

Query Translation using Concepts Similarity Based on Quran Ontology for Cross-Language Information Retrieval

¹Zulaini Yahya, ¹Muhamad Taufik Abdullah,
¹Azreen Azman and ²Rabiah Abdul Kadir

¹Department of Multimedia,

²Department of Computer Science,

Faculty of Computer Science and Information Technology,

Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

Received 2012-05-02, Revised 2012-09-06; Accepted 2013-06-21

ABSTRACT

In Cross-Language Information Retrieval (CLIR) process, the translation effects have a direct impact on the accuracy of follow-up retrieval results. In dictionary-based approach, we are dealing with the words that have more than one meaning which can decrease the retrieval performance if the query translation return an incorrect translations. These issues need to be overcome using efficient technique. In this study we proposed a Cross-Language Information Retrieval (CLIR) method based on domain ontology using Quran concepts for disambiguating translation of the query and to improve the dictionary-based query translation. For experimentation, we use Quran ontology written in English and Malay languages as a bilingual parallel-corpora and Quran concepts as a resource for cross-language query translation along with dictionary-based translation. For evaluation, we measure the performance of three IR systems. IR₁ is natural language query IR, IR₂ is natural language query CLIR based on dictionary (as a Baseline) and IR₃ is the retrieval of this research proposed method using Mean Average Precision (MAP) and average precision at 11 points of recall. The experimental result shows that our proposed method brings significant improvement in retrieval accuracy for English document collections, but deficient for Malay document collections. The proposed CLIR method can obtain query expansion effect and improve retrieval performance in certain language.

Keywords: English Language, Malay Language, Bilingual Dictionary, Quran Concepts, Quran Ontology

1. INTRODUCTION

Nowadays, the usage of computers and the Internet has grown. More than one billion people use the Internet and get a lot of benefit from the available information. This information not only written in their native language but also in other non-native languages and expanded rapidly followed the growth of internet information. Information Retrieval (IR) generally refers to the process that user searches the needed information from a large number of documents. Traditional IR is implemented mainly for monolingual documents and only

supports the retrieval of documents that are written in the same language as the query. Cross-Language Information Retrieval (CLIR) is intended to matching the user query written in one language with the documents written in other languages. In CLIR, systems automatically search documents written in other languages.

Translation in CLIR can be performed either on the query, document or both into an Interlingua representation (Saralegi and Lacalle, 2009; Sheridan and Ballerini, 1996). Most CLIR systems used the query translation approach to reduce difficulty in translating a large document collection and multi-lingual translation.

Corresponding Author: Zulaini Yahya, Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Translation methods, they can be classified into three main groups: Machine Translation (MT)-based, parallel corpora-based and Bilingual Machine Readable Dictionary (MRD)-based (Sheridan and Ballerini, 1996). MRD-based query translation has been a common method in CLIR system. In these methods, we face with the problem of translation ambiguity in which a single word in one language has more than one translation in the other language (Pourmahmoud and Shamsfard, 2008). By using information external to the query, like ontologies and document collections, the effect of ambiguity can be reduced (Lilleng and Tomassen, 2007). Ontology is a formal, explicit specification of a shared conceptualization. It contains a set of distinct and identified concepts related by a set of relations (Pourmahmoud and Shamsfard, 2008) and can be implemented in translation systems to extract conceptual relations for monolingual and CLIR (Abusalah *et al.*, 2005). A bilingual ontology consists of an ontology and a bilingual dictionary, can be used to annotate the documents and queries (Pourmahmoud and Shamsfard, 2008).

Based on this approach, we proposed another CLIR method, English-Malay query translation based on Quran ontology using Quran concepts. The performance of the proposed CLIR method will be compared to mono IR and MRD-based query translation approach.

The structure of this study is organized as follows. Section 2 explained in detail the proposed CLIR architecture and the experimentation setup to evaluate performance. In section 3, the results of the experimentation will be presented and analyzed in section 4. Finally, the conclusions and the future works are given in section 5.

1.1. Related Work

CLIR provides a convenient way to solve the translation problem between two or more languages. Many researchers have focused on query translation approach mainly due to the lower requirements of memory and processing resources (Hull and Grefenstette, 1996). Documents translation approach and the use of Interlingua representation approach dealing with a huge document collections and impractical to translate, but with this two approach more context information for disambiguation can be provided (Kimura *et al.*, 2008). The richer context information is useful for dealing with disambiguation problems (Saralegi and Lacalle, 2009). In MT-based approach, the MT system is usually use a complete sentence, but in IR, the query sentence often comprises some query keywords, tend to be very short, lack of context and lack of the integrity of semantic information. In dictionary-

based approach, processed queries are translated by looking up terms in bilingual MRD. It's a simple method, but hard to dealing with the words that have more than one meaning. The major problem in bilingual dictionary approach is translation ambiguity (as in the case for MT system) in addition to problems of word inflection, problems of translating word compound, phrases, proper names, spelling variants and special term (Abusalah *et al.*, 2005). In CLIR process, the translation effects have a direct impact on the accuracy of follow-up retrieval results. MT-based and dictionary-based approach cannot achieve satisfactory result.

Various topics in multiple languages available through the internet and can be used as a domain, which can improve the retrieval effectiveness. Cheng *et al.* (2008) adopt bilingual domain ontology and Muller and Gurevych (2008) used Wikipedia and Wiktionary as a specific domain. Closed or restricted domains CLIR approaches do traditionally produce a better result compared to CLIR used in open domains (Lilleng and Tomassen, 2007) Despite these promising results, they are highly depended on a fairly common terminology being used.

Lilleng and Tomassen (2007) and Tomassen *et al.* (2006) investigates query translation in CLIR, caused by ambiguity and polysemy, based on feature vectors and their method uses context during the translation of queries. Tomassen *et al.* (2006) Use a query enrichment approach to associate every concept of the ontology with a feature vector to tailor these concepts to the specific terminology used in the document collection. Cheng *et al.* (2008) brings forward a Web CLIR model based on the domain ontology, to describe the relevant domain knowledge in different kinds of languages, comprehend and extend query terms to improve the average precision/recall of retrieval. Pourmahmoud and Shamsfard (2008) use bilingual ontologies to annotate the documents and queries. Polpinij (2009) propose a method for simplifying ambiguity of requirement specification documents through two concepts of ontology-based probabilistic text processing: Text classification and Text Filtering to reduced ambiguity translation.

In order to deal with the translation selection problem affecting queries derived from bilingual MRD, there are several methods proposed in the literature. Structured queries as an extended approach to tackle the problem of ambiguity, also called Pirkola's method (Pirkola, 1998) Probabilistic structured queries, allows to weight the different translation candidates offering better performance (Darwish and Oard, 2003). In Ballesteros and Bruce (2008) the co-occurrence method is significantly better at disambiguating than the

parallel corpus-based technique. Gao *et al.* (2002) the basic co-occurrence is extended by adding a decaying factor. Liu *et al.* (2005) propose a co-occurrence method. Monz and Dorr (2005) introduces an iterative co-occurrence method which combines term association measures with an iterative machine learning approach based on expectation maximization.

CLIR systems that have combined query translation to domain ontology could show a better results as domain ontology can reduce ambiguity translation. This research will focused on translation using query translation approach in MRD-based method and bilingual domain ontology. In this approach, a sources language query is first translated into the target language using bilingual dictionary and the translated query is then disambiguated. In relation to Web retrieval, we used English and Malay Quran concepts to resolve ambiguities.

2. MATERIALS AND METHODS

2.1. Cross Language Information Retrieval Using Quran Ontology

2.1.1. Quran Ontology

Quran ontology uses knowledge representation to define the key concepts in the Quran and shows the relationships between these concepts using predicate logic. The fundamental concepts in the ontology are based on the knowledge contained in traditional sources of Quran analysis, including hadith and tafsir. Concept is the name of the root entity in the Quran ontology. All other concepts and subcategories in the ontological hierarchy are organized under this root. It is consists of 300 linked concepts with 350 relations (Dukes, 2009).

2.2. Concepts Similarity

The proposed concepts similarity is applied using cosine similarity for computing the relevancy between concepts verses and the documents or verses in Quran document collections. In this approach, we use the cosine similarity between the document vector and a concepts vector as a measure of the score of the concepts for that document. The resulting scores can then be used to select the top-scoring concepts for a document. Each concept may share the same main verse. Based on this main verse, we calculate the concepts similarity for every document in Quran collections. Each document may have more than one concept depends on their main verse and score.

Based on this calculation, 5,695 English documents have concepts and another 541 document does not belong to any concepts. 5,999 Malay documents have concepts and another 237 document does not belong to any concepts.

2.3. Bilingual Ontology

Based on category, subcategories and related concepts in Quran ontology, we build a bilingual ontology, consists of Quran concepts and document concepts. The Quran concepts will contain a list of English and Malay Quran concepts as shown in **Table 1**. The document concepts will contain a list of English and Malay document and their related concepts as shown in **Table 2**.

2.4. Bilingual Dictionary

In bilingual dictionary, each word or phrase in source language is translated into the target language by one, or often several words or phrases. Bilingual dictionaries can be unidirectional or bidirectional (Rais *et al.*, 2011). In this research we also use unidirectional dictionary for Malay-English and English-Malay translation. Dictionaries information may disambiguate, not useful and not complete. To reduce this type of problems, we use term in the bilingual dictionary together with term and their translation from the bilingual ontology to build a new combination bilingual dictionary. This bilingual dictionary will contain a word list from bilingual dictionary and bilingual ontology, together with their translations as shown in **Table 3**.

We adopted two basics approaches from Rais *et al.* (2011) work: (1) using the first translation listed in the dictionary, motivated by the fact that the first translation is often the most frequently used; and (2) using all the translation candidates, motivated by the fact that when all the translation candidates 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 query and translation candidates have the same semantic meanings and (2) decrease in retrieval performance if query translation return an incorrect translations.

Table 1. Examples of english and malay Quran concepts

Quran Concepts	
English	Malay
Artifact	Artifak
Place of Worship	Rumah Ibadat
Mosque	Masjid

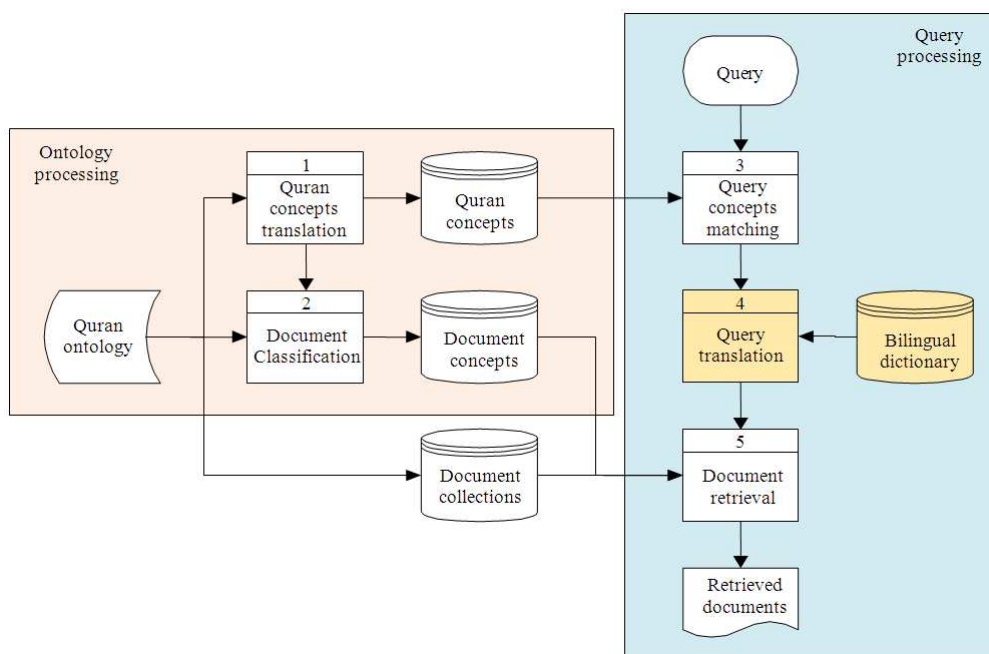


Fig. 1. Architecture design of the proposed CLIR method

Table 2. Examples of english and malay document concepts

English Document		Malay Document	
Verse	Concepts	Verse	Concepts
01:04	Saqar, Last Day, Day of Resurrection, Christianity Kristian	01:04	Neraka Saqar, Hari Akhirat, Hari Kiamat,
01:05	Church, Synagogue, Monastery, Ink	01:05	Gereja, Tempat Sembahyang, Biara, Dakwat

Table 3. English-malay bilingual dictionaries

Word (English)	Translation (Malay)	Source
chapter information	fasal, babak, bab maklumat, khabar, penerangan	bilingual dictionary
hudud	hudud	bilingual ontology
fasting	berpuasa	

2.5. The Proposed CLIR Architecture Design

Figure 1 illustrates the outline of the proposed CLIR architecture design. It consists of Quran concepts translation, document classification, query concepts matching, query translation and document retrieval.

In ontology processing, we build bilingual ontology which includes list of English and Malay concepts and a list of English document collections and their related concepts. To match concepts between English and Malay languages firstly we translate the English concepts using English-Malay-Arabic dictionary and Quran collections to Malay concepts. Then we estimate the corresponding

concepts in the Malay language by comparing the related verses in the English concepts. For document classification, we calculate similarity scores between term in Quran concepts and term in document collections. The most similar concepts are assumed to be concepts for the document. Top ranking similar concepts can be used as the expansion keys.

In query processing, we use a bilingual dictionary and bilingual ontology to translate and to calculate the related concepts for each query. In query concepts matching, we calculate similarity scores between term in Quran concepts and term in queries.

Query translation will be made by using the bilingual dictionary for first translation and all translation candidates. Document retrieval retrieves documents using the translated query and query concepts.

For the evaluation purposes, we measure the performance of three IR systems. IR₁ is natural language query IR, IR₂ is natural language query CLIR based on dictionary and IR₃ is the retrieval of this research proposed method, first, input Malay (also English) terms, match concepts and translate these into English (also Malay) candidates, then query information by use of the proposed methods.

The experimentation was setup to test these five approaches: (1) IR query retrieval, (2) CLIR query translation using first translation, (3) CLIR query translation using all translation candidates, (4) CLIR query translation using first translation and Quran concepts and (5) CLIR query translation using all translation candidates and Quran concepts. The IR and CLIR performance using these approaches were evaluated using Mean Average Precision (MAP) and average precision at 11 points of recall.

2.6. Experimentation

We conducted experiment on the proposed CLIR method using English and Malay version of Quran concepts, collected and translated from "Ontology of Quranic Concepts" by Dukes (2009). In these experiments, we used Malay queries and retrieved English documents and vice versa. To evaluate the effectiveness of the proposed CLIR method, we used Malay-English documents sets from actual Malay Quran collection and actual English Quran collection from Abdullah (2006) to verify the proposed method. The English-Malay corpora contain 6,236 documents in each language. The set of queries and relevance judgments adopted in this experiment was collected by Abdullah (2006). To resolve ambiguities, we use English and Malay Quran concepts from Dukes (2009). For experiments purpose, we used 36 Malay and English queries covering a number of major issues in Quran ontology.

For IR₁, Mono IR system experiments, we build a simple program to retrieve documents in query language.

For IR₂, Bilingual CLIR system experiments, we used unidirectional bilingual dictionaries for Malay-English (also English-Malay) queries translation. The

English-Malay bilingual dictionary contains 22,279 entries and Malay-English bilingual dictionary contains 21,209 entries. The dictionaries were collected from the internet, Quran ontology and other available collections. Two basic translation approaches were tested in this experiment: IR₂1, using first translation listed in the dictionary and IR₂2, using all the translation candidates for each query.

For IR₃, proposed CLIR method experiments, we used unidirectional bilingual dictionaries and the Quran concepts for concepts matching. Two approaches were tested in this experiment: IR₃1, using first translation and Quran concepts and IR₃2, using all translation candidates and Quran concepts. To compare the performance of CLIR proposed method, we used IR₂ as a Baseline.

The results of this experiment are analyzed using Mean Average Precision (MAP) and average precision at 11 points of recall for English-Malay retrieval and vice versa as measures of retrieval effectiveness.

3. RESULTS

Table 4 shows the performance of mono IR (IR₁), dictionary-based CLIR systems (IR₂) and concepts similarity approach (IR₃) for English and Malay document collections. The effectiveness of an information retrieval system is evaluated in terms of MAP by applying different translation approaches.

As shown in **Table 4**, the English document MAP results for IR₂ were 2.1 and 5.1% lower than mono IR₁ result and the Malay document MAP results for IR₂ were 3.4 and 7.5% lower than mono IR₁ result, as expected. The English document MAP result for IR₃ is higher than IR₂ by 2 and 0.6% and the Malay document MAP result for IR₃ is lower than IR₂ by 2.5 and 2.8%, respectively.

Average precision at 11 points of recall curves for mono IR and dictionary-based CLIR methods for English and Malay document collections are also depicted in **Fig. 2 and 3**. As can be seen in this figure, dictionary-based CLIR approach in IR₂, showed decreasing in retrieval performance compared to Mono IR system in IR₁. Dictionary-based CLIR using first translation, obtained an improvement in retrieval performance compared to dictionary-based CLIR using all translation candidates for English and Malay document collections.

Table 4. MAP result for English and Malay document

Method	English Document			Malay Document			
	Approaches	Mono IR	First Translation	All Translation	Mono IR	First Translation	All Translation
IR ₁		0.143			0.175		
IR ₂ (Baseline)	IR ₂ 1 and IR ₂ 2		0.122	0.092		0.141	0.100
IR ₃	IR ₃ 1 and IR ₃ 2		0.142	0.098		0.085	0.072

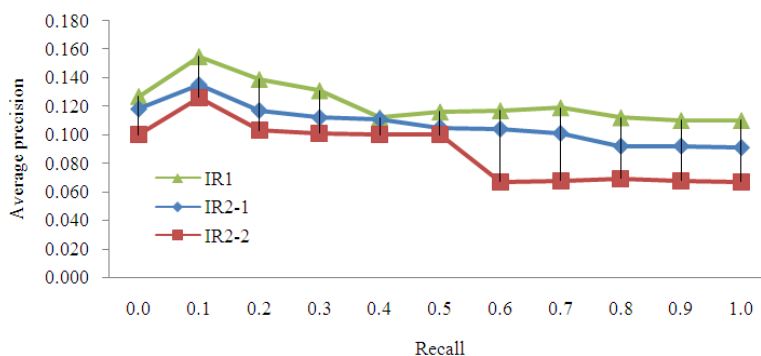


Fig. 2. Mono IR and dictionary-based CLIR approach: Average precision at 11 points of recall for English document collections

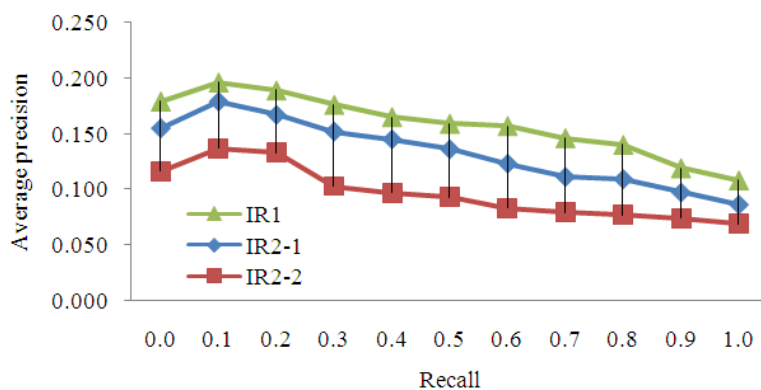


Fig. 3. Mono IR and dictionary-based CLIR approach: Average precision at 11 points of recall for Malay document collections

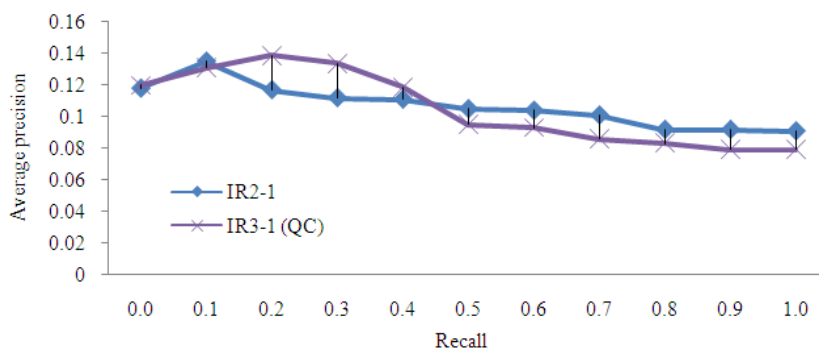


Fig. 4. Concepts similarity approach: Average precision at 11 points of recall for English document collections using first translation approach

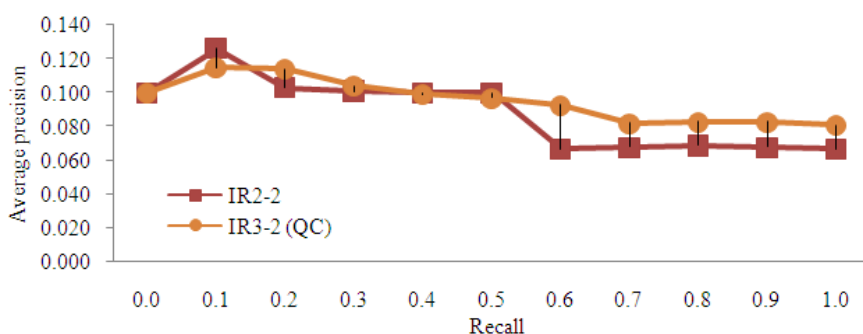


Fig. 5. Concepts similarity approach: Average precision at 11 points of recall for Malay document collections using first translation approach

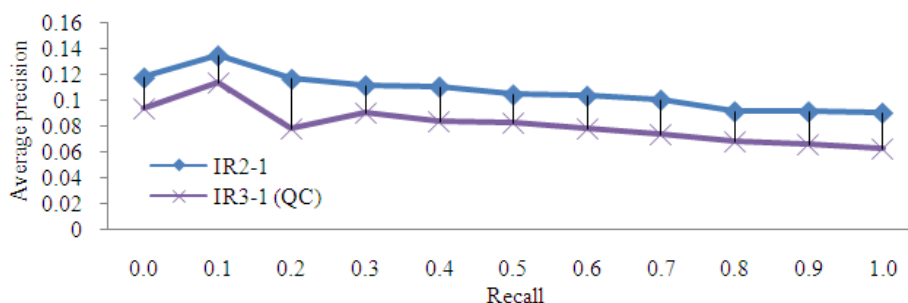


Fig. 6. Concepts similarity approach: Average precision at 11 points of recall for Malay document collections using first translation approach

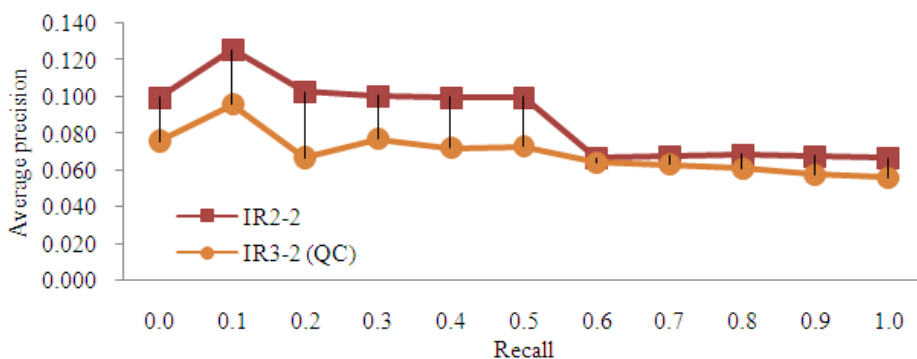


Fig. 7. Concepts Similarity Approach: Average precision at 11 points of recall for Malay document collections using all translation approach

Translation approach in IR₂, showed that CLIR query translation using first translation listed in the dictionary is obtained a better result compared to using all translation candidates listed in the dictionary either in English or Malay document collections. As shown in Table 4, the MAP results for using first translation were 3% and 4.1% higher than using all translation candidates result for English and Malay document collections.

Average precision at 11 points of recall curves for dictionary-based and proposed CLIR methods for English and Malay document collections by concepts similarity approach using first translation listed in the dictionary are depicted in Fig. 4 and 6 and using all translation listed in the dictionary are depicted in Fig. 5 and 7.

4. DISCUSSION

The main purpose of experiment is the comparison between dictionary-based query translation methods which were applied in the benchmark from Rais *et al.* (2011) and the proposed CLIR query translation method for retrieving relevant document in CLIR approaches. The proposed CLIR method considers Quran concepts in addition to query expansion and translation ambiguity.

To compare the performance CLIR proposed method, we used IR₂ as a Baseline (benchmark).

The result in **Table 4**, shows that direct translation in dictionary-based CLIR approach, involved in returning many possibility results which can affect the decreasing in retrieval performance and by limiting to one translation with the most frequently used term, will prevent us from receiving irrelevant documents using unrelated term.

As can be seen in **Fig. 4 and 5**, ontology based on concepts similarity approach in IR₃, either using translation concepts in both first translation and all translation approach showed that CLIR query translation using IR₃, obtained query expansion effect and improve retrieval performance compared to IR₂. Using first translation approach, obtained a better result compared to using all translation candidates approach.

As can be seen in **Fig. 6 and 7**, ontology based on concepts similarity approach in IR₃, using translation concepts in both first translation and all translation approach showed that CLIR query translation using IR₃, obtained query expansion effect and improve retrieval performance compared to IR₂. Using first translation approach, obtained a better result compared to using all translation candidates approach.

5. CONCLUSION

In this study we evaluate the effectiveness of query translation using bilingual dictionary and bilingual ontology for CLIR system.

Translation approach in IR₂ and IR₃, showed that CLIR query translation using first translation listed in the dictionary is obtained a better result compared to using all translation candidates listed in the dictionary either in English or Malay document collections. This result shows that by limiting to one translation with the most frequently used term, will prevent us from receiving irrelevant documents using unrelated term.

Ontology approaches in IR₃, showed that CLIR query translation using IR₃, obtained query expansion effect and improve retrieval performance compared to IR₂

for English document collections, but not for Malay document collections. CLIR query translation for Malay document collections using IR₃, is deficient than IR₂.

There are two problems have been identified in using IR₃ approach that caused increasing and decreasing in CLIR performance. The problems appear in concept matching after query translation and document retrieval. The query concepts may not in the translated query concepts list. Therefore, no relevant document will be return. In document retrieval, we assigned document concepts by calculating the similarity scores between concepts and document. One verse may have different concepts in different languages. Therefore, the possibility not to retrieve a relevant document will happen. To reduce this type of problem we may assign same concepts for every verse in English and Malay document collections by using a different technique as a future work study.

This result shows that by adding the bilingual ontology into bilingual dictionary system, using concepts similarity, can obtain query expansion effect and improve retrieval performance in certain language.

6. REFERENCES

- Abdullah, M.T., 2006. Monolingual and Cross-Language Information Retrieval Approaches for Malay and English Documents. Ph. D. Thesis, Universiti Putra Malaysia.
- Abusalah, M., J. Tait and M. Oakes, 2005. Literature review of cross language information retrieval. *World Acad. Sci. Eng. Technol.*, 4: 175-177.
- Ballesteros, L. and W.C. Bruce, 2008. Resolving ambiguity for cross-language retrieval. *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 24-28, ACM Press, Melbourne, Australia, pp: 64-71. DOI: 10.1145/290941.290958
- Cheng, X., H. Guo, Y. Wang and W. He, 2008. Research on web cross language information retrieval based on domain ontology. *Proceeding of the ISECS International Colloquium on Computing, Communication, Control and Management*, Aug. 3-4, IEEE Xplore Press, Guangzhou, pp: 622-626. DOI: 10.1109/CCCM.2008.291
- Darwish, K. and D.W. Oard, 2003. Probabilistic structured query methods. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 28-Aug. 01, ACM Press, Toronto, ON, Canada, pp: 338-344. DOI: 10.1145/860435.860497

- Dukes, K., 2009. The Quranic Arabic Corpus-Ontology of Quranic Concepts.
- Gao, J., J.Y. Nie, H. He, W. Chen and M. Zhou, 2002. Resolving query translation ambiguity using a decaying co-occurrence model and syntactic dependence relations. Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Aug. 11-15, ACM Press, Tampere, Finland, pp: 183-190. DOI: 10.1145/564376.564409
- Hull, D.A. and G. Grefenstette, 1996. Querying across languages: A dictionary-based approach to multilingual information retrieval. Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Aug. 18-22, ACM Press, Zurich, Switzerland, pp: 49-57. DOI: 10.1145/243199.243212
- Kimura, F., A. Maeda, K. Hatano and J. Miyazaki, 2008. Cross-language information retrieval by domain restriction using web directory structure. Proceedings of the Annual Hawaii International Conference on System Sciences, Jun. 7-10, IEEE Xplore Press, Waikoloa, HI., pp: 135-135. DOI: 10.1109/HICSS.2008.108
- Lilleng, J. and S.L. Tomassen, 2007. Cross-lingual information retrieval by feature vectors. Lect. Not. Comput. Sci., 4592: 229-239. DOI: 10.1007/978-3-540-73351-5_20
- Liu, Y., R. Jin and J.Y. Chai, 2005. A maximum coherence model for dictionary-based cross-language information retrieval. Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Aug. 15-19, ACM Press, Salvador, Brazil, pp: 536-543. DOI: 10.1145/1076034.1076125
- Monz, C. and B.J. Dorr, 2005. Iterative translation disambiguation for cross-language information retrieval. Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development In Information Retrieval, Aug. 15-19, ACM Press, Salvador, Brazil, pp: 520-527. DOI: 10.1145/1076034.1076123
- Muller, C. and I. Gurevych, 2008. Using wikipedia and wiktionary in domain-specific information retrieval. Proceedings of the 9th Cross-Language Evaluation forum Conference on Evaluating Systems for Multilingual and Multimodal Information Access, (CLEF' 08), Springer-Verlag Berlin, Heidelberg, pp: 219-226.
- Pirkola, A., 1998. The Effects of query structure and dictionary setups in dictionary-based cross-language information retrieval. Proceedings of the 21th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Aug. 24-28, ACM Press, Melbourne, Australia, pp: 55-63. DOI: 10.1145/290941.290957
- Polpinij, J., 2009. An ontology-based text processing approach for simplifying ambiguity of requirement specifications. Proceedings of the IEEE Asia-Pacific Services Computing Conference, Dec. 7-11, IEEE Xplore Press, Singapore, pp: 219-226. DOI: 10.1109/APSCC.2009.5394119
- Pourmahmoud, S. and M. Shamsfard, 2008. Semantic cross-lingual information retrieval. Proceedings of the 23rd International Symposium on Computer and Information Sciences, Oct. 27-29, IEEE Xplore Press, Istanbul, pp: 1-4. DOI: 10.1109/ISCIS.2008.4717868
- Rais, N.H., M.T. Abdullah and R.A. Kadir, 2011. Bilingual dictionary approach for malay-english cross-language information retrieval. J. Commun. Comput., 8: 354-360.
- Saralegi, X. and M.L.D. Lacalle, 2009. Comparing different approaches to treat translation ambiguity in CLIR: Structured queries vs. target co-occurrence based selection. Proceedings of the 20th International Workshop on Database and Expert Systems Applications, Aug. 31-Sept. 04, IEEE Xplore Press, Linz, pp: 398-404. DOI: 10.1109/DEXA.2009.58
- Sheridan, P. and J.P. Ballerini, 1996. Experiments in multilingual information retrieval using the spider system. Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Aug. 18-22, ACM Press, Zurich, Switzerland, pp: 58-65. DOI: 10.1145/243199.243213
- Tomassen, S.L., J.A. Gulla and D. Strasunskas, 2006. Document space adapted ontology: Application in query enrichment. Proceedings of the 11th International Conference on Applications of Natural Language to Information Systems, May 31-Jun. 02, Springer Berlin Heidelberg, Klagenfurt, Austria, pp: 46-57. DOI: 10.1007/11765448_5