabic, both static and generative (transliteration) and explore their effectiveness in the context of the TREC 2002 cross language IR task. We support a conclusion that a combination of static translation resources plus transliteration provides a successful solution.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *Indexing methods, Linguistic processing*.

General Terms: Algorithms, Experimentation, Performance.

Keywords: Cross language information retrieval, CLIR, crosslingual, Arabic, transliteration, proper names.

1. INTRODUCTION

Cross language IR requires text resources that define the correspondence between words in the two languages. The primary resource for almost all approaches is a bilingual lexicon or dictionary. Bilingual dictionaries have been available for a long time for many language pairs, and with conversion to machine-readable form, these can be used for IR.

Machine readable dictionaries are not available for all languages pairs, and those that are available are often of limited value for CLIR. Their coverage can be limited in that they often do not contain names, numbers, technical terms, and acronyms. Their translations may not reflect current usage. In addition, they do not provide a way to directly relate inflected forms in the two languages, so that stemming or more sophisticated morphological analysis can be required, particularly for highly inflected languages. The morphological problem is particularly acute for Arabic, where traditional dictionaries are arranged by roots, so one must do morphological analysis to look up an Arabic word.

CIIR Technical Report IR-278

Bilingual dictionaries can be induced from parallel corpora – collections of the same documents in two languages. Given large enough collections of parallel corpora, dictionaries constructed this way can be more effective than manually constructed dictionaries because they cover words which are not typically found in bilingual lexicons, like proper names, numbers, technical terms, and acronyms. They include inflected forms of words. In addition, they include translation probabilities, required by many translation models. However, parallel corpora are not always available. Even when available, they are never complete.

We assume a general context in which one may have a traditional bilingual dictionary in machine readable form, and in which a parallel corpus may or may not be available. Other specialized bilingual lexicons may be available. There will always be out-of-vocabulary words, that is, words not covered by these *static* resources. Thus, there is always a need for a component that can generate translations for unknown words, or estimate translation probabilities for pairs of strings in the two languages.

We are interested here in getting an overview of the role of various kinds of translation resources, both static and generative, in a typical cross language IR task. For convenience, we focus on proper nouns in this research, and experiment with a set of names involved in TREC queries for our experiments. Proper nouns are an important cause of out-of-vocabulary (*OOV*) errors in IR. By proper nouns, we mean names of people, places, and organizations, including acronyms –nouns that are typically capitalized in English. We refer to these as *names* in this paper, but it should be understood that this usage is not restricted to names of people.

Another problem with names in Arabic-English information retrieval is great variability in spelling. Whitaker [24], for example, identifies 32 different English spellings for the name of the Libyan leader Muammar Gaddafi, and four spellings for Al-Qaeda.

Simple IR systems do not know which words are names. We make the assumption that the facts about name translation also apply to the translation of other unknown words, particularly technical terms, and that solutions to the problem of name translation would be effective for translating other unknown words.

In the present study, we demonstrate the importance of proper names in the TREC 2002 (Arabic-English) cross language track by measuring how much performance degrades using bilingual lexicons that lack proper names. We then examine several different sources of proper name translations from English to Arabic, both static and generative, including our own transliteration system. We explore the effectiveness of each of these sources on the information retrieval task. Finally, we demonstrate that a combination of static resources and transliteration of unknown names is an effective approach.

2. PREVIOUS RESEACH

TREC has had a cross language track since 1997. TREC-6 had documents in English, French, and German, and queries in English, French, German, Spanish and Dutch [21]. It was noted that proper names accounted for large proportion, 49%, of a sample of corpus words that had no translations in a large English-Spanish machine readable dictionary [8]. Many TREC participants added untranslated words to the translated query without modification. This worked for Spanish, and for many western European languages, because names are often rendered the same way in these different languages.

However, many other words are rendered similarly, but not identically in different languages, particularly when the languages are less closely related. Language pairs like English and Finnish contain examples like *European* and *Euroopan*, *pharmacology* and *farmakologian*, or *calcitonin* and *kalsitoniini*. Researchers participating in CLEF [15][16], have applied and developed ways to relate such pairs of words, using approximate string-matching techniques developed for monolingual name matching and for error correction, such as Soundex, Phonix, the Damerau-Levenstein metric, and n-grams. Pfeifer et al [17] showed n-grams to be the most effective of these. Pirkola, et al [20] have generalized n-grams to s-grams, or skip-grams, which considers nonadjacent pairs of characters.

Language pairs with different orthographies introduce an additional complication to this problem. It is not sufficient to use the untranslated query term in these cases, because the untranslated term will match nothing in the collection. Approximate string matching is also inadequate when the strings are from different alphabets. The query term must be somehow rendered in the orthography of the other language. The process of converting a word from one orthography into another is called *transliteration*.

There are two senses in which the term transliteration can be used. The first is a one-for-one character mapping, in which each character in one alphabet is invariably substituted with a specific character in the other alphabet. The Buckwalter transliteration [6], is one such system, which renders Arabic words in Roman characters. Its usage can be seen used in the pronunciation guide in Table 4. Although it is simple and deterministic, it produces a pseudo-phonetic notation rather than a likely English or European spelling. For example, the Arabic character "shin" (ش) is represented by the dollar sign "$".

This one-for-one substitution is different from the sense of transliteration in the present work, in which Arabic words are rendered in Roman characters, or European words in Arabic characters, in a way that will cause a reader to produce an approximately correct pronunciation of the name. The goal is to produce "correct" Arabic spelling(s) as they would be found in published text, and also handle the variability addressed by the approximate string-matching techniques described above. This kind of transliteration can be viewed as an approximate string matching technique for the case where the two strings are from different alphabets.

Although many groups have developed English/Arabic transliteration systems, little is published about them, and no work evaluates their effectiveness for IR. Online machine translation engines include transliteration for unknown words [1][23], but no information is available about how this is done, or how well it works. Darwish et al. [7] described a transliterator used for TREC-2001, but provided no evaluation of its effectiveness.

Stalls and Knight [22] and Al-Onaizan and Knight [3] have also produced transliterators for Arabic/English which receive more extensive description and evaluation. Their work includes evaluation of how well the transliterators can match a source spelling, and Al-Onaizan evaluates the transliterations in terms of their reasonableness according to human judges, but no study measures their performance on a retrieval task or on other NLP tasks.

3. EXPERIMENTAL METHOD

We have chosen to assess the quality of different sources of query names via a retrieval task. This may seem indirect, compared to counting the number of correct translations found in each of the sources, but "number correctly translated" would not be as good a measure for several reasons: the notion of a "correct translation" is vague for names in Arabic – we care more about whether a translation matches the rendering found in the corpus being searched. Many reasonable and correct name transliterations are useless because a different form is found in a corpus of interest. Conversely, technically incorrect transliterations can suffice, provided that after normalization and stemming they match the forms found in the normalized and stemmed corpus.

All the experiments carried out here involve the same set of English queries and the same query expansion, translation, and retrieval method. The only difference among different experimental conditions is in what dictionaries are used in query translation.

3.1 IR Experiments

All experiments were carried out using the TREC 2001 collection of 383,872 Arabic newspaper articles from Agence France Presse (AFP), and the 50 TREC 2002 topics, 26-75. Queries were formed from the title and description fields of the English versions of the topics.

Although we used both INQUERY and Language Modeling approaches in our official TREC submissions, we used INQUERY here to expedite the large number of experimental conditions to be run. In general, we have found that we obtain comparable performance with these two approaches, and gain some effectiveness by combining them [ref to be added later for anonymity].

Arabic articles in the collection were converted from Unicode UTF-8 encoding to Windows Arabic (CP1256). Simple tokenization broke up text into words at white space or punctuation characters, which included 5 Arabic punctuation characters not found in English text. Arabic text was stemmed using the light stemmer, described in [ref to be added later]. Stop words from the INQUERY stop word list were removed. Tokens of more than one character in length were indexed.

For cross language querying of the Arabic collection, we used structural query translation [5], sometimes called the Pirkola method [18], a dictionary-based query translation method in which multiple translations of a term are wrapped in an INQUERY #synonym operator. This has the effect of treating the set of translations as a single term in retrieval, whose term frequency is the sum of frequencies of all the different translations, and whose document frequency is the number of documents in the union of the sets of documents containing each translation.

Retrieval experiments contained the follow steps:

- English queries were tokenized, lower cased, and stop words were removed.
- English queries were expanded as follows. Using the English query, the top 10 documents were retrieved (via INQUERY) from a collection of AP news articles from 1994 through 1998 in the Linguistic Data Consortium's NA News corpus. This corpus was indexed without stemming, but normalized to lower case. The top 10 documents received an expansion score which was the sum across the ten documents of the INQUERY belief score for the term in the document. The 5 new terms with the highest expansion score were added to the query. Final weights for all terms were set to $2w_o + w_e$ where w_o is the original weight in the unexpanded query and w_e=1.
- Each English word was looked up in the bilingual lexicon appropriate to the experimental condition. All the alternative translations were placed inside a #synonym operator, then the translations or #syn's for all the query words were gathered under a weighted sum operator. The weight on each translation (term or #syn) was the weight of the English source word in the expanded English query.
- The structured Arabic queries were expanded much like the English queries. Using the unexpanded query, 10 documents were retrieved from the AFP corpus. Terms from these documents were ranked and the top 50 terms were added to the query, receiving weights as in English query expansion.
- The expanded Arabic query was submitted to the AFP collection.

English query expansion added many proper names to the queries. Our experiments focus on the names found in the original and expanded English queries, which number 241. We refer to this set as the *query names*.

3.2 Baseline Dictionaries

Most of the work reported here begins with the English-Arabic bilingual lexicon used in our previous work (TREC 2001, TREC 2002), which we call the *Inhouse* dictionary. Most of the words were obtained by querying an online bilingual dictionary via a cgi script that requested an English translation for each Arabic word in the AFP corpus. This dictionary contains 775,000 translations for 50,000 English words. In addition, it includes proper names derived from the static, AlMisbar, NMSU, and Tarjim sources described below. The *Inhouse* dictionary does not include any transliterations from our transliteration system, or translations from the United Nations parallel corpus, also described below.

A second bilingual lexicon, *UNdict*, is derived from the United Nations parallel English-Arabic corpus. The Arabic was stemmed using our light stemmer, and the English was normalized to lower case. GIZA++ software [11] was used to train a statistical translation model from this corpus, yielding a bilingual corpus with translation probabilities *P(a|e)*, the probability of an Arabic word ***a*** given the English word ***e***. We retained translations with a probability of .15 or higher. It should be noted that this is related to, but not the same as, the "standard parallel corpus" dictionary distributed for TREC-2002 [10]. Our dictionary is trained on the same United Nations corpus, using the same GIZA++ parameters. It differed in that we used a different Arabic stemmer and we did not stem the English text before training. English text was normalized to lower case. The *UNdict* dictionary is used both to replicate some of the results found with the *Inhouse* dictionary, and as a source of name translations for the experiments in Section 4.3

3.3 Sources of Proper Names

Table 1 provides an overview of the sources of proper names compared in the present experiments.

Table 1: Sources of names in experiments

Source	Number of Query Names having translations	Avg. Number of Translations per Query Name
Static	78	4.3
NMSU	94	1.1
AlMisbar	241	1
Tarjim	241	1
UN	190	1.46
Translit	241	various

Static refers to the subset of query names and translations that already existed in the dictionary that we used for TREC-2001 before we added translations for TREC-2002 query words by consulting online machine translation engines. It consisted primarily of a small bilingual lexicon of country and city names that derived from a list of world cities found on the web [25]. This list had 489 entries, and listed the names of most countries of the world, their capitals, and a few other major cities. To get the Arabic translations, we used the Sakhr SET engine, an earlier version of Tarjim [23], which performed machine translation from English to Arabic. This list of place names (and only this list, which was made independently of the queries) was hand corrected by an Arabic speaking consultant.

NMSU is a parallel list of 148,599 English and Arabic proper nouns obtained from the web page of the CRL at New Mexico State University [14]. It covered only 94 of the 241 query names.

AlMisbar [2] and *Tarjim* [23] are both web sites which provide English-Arabic online machine translation. We submitted the list of 241 query names to the machine translation engine at each of these sites. Each includes a transliteration component which generates a translation for words that are not found in its dictionary. Each returned exactly one Arabic translation for each English name submitted.

UN refers to query name translations found in the *UNdict*.

Translit refers to a statistical model for English to Arabic transliteration under development at our lab. It is one of class of extremely simple models in which a transliteration is assigned a score that estimates the probability that an Arabic word *A* is a correct transliteration of an English word *E*, *P(A|E)*, as a function of translation probabilities of segments making up the word:

$$P(A \mid E) = \prod_i a_i \mid e_i$$

where the a_i are segments making up the Arabic word and e_i are segments making up the English word. The segments include a combination of unigrams and commonly occurring n-grams. On the English side, the alphabet consists of the 26 letters plus beginning and ending markers. On the Arabic side, the alphabet also consisted (coincidentally) of 26 Arabic characters plus beginning and ending markers. English words were normalized to lower case, and Arabic words were normalized by removing diacritics, replacing أ, إ, and آ with bare alif ا, replacing final ى with ي, and replacing final ة with ه.

The transliterator is part of a larger project exploring how best to design transliteration models that can be trained automatically, and is the subject of a forthcoming report. The particular instantiation of the model used here is hand-crafted for the purpose of providing a high-quality benchmark against which to measure the performance of the automatically trained transliteration systems.

The transliterator proposes a ranked list of transliterations for each English word. We can truncate these lists to any desired length to get a specified number of transliterations for an English word.

4. RESULTS

4.1 Importance of Proper Names

The first set of experiments demonstrates the importance of proper names in this set of TREC queries. We tested the 50 TREC 2002 queries using the *Inhouse* dictionary and using a dictionary from which the 241 query names have been removed. The same experiment was conducted using the *UNdict* dictionary, and the *UNdict* dictionary after removing translations for the 241 query names. Table 2 shows the mean average precision for each of these four dictionaries.

Table 2: Effect of removing query names from bilingual lexicons

Dictionary	Full Dictionary	No Names
Inhouse	.3330	.1433 (-57.0)
UNdict	.3161	.1458 (-56.4)

Although this experiment is somewhat simple-minded, it is nevertheless useful. We expected performance to be seriously degraded when the dictionaries did not contain the query names, but we did not expect the degradation to be so large. Performance is reduced more than 50% when the proper names are missing. The performance of the *Noname* dictionary – the *Inhouse* dictionary minus query names – provides a useful baseline for subsequent experiments.

4.2 Transliteration of Proper Names

In this section we look at the effectiveness of transliteration for translating unknown words, and we explore how many alternatives should be included from the transliterator. It is clear that more than one alternative spelling should be generated, because multiple spellings can be found for the same name in Arabic text. With more alternative spellings generated, the output is more likely to include the spelling(s) that are found in the corpus. At the same time, the larger set of alternative spellings is more likely to cause false hits, matching the wrong words in the corpus.

The baseline dictionary for this experiment is *Noname*, which is the *Inhouse* dictionary after all the translations for the 241 query names have been removed. Additional dictionaries were made by adding transliterations for the query names to *Noname*. For each of the 241 query names, the transliterator generated all the possible transliterations and ranked them. Dictionaries *Translit1*, *Translit5*, *Translit10*, *Translit 20*, and *Translit30* were made by adding the 1, 5, 10, 20, or 30 top-ranked transliterations to *Noname* dictionary. Table 3 shows mean average precision on the IR task using each of these dictionaries. Note that it is not possible to generate 20 or 30 distinct transliterations for all words. The column labeled *Raw* indicates the average number of transliterations actually generated per word in the set. The results show that performance improves with up to twenty alternatives and then levels off.

Table 3: Effectiveness of different numbers of transliterations

Dictionary	Number of Translations per query name			Mean average precision
	Raw	After Stem	In corpus	
Noname	0	0	0	.1433
Translit1	1	1	.6	.1934
Translit5	4.9	4.3	2.0	.2459
Translit10	9.4	8.2	3.1	.2779
Translit20	17.0	14.5	4.6	.3026
Translit30	23.3	19.5	5.4	.3018

It might seem that a danger of generating 20 variants spellings for a name would be false hits: that many of the spellings would match unrelated Arabic words in the corpus. Table 3 suggests that spurious matches with bad transliterations are probably not a big problem, even when generating such large numbers of alternatives. First, after stemming some of the alternatives are identical, so there are fewer overall alternatives, as seen in the column labeled *After Stem*. Although it is beyond the skills of the researchers to examine all the occurrences of these transliterations in corpus documents and determine which are translations of the

English term and which are false hits, we can easily see in the column labeled *In corpus* that most of the transliteration alternatives do not occur at all in the corpus, and consequently cannot cause false hits.

Based on these results, 20 transliterations are used in all subsequent transliteration conditions.

4.3 Comparison of Individual Name Sources

Before presenting retrieval results based on the different name sources, it is informative to look at an example that illustrates how much variation is found in the Arabic rendering of an English name. Table 4 lists 6 distinct Arabic spellings of the name *Clinton* found in our set of bilingual English-Arabic resources. Each row contains a different Arabic spelling for the word and each column indicates a source of translations. An *x* in a cell means that the indicated source included that spelling in its set of translations.

Table 4: Arabic spellings of *Clinton* from different sources

Arabic Spelling	**Pronun-ciation Guide**	AFP	NMSU	Tarjim	Al Misbar	UN	Translit 1	Translit 5
كلينتون	klyntwn	x		x		x	x	x
كلينتن	klyntn	x*						x
كلينطون	klynTwn		x					x
كلنتن	klntn				x			
كلنتون	klntwn					x		x
كلاينتون	klAyntwn							x

* This spelling was rare, found in only 6 AFP documents

Five of the six spellings (all except the last) are reasonable, and consistent with the way many other English names are rendered in Arabic. However, only the first is useful for retrieval from the AFP collection. This example is particularly striking because one might expect the spelling of such a widely used name to be fairly standardized.

The retrieval experiment in this section compares dictionaries made by adding query name translations from individual sources to the *Noname* dictionary. Thus, the *Static* dictionary contains static translations for query names added to the *Noname* dictionary, this *UN* dictionary contains the query name translations from *UNdict* added to the *Noname* dictionary, *Translit20* contains the top 20 transliterations of the query names added to the *Noname* dictionary, and so forth.

Table 5 shows the retrieval performance based on the different sources, for expanded queries consisting of title and description. The column labeled *Coverage* indicates how many of the query names have at least one translation in the indicated dictionary.

Each of the query name sources provides a substantial and significant improvement in performance over the *Noname* dictionary. In general, resources with translations for more query names support more effective retrieval, but some resources appear to be of higher quality than others.

Table 5: Retrieval effectiveness and coverage of different query name sources

Dictionary	**Coverage**	**Mean average precision**	**% above Noname**
Noname	0	.1433	
Static	78	.2640	(+84%)
NMSU	94	.2337	(+63%)
Almisbar	241	.3033	(+112%)
Tarjim	241	.3187	(+122%)
Translit20	241	.3026	(+111%)
UN	190	.3276	(+128%)
Inhouse	241	.3330	(+132%)

The Almisbar and Tarjim sources appear to be very high quality but their effect is difficult to interpret because their translations are actually a mixture of manual translations and transliterations. Presumably the static translations were entered manually, and appeared to be accurate. On the other hand, the transliterations generated by these two systems were not very useful because they produced only one transliteration each.

We performed a small direct test of transliteration accuracy to compare the two online systems with our own. We started with a test set of 450 English names we had been using to compare different versions of our own transliterator. For each of these names, we had found the exact Arabic spelling used in the AFP corpus, considered correct for the purposes of this experiment. We looked up all these words in the AlMisbar system with their transliteration turned off, and found that 132 of the words did not have translations, and used these to test their transliteration against ours. We were not able to distinguish transliterations from other translations using Tarjim, but we tested it against the same set of 132 names anyway. The results can be seen in Table 6.

Table 6: Accuracy of different sources of transliterations

Number of Alternatives	Source of Transliterations		
	Translit	AlMisbar	Tarjim
1	37%	35%	30%
5	65%		
10	78%		
20	80%		
30	80%		

The table shows a large increase in accuracy as more alternative transliterations are added, and the three systems are comparable when only one transliteration is produced.

The next set of experiments attempts to look at the retrieval effects of manual translations and transliterations separately.

4.4 A Reasonable Combination

In this section we investigate the performance of a systematic combination of static and generative resources in a manner that is applicable to an operational CLIR system. The baseline here is the *Static* dictionary, which contains all the words collected for TREC prior to receiving the 2002 queries. *Static* excludes any query words looked up in online resources specifically for TREC 2002. Static excludes Recall that the online resources (Tarjim and Almisbar) contain some translations that are transliterations. Because these are excluded, the static dictionary does not have any translations that were generated by transliteration. The static dictionary does not have particularly good coverage of query names, or query words.

To the static dictionary were added one of two sets of transliterations: *Translit20,* the top 20 transliterations our system generates for all the query names, and *Translit_unk20*, the top transliterations for only the unknown query names, that is, the names that do not already have at least one translation in the static dictionary.

Table 7 shows the performance of these two systems compared to the static dictionary alone.

Table 7: Retrieval effectiveness of static dictionary plus transliterations of all names or of unknown names

Dictionary	Mean avg. Precision	% above Static
Static	.2640	
Static+translit20	.3265	(+24%)
Static+translit_unk20	.3277	(+24%)

These values confirm that it is a reasonable strategy in CLIR system to transliterate names, or to transliterate only the unknown names. These data do not suggest either alternative as a better choice.

We performed a similar experiment with a different static baseline. A new, larger baseline dictionary, *Staticbig,* was built, containing only non-transliteration translations from Tarjim and Almisbar.

To the *Staticbig* dictionary we added either *Translit20,* the top 20 transliterations our system generates for the all the query names, or *Translit_unk20,* the top transliterations for only the unknown names. Note that the set of unknown names for *Staticbig* is smaller than the set of unknown names for *Static*. *Staticbig* covers 165 of the query names, compared to *Static*'s coverage of only 78 names.

Table 8 shows the results of the experiment with the larger static dictionary as a baseline. The results are similar to those using *Static* as a baseline. Although it appears that translating only unknown names is more effective than translating all the names, the difference is not statistically significant.

Table 8: Retrieval effectiveness of larger static dictionary *Staticbig* plus transliterations of all names or of unknown names

Dictionary	Mean avg. Precision	% above Staticbig
Staticbig	.3315	
Staticbig+translit20	.3421	(+3%)
Staticbig+translit_unk20	.3519	(+6%)
Inhouse	.3330	0

5. DISCUSSION

The goal of these experiments was to gain some understanding of the relative effectiveness for information retrieval of different sources of name translations available for Arabic, with the idea that a similar situation would exist for other languages. The results support the following generalizations:

- Proper names are an extremely important component in cross language IR. Mean average precision degrades more than 50% using the typical bilingual dictionary that does not include proper names. Perhaps not all cross language tasks rely so heavily upon proper names, but the TREC 2002 cross language task is not atypical. We assert that any task involving searching, tracking, or extracting information from news items would share this reliance.
- Sources of proper names vary in quality, at least for a language pair like English and Arabic, in which there is tremendous variation, both in how English (and other western European) names are rendered in Arabic, and in how Arabic names are rendered in English. One cannot assume a given source of names translations will have useful spellings. Coverage can be poor, and the translations may not match the spellings in the corpora being searched.
- A good strategy is to use transliteration for proper names, or to use transliteration for unknown proper names. In future, we would like to investigate the related question of whether it is beneficial to generate transliterations for any unknown words, whether or not they are names.
- It is better to generate multiple alternative transliterations for unknown words rather than one. It is safe to include up to 20 alternative spellings for the unknown query words.

6. ACKNOWLEDGMENTS

This work was supported in part by the Center for Intelligent Information Retrieval and in part by SPAWARSYSCEN-SD grant numbers N66001-99-1-8912 and N66001-02-1-8903.

Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsor.

7. REFERENCES

[1] AbdulJaleel, N. and Larkey, L. S. English to Arabic Transliteration for Information Retrieval: A Statistical Approach," CIIR Technical Report IR-261.

[2] AlMisbar online translation. http://www.almisbar.com/salam_trans.html

[3] Al-Onaizan, Y. and Knight, K. Machine transliteration of names in Arabic text. Proceedings of the ACL conference workshop on computational approaches to Semitic languages, 2002 .

[4] Arbabi, M., Fischthal, S. M., Cheng, V. C., and Bar, E. Algorithms for Arabic name transliteration. IBM Journal of research and Development, 38 (2), 183-193, 1994.

[5] Ballesteros, L. and Croft, W.B. Resolving ambiguity for cross-language retrieval. SIGIR '98, 64-71, 1998.

[6] Buckwalter, T. Qamus: Arabic lexicography. http://members.aol.com/ArabicLexicons/.

[7] Darwish, K., Doermann, D., Jones, R., Oard, D., and Rautiainen, M. TREC-10 experiments at Maryland: CLIR and video. In TREC 2001. Gaithersburg: NIST, 549-561, 2001.

[8] Davis, M. W. and Ogden, W. C. Free resources and advanced alignment for cross-language text retrieval. In Proceedings of the sixth text retrieval conference (TREC-6), E. M. Voorhees and D. K. Harman (eds.). Gaithersburg: NIST Special Publication 500-240, 385-394, 1998.

[9] Gey, F. C. and Chen, A. Phrase discovery for English and cross-language retrieval at TREC-6. In Proceedings of the sixth text retrieval conference (TREC-6), E. M. Voorhees and D. K. Harman (eds.). Gaithersburg: NIST Special Publication 500-240, 637-647, 1998.

[10] Gey, F. C. and Oard, D. W. The TREC 2002 Arabic/English CLIR Track. In TREC2002 notebook, 81-93, 2002.

[11] GIZA++. http://www-i6.informatik.rwth-aachen.de/Colleagues/och/software/GIZA++.html

[12] Knight, K. and Graehl, J. Machine transliteration. Computational Linguistics, 24(4), 1998.

[13] Kraaij, W. TNO at CLEF-2001: Comparing translation resources. In Cross-language information retrieval and evaluation: Proceedings of the CLEF 2001 workshop, C. Peters (ed.), Springer, 2001.

[14] NMSU Arabic Proper Names Dictionary. http://crl.nmsu.edu/~ahmed/downloads.html.

[15] Peters, C. Cross-language information retrieval and evaluation, Workshop of cross-language evaluation forum, CLEF 2000, Lisbon, Portugal, September 21-22, 2000, Springer, 2001.

[16] Peters, C., Braschler, M., Gonzalo, J., and Kluck, M. Evaluation of cross-language information retrieval systems: Second workshop of the cross-language evaluation forum, CLEF 2001, Darmstadt, Germany, September 3-4, 2001. Springer, 2002.

[17] Pfeifer, U., Poersch, T., and Fuhr, N. Retrieval effectiveness of proper name search methods. Information Processing and Management, 32 (6), 667-679, 1996.

[18] Pirkola, A. The effects of query structure and dictionary setups in dictionary-based cross-language information retrieval. In Proceedings of SIGIR 98 (Melbourne, Australia, Aug 1998), ACM Press 55-63.

[19] Pirkola, A. and Järvelin, K. Employing the resolution power of search keys. Journal of the American Society for Information Science, 52 (7), 575-583, 2001.

[20] Pirkola, A., Keskustalo, H., Leppänen, E., Känsälä, A.-P., and Järvelin, K. Targeted s-gram matching: A novel n-gram matching technique for cross- and monolingual word form variants. Information Research, 7 (2), 2002.

[21] Schäuble, P. and Sheridan, P. Cross-language information retrieval (CLIR) track overview. In Proceedings of the sixth text retrieval conference (TREC-6), E. M. Voorhees and D. K. Harman (eds.). Gaithersburg: NIST Special Publication 500-240, 31-43, 1998.

[22] Stalls, B. G. and Knight, K. Translating names and technical terms in Arabic text. Proc of the COLING/ACL Workshop on Computational Approaches to Semitic Languages, 1998. http://citeseer.nj.nec.com/glover98translating.html

[23] Tarjim online translation. http://tarjim.ajeeb.com/ajeeb/.

[24] Whitaker, B. Arabic words and the Roman alphabet. http://www.al-bab.com/arab/language/roman1.htm

[25] World Cities. http://www.fourmilab.ch/earthview/cities.html.