

Query Translation and Quran Result in TreeMap

*Mohd Amin Mohd Yunus*¹, *Aida Mustapha*, and *Noor Azah Samsudin*

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400 Batu Pahat, Johor, Malaysia

Abstract. This paper presents the view of TreeMap representation with retrieval results more attractive on eyes. It is also found that TreeMap more friendly to present Quran results for keyword (K) and queryword (Q). This research also is being applied in retrieving Malay verses Quran translation.

1 Introduction

Cross Language Information Retrieval (CLIR) systems results can be represented automatically with the use of TreeMap (University of Maryland, 2000). This gives rise to the following approach to present the result in CLIR. Start with queries, documents, and relevance judgments with result visualization representation in a single language. In order to come out with TreeMap, the queries have been translated into another language by translation dictionary. These translated queries are helpful for the performance of the CLIR system. In order to conduct the experiments, data regarding the Malay and English Quran translation are gathered accordingly as two sets of collection respectively. The Malay and English dictionary files are useful for translating the query. A Malay and English stopword files are used for removing the unnecessary words in the query.

2 Related Work

Information visualization is a branch of computer graphics and user interface which are concerned with the presentation of interactive or animated digital images to users to understand data. For example, scientists interpret potentially huge quantities of laboratory or simulation data or the results from sensors out in the field to aid reasoning, hypothesis building and cognition. The field of data mining offers many abstract visualizations related to these visualization types. They are active research areas, drawing on theory in information graphics, computer graphics, human-computer interaction and cognitive science. This is quite similar to this paper in which the use of TreeMap is presented and discussed. The flexibility of their system to visualize different terms will be discussed. The development of the system involves two stages. The first stage is the processing of 6236 documents of the Malay translated Quran to create a database of all terms. The second stage is the creation of the web-based system using the preprocessing data created in the first

¹ Corresponding author: aminy@uthm.edu.my

stage. Both stages are explored. The visualization will help to improve the understanding of relationship between the terms in this specific domain [3].

The tools of computer interface design are familiar to most computer users today such as windows, menus, icons, dialog boxes, and so on. These make use of bitmapped display and computer graphics to provide a more accessible interface than command-line-based displays. A less familiar but growing area is that of information visualization, which attempts to provide visual depictions of very large information spaces. TreeMap aids in this process by providing an abstract visual representation of the compressors of interest in a single screen display that can be grouped and arranged by geographic region, asset, field or platform. Each individual compressor is represented in the TreeMap by rectangular area whose characteristics are mapped to attributes of the compressors chosen by engineer [5].

3 Experimental Approach

This paper focus on the testing Malay query words are taken from Fatimah's collection as natural language queries [1,2]. Fatimah has obtained them by considering several guidelines. Each query would be separated and broken into keywords and replaced by target language. For example, if query is Malay, so it is called as source language and the target language is English or Arabic. The translation refers to the same index between Malay natural query language [1] and the translation of the natural query language in English. When the keyword is Malay, then reference is to the English word at the same index or when the keyword is English and then reference is to the Malay word at the same index. It is considering word by word in the text files. The query can be categorized into two which are keyword and queryword. If the query is keywords, the results retrieved according to word by word results and redundant document names existed if merged. But querywords, retrieved according to the whole words as one at all and only when no redundant or unique document names retrieved rankly. Query translation can replace the origin query in to another language of the query. This translation is most important for those languages to investigate those information retrieval results. Figure 1 explains the workflow of cross language information retrieval based on query translation in the retrieval process of Quranic documents collection. Query is processed by removing the meaningless words or stopwords. Then, query can be translated into another language if needed by the dictionary. Next step is to stem the query by a stemmer if needed according to accurate language.

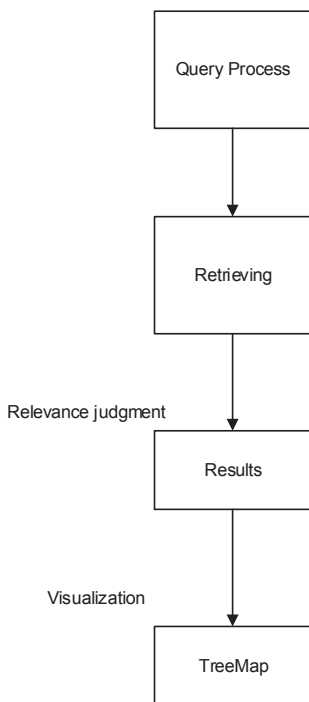


Fig. 1. Workflow of query translation retrieval and visualization results.

The query results consist of word by word result or keyword or the whole words or full phrase. The documents retrieval comes from al-Quran documents translation collection which is from Arabic, Malay and English collection. For this research, when the query is Malay words, the words are translated into English or Arabic vice versa. TreeMap display rows of data as groups of rectangles that can be arranged, sized and colored to graphically reveal underlying data patterns [4]. Thus total retrieve and relevant can be presented in terms of Surah names and document sentences or ayats or verses.

4 Result and Discussion

All results of the query(ies) translation are referred to the natural language queries of Malay [1] and then translated into Arabic and English queries in this study. In this paper actually concentrate on the comparison between translated query and normal query in retrieving documents by presenting TreeMap with K and Q accordingly as shown in Figure 3 and 4 based on Table 1 query and Figure 2 Normal Query Result.

Table 1. Natural language queries.

No.	Malay	English	Arabic
1	Kelahiran nabi isa	Prophet jesus birth	مولد النبي عيسى

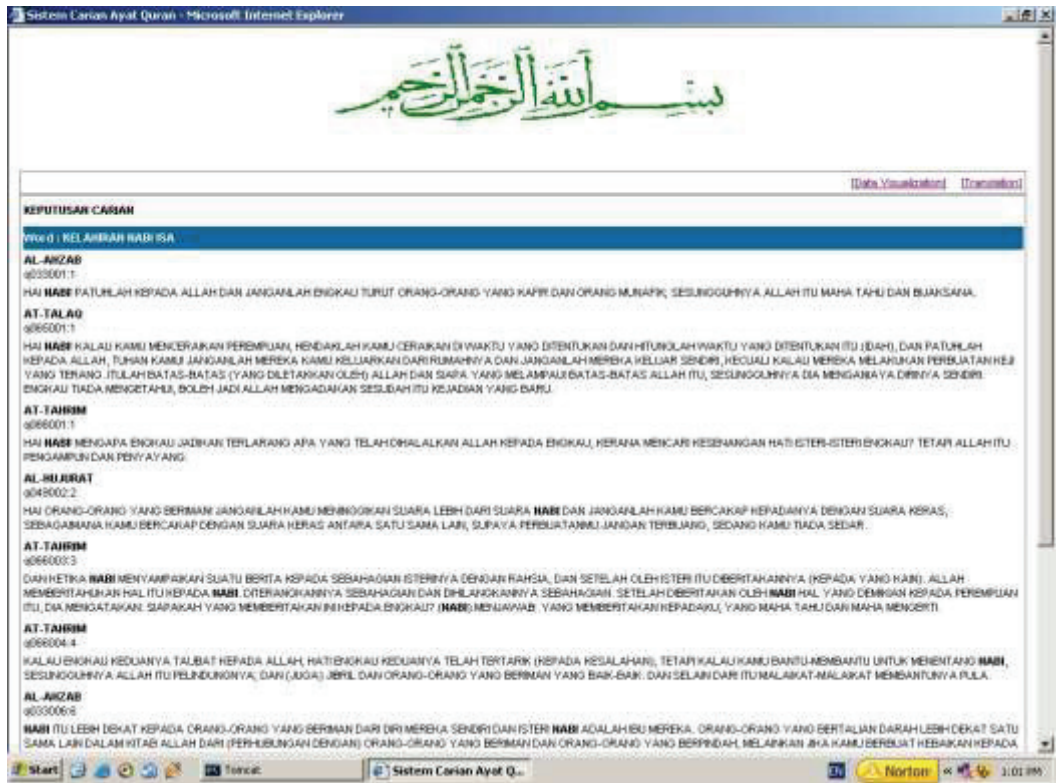


Fig. 2. Normal query result for “kelahiran Nabi Isa”.

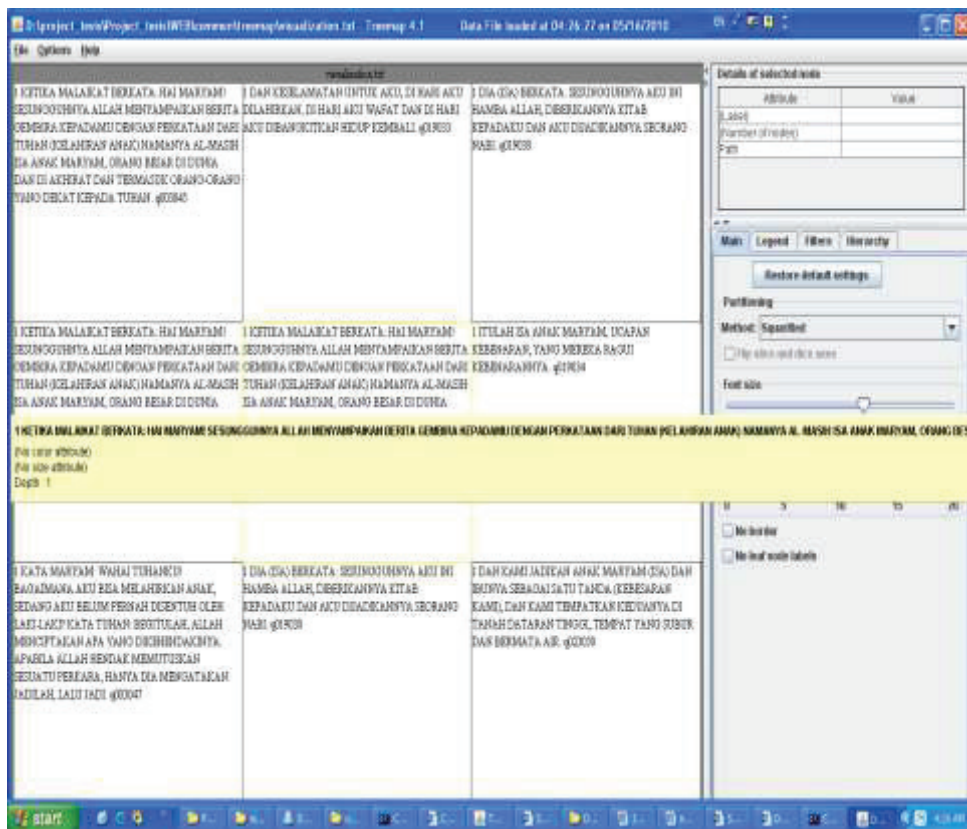


Fig. 3. Data visualization of K Result using TreeMap for “kelahiran Nabi Isa”.

1. A. Fatimah, A Malay Language Document Retrieval System: An Experiment Approach And Analysis, Tesis Ijazah Doktor Falsafah Universiti, Kebangsaan Malaysia (1995)
2. A. M. Y. Mohd, Short Query Translation: A Dictionary-Based Approach to Cross Language Information Retrieval. Tesis Master of Computer Science, Universiti Teknologi MARA. Malaysia (2008)
3. K. I. Normaly, A. R. Nurazzah, A. B. Zainab, Terms Visualization for Malay Translated Quran Documents, *Proceedings of the International Conference on Electrical Engineering and Informatics*, Institut Teknologi Bandung, Indonesia, 17-19 (2007)
4. University of Maryland, TreeMap. www.cs.umd.edu/hcil/treemap (2000)
5. C. Plaisant, G. Chintalapani, C. Lukehart, D. Schiro, J. Ryan, Using Visualization Tools to Gain Insight Into Your Data, The SPE Annual Technical Conference and Exhibition. Colorado: Denver (2003)