

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Multiword Phrases Indexing for Malay-English Cross-Language Information Retrieval

N.H. Rais, M.T. Abdullah and R.A. Kadir

Department of Multimedia, Faculty of Computer Science and Information Technology,  
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

---

**Abstract:** Cross-Language Information Retrieval (CLIR) is the process of providing queries in one language and returning documents relevant to that query which is written in a different language. A popular approach to CLIR is to translate the query into the language of the documents being retrieved. One of the simplest and most effective methods for query translation is to perform dictionary look-up based on a bilingual dictionary. However, lack of dictionary coverage prune two problems: proper names and compound words handling. Relevance concept words consist of proper names and compound words, were applied in document and query indexing and query translation processes. We believed by using concept-based indexing and translations makes proper names and compound words translation possible. A series of experiments conducted to test the compound words and proper names translation methods in CLIR system. The best retrieval performance obtained from the combination of query translation approach-select all translations listed in the dictionary, alternative weighting scheme and proper names identification and translation. For both Malay and English document collection, these approaches outperformed query translation approach, select all translations listed in the dictionary, by 1.0 and 9%. The results show that proper names and compound words translations were important in query translation for Malay-English CLIR.

**Key words:** Concept-based IR, cross-language information retrieval, query translation, bilingual dictionary, proper names identification and translation

---

### INTRODUCTION

Information Retrieval (IR) deals with the representation, storage, organization of and access to information items (Baeza-Yates and Ribeiro-Neto, 1999). An IR system prepares for retrieval by indexing documents and formulating queries, resulting in document representations and query representations, respectively; the system then matches the representations and displays the documents found and the user selects the relevant items. Kasam and Hyuk-Chul (2006) and Premalatha and Natarajan (2010) applied vector space model to compute similarity matching between documents and query.

Cross Language Information Retrieval (CLIR) is a special case of Information Retrieval (IR). CLIR addresses this initial task of filtering, selecting and ranking documents that might be relevant to a query expressed in a different language. Abdullah *et al.* (2003) and Abdullah (2006) applied latent semantic indexing approach on Malay-English CLIR. The main translation approaches in

CLIR are: query translation (Pirkola *et al.*, 2001; Fujii and Ishikawa, 2001; Rais *et al.*, 2010), document translation (Hayurani *et al.*, 2007) and hybrid method (Bian and Chen, 1998; Kishida and Kando, 2006; Rahman *et al.*, 2006), where both query and documents were translated. Since it is more economical to translate queries rather than documents, query translation often preferred.

There are three main resources for query translation, Machine Translation (MT), bilingual dictionary and parallel or comparable corpora. In bilingual dictionary, each word or phrase in source language is translated into the target language by one or often several words or phrases.

The main problems reported in direct translation dictionary-based CLIR are: (1) the problem of inflection, (2) translation ambiguity, (3) compounds and phrases and their handling, (4) proper names and other untranslatable words and their handling and (5) lack of structuring (Pirkola *et al.*, 2001). However, we will focus on two problems in dictionary-based system: compounds translation and proper names translation.

---

**Corresponding Author:** Muhamad Taufik Abdullah, Department of Multimedia,  
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia,  
43400 UPM Serdang, Selangor, Malaysia Tel: 03-89471723

Particularly, dictionaries do not contain proper names, such as the names of capital cities and countries. However, in document collection consists of news documents, proper names are considered important text elements. A common methods used to handle untranslatable keywords is to include them as such in the target language. Proper names should be identified and often need to be translated (Grefenstette, 1998). A compound word is defined as a word form from two or more root words. From the IR point of view, compounds may content bearing words in natural language sentences and therefore important for retrieval result. Compound words handling does not applied in direct translation. These compound words should be translated as whole. English-Japanese CLIR system proposed by Fujii and Ishikawa (1999), focusing mainly in translation of technical compound words.

In the cognitive view of the world, there exists the presumption that the meaning of a text (word) depends on conceptual relationships to objects in the world rather than to linguistic or contextual relations found in texts or dictionaries. A new generation information retrieval model is drawn from this view. We call it concept-based information retrieval model. Sets of words, names, noun-phrases, terms, etc. will be mapped to the concepts they encode (Haav and Lubi, 2001).

Concept-based or context-based IR was applied to solve other problems. Wei *et al.* (2009) proposed an ontology-based system to solve problems raised in the manufacturing design. Prasannakumari (2010) deals with contextual information retrieval from multimedia databases which have feature descriptors and metadata for the data items in it. Lilleng and Tomassen (2007) analyzed the query translation in cross lingual IR based on feature vectors and usage of context information during the query translation. They pointed out that by using information external to the query, such as the ontologies and document collections, the effect of disambiguation and polysemi can be reduced.

This study focused on bilingual dictionary approaches for Malay-English CLIR using relevance concept words. These approaches were query translation, alternative weighting scheme and proper names identification and translation.

**DICTIONARY-BASED CLIR USING RELEVANCE CONCEPT WORDS**

**Relevance concept words:** For experiment purpose, list of concepts word were built manually, consists of proper names such as country names, capital city names, people, title, places and events and compound words. There were

282 entries in Malay and English concepts word based on 35 queries created for this experiment. These lists of concepts were used in document and query indexing process and query translation process.

**Example 1:**

- **Malay concepts word:** cuti bersalin, Pusat Konvensyen Kuala Lumpur, Wilayah Ekonomi Pantai Timur, Perdana Menteri
- **English concepts word:** maternity leave, Kuala Lumpur Convention Centre, East Coast Economy Region, Prime Minister

**Concept-based indexing:** In concept-based IR, sets of words, names, noun phrases are mapped into the concepts they represent. In these approaches, a document is represented as a set of concepts. Documents and queries indexing involved four main tasks: tokenization, create multiword phrase, concept identification and term weighting, as shown in Fig. 1. First, document texts were tokenized into single terms in tokenization process. Then, we combine 2 to 5 single terms to create multiword phrases. In concept identification, the phrases created in previous step were identified as concepts based on relevance concepts word.

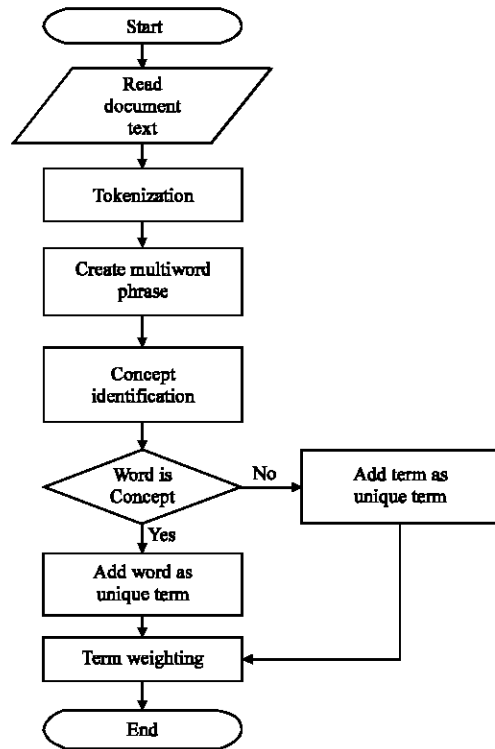


Fig. 1: Documents and queries indexing processes

Term weighting process takes place to assign weights to each unique concept. We were using tf.idf weighting scheme.

**Dictionary-based cross-language information retrieval:**

Many available bilingual dictionaries do not contain useful information to help the appropriate translation words or expression. Dictionaries are organized according to different principles. For this study, bilingual dictionary is a word list, together with their translations as follows:

**Malay-English dictionary:**

Kemenyan : Benzoin; a type of aromatic resin of a tree  
Kemerdekaan : Liberty; independence  
Kemerostotan : Slump; decadence  
Kemesraan : Absorption; love; feeling of being intimate

**English-Malay dictionary:**

Accidentally : Secara kebetulan; secara tak sengaja  
Accidents : Kemalangan; sesuatu yang terjadi; Nahas  
Acclaim : Bersorak; memuji  
Acclamation : Sambutan meriah

**Translation using bilingual dictionary faces two translation problems:**

How to translate and how to prune alternatives. There are two basics approaches have been proposed in the early study: (1) using the first translation listed in the dictionary and (2) using all the translations listed in the dictionary for each query words. The first approach is motivated by the fact that the first translation is often the most frequently used. However, this assumption on the organization of the dictionary is not true in many dictionaries. In our Malay-English bilingual dictionary, term kemerdekaan has listed liberty as the first translation candidate and independence as its second translation candidate, even though independence is most frequently used translation for kemerdekaan.

On the other hand, the second approach is motivated by the fact that when all the translations are used, one can include all the possible expressions in the target language and obtain query expansion effect. There are two results can be obtained from this approach: (1) improvement in retrieval performance if all translation candidates included in the query translation have the same semantics and (2) dropped in retrieval performance if incorrect translations were included in the query translation.

Term weighting in both documents and queries is an important aspect in IR. In CLIR, if a word retains many translations, the weight of that word would be artificially inflated if the query is simply sent to an IR engine. A simple solution to this problem is to divide the term weights for translations per source term, i.e., the weight of a translation becomes  $1/n$ , where  $n$  is the number of translation for the source term (Fujii and Ishikawa, 2001). Example 1 shows English to Malay language translation of Query 1 and Example 2 shows English to Malay language translation of Query 9.

**Example 2:** Malay to English query translation:

- Query 1: Sambutan kemerdekaan Malaysia ke-50
- Query translation approach, Method 1: Select the first translation listed in the dictionary
- Reception independence Malaysia
- Query translation approach, Method 2: Select all translation listed in the dictionary
- Reception welcome response celebration independence Malaysia

**Alternative term-weighting scheme:**

- Liberty (0.5) independence (0.5) Malaysia (1.0) reception (0.25) welcome (0.25) response (0.25) celebration (0.25)

**Example 3:** English to Malay query translation:

- Query 9: Accidents involving Nuri helicopter
- Query translation approach, Method 1: Select the first translation listed in the dictionary
- Kemalangan melibatkan nuri helikopter
- Query translation approach, Method 2: Select all translation listed in the dictionary
- Kemalangan sesuatu yang terjadi nahas melibatkan membabitkan mengenai nuri helikopter

**Alternative term-weighting scheme:**

- Kemalangan (0.334) sesuatu (0.334) yang (0.334) terjadi (0.334) nahas (0.334)
- Melibatkan (0.334) membabitkan (0.334) mengenai (0.334)
- Nuri (1.0)
- Helikopter (1.0)

**Concept-based translation using bilingual dictionary:**

Concept-based translation process consists of

concept-based tokenization, general dictionary look-up, proper name identification, proper name dictionary look-up, query construction and query weighting. First, punctuation marks, digits and hyphens were removed from user query before concept-based tokenization take place. Concept-based tokenization will produces query keywords based on relevance concept words. All relevance concept words existed in the query will be treated as one keyword.

There are two types of dictionaries being used in this experiment, general dictionary that contains root words and compound words and proper names dictionary that contains translation for proper names. General dictionary look-up was take place after tokenization task to translate root words and compound words. Concept-based tokenization makes compound words translation possible.

**Example 4:**

- Query 28: Pelancaran kapal selam pertama negara
- Concept-based tokenization: (pelancaran) (kapal selam) (pertama) (negara)
- General dictionary look-up: (launching) (submarine) (first) (country)

**Proper names identification and translation:** Proper names such as country names, city names, titles, events and places were written in capital letter in text. The simplest method to detect proper names in the text is by detection on capital letter. In order to allow proper names detection, query terms will not be converted to lowercase during tokenization process. Proper names identification based on relevance concept words take place right after concept-based tokenization. Proper names translations were applied in proper names translation look-up.

**Example 5:** Proper names handling:

- Query 8: Anugerah seni bina Aga Khan di Pusat Konvensyen Kuala Lumpur (KLCC)
- Concept-based Tokenization: (anugerah) (seni bina) (Aga Khan) (di) (Pusat Konvensyen Kuala Lumpur) (KLCC)
- Proper names identification: (Aga Khan) (Pusat Konvensyen Kuala Lumpur) (KLCC)
- Proper names dictionary look-up: (Aga Khan) (Kuala Lumpur Convention Centre) (KLCC)

The experimentation was setup to test these six approaches: (1) query translation-select first translation

listed in dictionary for each term source; (2) query translation-selects all translations listed in the dictionary for each term source; (3) query translation-select all translation listed in the dictionary and alternative term weighting scheme; (4) query translation-select first translation and proper name handling; (5) query translation-select all translation listed in the dictionary, alternative weighting scheme and proper name handling and (6) query translation-select all translations listed in the dictionary and proper name handling. The CLIR performances using these approaches were evaluated using Mean Average Precision (MAP) and average recall-precision graph.

## EXPERIMENTATIONS

To evaluate the effectiveness of query translation approaches in concept-based CLIR system for Malay-English language pair, we conducted a series of experiments using Malay-English alignment corpora. The English-Malay corpus contains 1, 446 news articles collected from Bernama News ([www.bernama.com](http://www.bernama.com)). For experiments purpose, we created 35 Malay queries covering a number of major events occurred in Malaysia during the period when the newspaper collection was built. For English monolingual IR, 35 Malay queries were manually translated to English language. The relevance judgments for Malay-English news collection were established manually. We then built term-document matrix for Malay and English documents.

For CLIR experiments, we used unidirectional bilingual dictionaries for Malay-English queries translation. The Malay-English bilingual dictionary contains 21,176 entries and English-Malay dictionary contains 22,228 entries. The dictionaries were collected from translation website and were manually edited. A program was built to translate 35 queries from Malay to English and vice versa. The program involves tokenization, dictionary look-up, query construction and query weighting.

Two basic translation approaches for query translation were tested in this experiment: (1) Method 1: Selecting the first translation listed in the dictionary, labeled as CLIR1 and (2) Method 2: Selecting all the translations for each query, labeled as CLIR2. The combination of query translation approach, Method2 and alternative weighting scheme was tested in this experiment, labeled as CLIR3. As shown in Fig. 2, CLIR1 and CLIR2 involve four processes: concept-based tokenization, general dictionary look-up, query

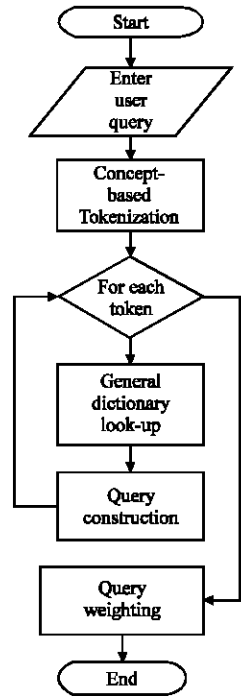


Fig. 2: Query translation processes occur in experiments CLIR1 and CLIR2. In experiment CLIR1, query translation method applied during general dictionary look-up is selects the first translation listed in the dictionary. In experiment CLIR2, query translation method, selects all translation listed in the dictionary were applied during general dictionary look-up process

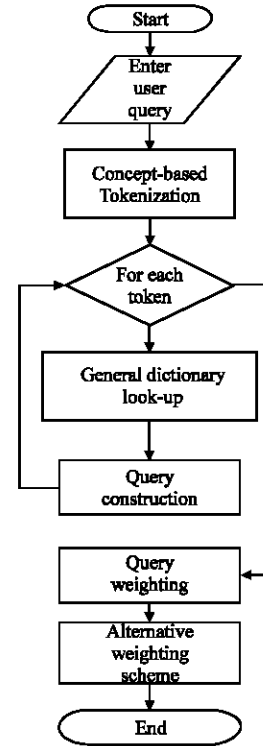


Fig. 3: Query translation processes occur in experiment CLIR3 consists of concept-based tokenization, query translation, selects all translations listed in the dictionary applied in general dictionary look-up, alternative weighting scheme, query weighting and query construction

construction and query weighting. For CLIR1, the first translation listed in the dictionary was selected in dictionary look-up process. While in CLIR2, all translations listed in the dictionary were selected. Figure 3 shows processes take place in CLIR3. Alternative weighting scheme was added to the query processes. In CLIR4, proper name identification and translation process were combined with query translation approach-method 1, as shown in Fig. 4. In CLIR5, query translation approach-method 2 was combined with alternative weighting and proper name handling. In CLIR6, query translation approach-method 2 and proper name handling. Figure 5 and 6 shows processes take place in experiments CLIR5 and CLIR6.

The experiment results were evaluated using Mean Average Precision (MAP) and Average Recall-Precision graph. We compared MAP results for CLIR with MAP result for monolingual IR, respectively.

## RESULTS

As can be expected, the retrieval performance of the direct translation method is lower than that of the equivalent monolingual methods. Table 1 shows the MAP results for CLIR experiments. For Malay document collection, English queries were translated into Malay language automatically. The MAP results for CLIR1 and CLIR2 were lower than Baseline, 5.7 and 4.7%, respectively. Query translation approach, select all translations listed in the dictionary outperformed query translation approach select the first translation listed in the dictionary result by 1%.

For English document collection, Malay queries were translated into English language. As shown in Table 1, the MAP results for CLIR1 and CLIR2 were 7.4 and 9.6% lower than Baseline result. For English document collection, query translation approach, selects the first translation listed in the dictionary outperformed CLIR using query translation approach, select all translation listed in the dictionary by 2.2%.

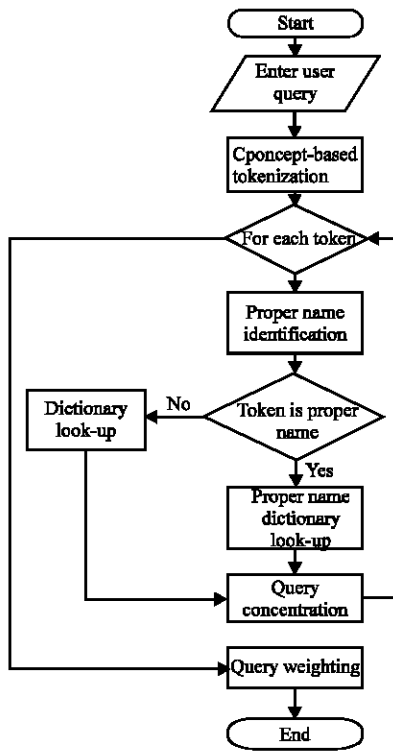


Fig. 4: Query translation events occur in experiment CLIR4 were concept-based tokenization, proper names identification, proper names dictionary look-up, general dictionary look-up, query construction and query weighting

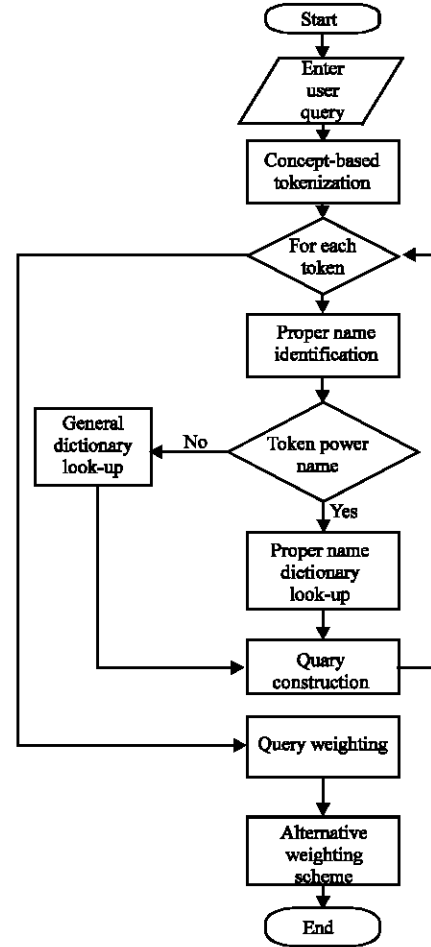


Fig. 5: Query translation processes occur in experiment CLIR5 were concept-based tokenization, proper names identification, proper names dictionary look-up, query translation approach, select all translations listed in the dictionary applied during general dictionary look-up, query construction, query weighting and alternative weighting scheme

Table 1: Comparison of MAP results of experiment CLIR1, CLIR2, CLIR3, CLIR4, CLIR5 and CLIR6 with monolingual IR (baseline) for Malay and English document collections

Experiments	MAP results	
	Malay document	English document
Monolingual IR (Baseline)	0.811	0.746
CLIR1 (First translation)	0.754	0.672
CLIR2 (All translations)	0.764	0.650
CLIR3 (All translations + Alternative weighting)	0.772	0.714
CLIR4 (First translation + Proper name Handler)	0.752	0.703
CLIR5 (All translations + Alternative weighting + Proper name Handler)	0.774*	0.740*
CLIR6 (All translations + Proper name Handler)	0.769	0.710

\*Indicate the best result obtained from Cross-Language Information Retrieval (CLIR) experiments

Alternative weighting scheme applied with query translation approach, select all translations listed in the dictionary seems to improve MAP results for Malay and English document collection. For Malay collection, MAP result for experiment CLIR3 improves 0.8% compared to experiment CLIR2 result. For English collection, MAP

result for experiment CLIR3 improves 6.4% compared to experiment CLIR2 result.

Proper names identification and translation applied in experiments CLIR4, CLIR5 and CLIR6 can improve retrieval performance for Malay and English document collection. For Malay document collection, combination of query translation approach, select the first translation listed in the dictionary, proper names handling and alternative weighting scheme in experiment CLIR5 outperformed experiment CLIR3 by 0.2%. Experiment CLIR6 which was the combination of query translation approach, select all translations listed in the dictionary and proper names handling improved MAP result by 0.5%

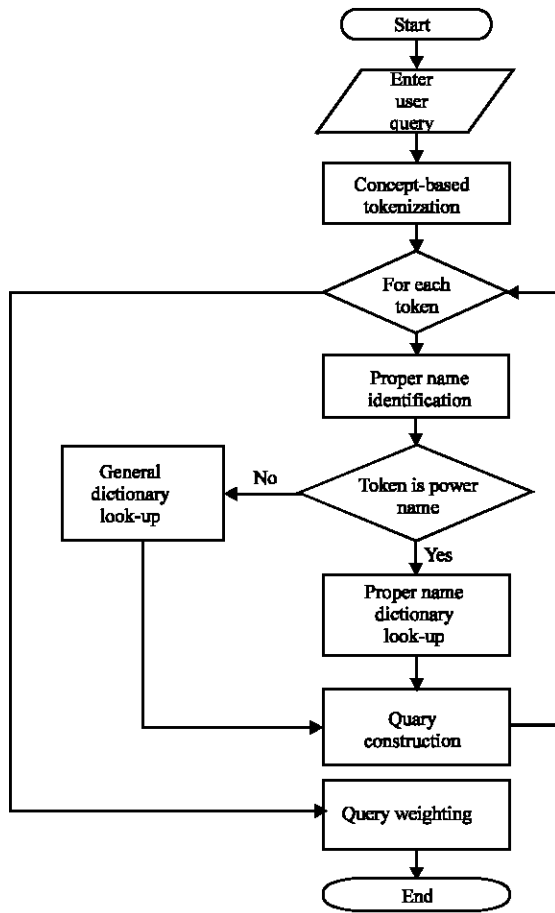


Fig. 6: Query translation events occur in experiment CLIR6 were concept-based tokenization, proper names identification, proper names dictionary look-up, query translation approach select the first translation listed in the dictionary applied in general dictionary look-up, query construction and query weighting

compared to MAP result of CLIR2. However, performance of experiment CLIR4 slightly dropped 0.2% compared to CLIR1 result.

For English document collection, the combination of query translation approach, select the first translation listed in the dictionary and proper names handling in experiment CLIR4 improve 3.1% compared to CLIR1 MAP result. Experiment CLIR5 outperformed CLIR3 MAP result by 2.6% and experiment CLIR6 outperformed CLIR2 result by 6%.

The best MAP results for Malay and English document collection obtained from experiment CLIR5, the combination of query translation approach, select all translations listed in the dictionary, alternative weighting

scheme and proper names identification and translation. Table 2 and 3 show average precision at 11-recall points for CLIR experiments for Malay and English document collections.

## DISCUSSION AND CONCLUSION

In this study we evaluated the effectiveness of bilingual dictionary approaches for Malay-English CLIR using relevance concepts word. For experiment purposed, relevance concepts words consists of compound words and proper names were created based on 35 queries used in the experiment. Relevance concepts words were applied in document/query indexing and query translation tasks. Concept-based indexing makes compound word and proper names translation possible.

For English and Malay document collections, as shown in Table 2 and 3, the best retrieval performance obtained from experiment CLIR5, the combination of query translation approach; select all translations listed in the dictionary, alternative weighting scheme and proper names identification and translation. These approaches solved three query translation problems discussed by Pirkola *et al.* (2001): translation ambiguity due to multiple translations obtained for one query term, compound words translation and proper names identification and translations.

Rais *et al.* (2010) reported the characteristics of compound words in Malay and English languages. In Malay language, compound words either written separately or link together when they are bound by circumfix or when they are considered as stable words. In English language, compound words in English language are written in three ways: hyphenated compound, an open compound and a solid compound. Included all compound words in bilingual dictionaries is the easiest yet effective way to translated compound words. Compound words such as kapal terbang and kapal selam were considered as solid compounds but written separately in text. By using concept-based indexing in documents and queries, these compounds were taken in as one index item and can be translated directly by dictionary. However, since the new compound words emerge every day, we need an approach to automatically acquire new compounds as suggested in (Zhang and Isahara, 2004).

In the study of Grefenstette (1998), proper names in texts often need to be translated. Country names, city names and events, often written with capital letter at the beginning of each character. The simplest way to identified proper names from document text was by detection of capital letter. However, these approach cannot detect proper names written at the beginning of



Table 2: Average precision at 11-recall points for monolingual IR, experiments CLIR1, CLIR2, CLIR3, CLIR4, CLIR5 and CLIR6 for Malay document collection

Recall	Precision						
	Monolingual IR	CLIR1	CLIR2	CLIR3	CLIR4	CLIR5	CLIR6
0.0	0.914	0.818	0.866	0.849	0.818	0.848	0.862
0.1	0.902	0.812	0.863	0.846	0.816	0.846	0.867
0.2	0.895	0.810	0.859	0.854	0.810	0.846	0.852
0.3	0.889	0.818	0.835	0.850	0.815	0.843	0.832
0.4	0.874	0.811	0.819	0.847	0.812	0.842	0.824
0.5	0.873	0.808	0.811	0.832	0.798	0.821	0.811
0.6	0.858	0.786	0.806	0.805	0.775	0.797	0.803
0.7	0.850	0.773	0.788	0.780	0.758	0.771	0.783
0.8	0.791	0.738	0.745	0.734	0.747	0.727	0.740
0.9	0.775	0.720	0.729	0.720	0.703	0.715	0.722
1.0	0.622	0.580	0.567	0.583	0.554	0.571	0.555
Average	0.840	0.770	0.790	0.791	0.764	0.784	0.786

Table 3: Average precision at 11-recall points for monolingual IR, experiment CLIR1, CLIR2, CLIR3, CLIR4, CLIR5 and CLIR6 for English document collection

Recall	Precision						
	Monolingual IR	CLIR1	CLIR2	CLIR3	CLIR4	CLIR5	CLIR6
0.0	0.853	0.789	0.739	0.797	0.818	0.832	0.815
0.1	0.847	0.771	0.748	0.794	0.800	0.820	0.819
0.2	0.842	0.750	0.746	0.768	0.781	0.810	0.805
0.3	0.809	0.737	0.744	0.771	0.756	0.801	0.794
0.4	0.791	0.731	0.724	0.763	0.756	0.772	0.775
0.5	0.772	0.699	0.678	0.738	0.720	0.744	0.728
0.6	0.761	0.688	0.665	0.735	0.705	0.743	0.712
0.7	0.734	0.675	0.652	0.707	0.696	0.717	0.697
0.8	0.694	0.627	0.610	0.636	0.645	0.674	0.648
0.9	0.675	0.610	0.582	0.610	0.627	0.643	0.618
1.0	0.519	0.481	0.458	0.510	0.507	0.549	0.519
Average	0.754	0.687	0.668	0.712	0.710	0.737	0.721

the sentences since capital letter were used at the beginning of each sentences. Proper names translation using dictionary was efficient depends on the coverage of the dictionary. It is possible to build proper names dictionary for Malay-English language pair but it require more times and cost. We suggest generating proper names dictionary automatically based on documents in the internet or document collection analysis.

**REFERENCES**

Abdullah, M.T., 2006. Monolingual and cross-language information retrieval approaches for Malay and English documents. Ph.D. Thesis, Universiti Putra Malaysia.

Abdullah, M.T., F. Ahmad, R. Mahmood and T.M.T. Sembok, 2003. Application of latent semantic indexing on malay-english cross language information retrieval. Lect. Notes Comput. Sci., 2911: 663-665.

Baeza-Yates, R. and B. Ribeiro-Neto, 1999. Modern Information Retrieval. Addison Wesley, London.

Bian, G.W. and H.H. Chen, 1998. Integrating query translation and document translation in a cross-language information retrieval system. Lect. Notes Comput. Sci.. 1529: 250-265.

Fujii, A. and T. Ishikawa, 1999. Cross-language information retrieval using compound word translation. Proceedings of 18th International Conference on Computer Processing of Oriental Languages, March 24-26, Tokushima, Japan, pp: 105-110.

Fujii, A. and T. Ishikawa, 2001. Japanese/English cross-language information retrieval: Exploration of query translation and transliteration. Comput. Hum., 35: 389-420.

Grefenstette, G., 1998. The Problem of Cross-Language Information Retrieval. In: Cross-Language Information Retrieval, Grefenstette, G. (Ed.). Kluwer Academic Publishers, Boston, pp: 1-9.

Haav, H.M. and T.L. Lubi, 2001. A survey of concept-based information retrieval tools on the web. Proc. East-Eur. Conf. ADBIS, 2: 29-41.

Hayurami, H., S. Sari and M. Adriani, 2007. Query and document translation for English-Indonesian cross language IR. Lect. Notes Comput. Sci. 4730: 57-61.

- Kasam, A. and K. Hyuk-Chul, 2006. Consolidation of diversifying terms weighting impact on IR system performances. *Inform. Technol. J.*, 5: 7-12.
- Kishida, K. and N. Kando, 2006. A hybrid approach to query and document translation using a pivot language for cross-language information retrieval. *Lect. Notes Comput. Sci.*, 4022: 93-101.
- Lilleng, J. and S.L. Tomassen, 2007. Cross-lingual information retrieval by feature vectors. *Lect. Notes Comput. Sci.*, 4592: 229-239.
- Pirkola, A., T. Hedlund, H. Keskustalo and K. Jarvelin, 2001. Dictionary-based cross-language information retrieval: Problems, methods and research findings. *Inf. Retrieval*, 4: 209-230.
- Prasannakumari, V., 2010. Contextual information retrieval for multi-media databases with learning by feedback using vector space model. *Asian J. Inform. Manage.*, 4: 12-18.
- Premalatha, K. and A.M. Natarajan, 2010. A literature review on document clustering. *Inform. Technol. J.*, 9: 993-1002.
- Rahman, S.A., N. Ahmad, H.A. Hashim and A.W. Dahalan, 2006. Real time on-line English-Malay Machine Translation (MT) system. Proceedings of the 3rd Real-Time Technology and Application Symposium, Dec. 5-6, IEEE Computer Society, pp: 1-7.
- Rais, N.H., M.T. Abdullah and R.A. Kadir, 2010. Query translation architecture for Malay-English cross-language information retrieval system. Proceedings of the International Symposium on Information Technology, June 15-17, Kuala Lumpur, pp: 990-993.
- Wei, S., M. Qin-Yi and G. Tian-Yi, 2009. An ontology-based manufacturing design system. *Inform. Technol. J.*, 8: 643-656.
- Zhang, Y. and H. Isahara, 2004. Acquiring compound word translations both automatically and dynamically. Proceedings of the PACLIC, Dec. 8-10, Waseda University, Tokyo, pp: 181-186.