An Improved Searching Method for Retrieving Local Knowledge Information

Wong Poh Lee[1], Mohd Azam Osman[1], Tarmiji Masron[2], and Abdullah Zawawi Talib[1]

[1]Universiti Sains Malaysia, Malaysia, wongpohlee@hotmail.com, {azam; azht}@usm.my
[2]Universiti Malaysia Sarawak, Malaysia, mtarmiji@unimas.my

ABSTRACT

Local knowledge information has been widely available for years. These information mainly resides in the form of paper or media which are mainly stored in archives for researchers. When a user tries to retrieve these information, a thorough search has to be done manually to locate specific search query from a user. A lot of these data are not stored digitally and mostly recordered in unstructured form due to the complexity of the data collection. Besides, the implementation of a search method for locating specific information is difficult due to the unstructured form of the data. In this paper, an improved search method which combines the search by selection and search by keywords methods is proposed for retrieving data which can be applied to both structured and unstructured data. The results of the search can be represented in the form of text and graphical charts. The case study presented in this paper is limited to the local knowledge information collected in several states in Malaysia.

Keywords: information retrieval, local knowledge, structured data.

I INTRODUCTION

Nowadays, having and managing a lot of information, whether they are in structured or unstructured forms is common among the researchers and the general public. These information grow day by day. Some of the information are stored digitally while some of them are stored in a traditional manner that is in the forms of document, video or journal. These information will be more meaningful and useful to the users if they are presented in a proper representation or in a well-organized documentation.

The problem of the existing searching methods is that the search engine is too general which would normally returns a lot of information regardless of whether they are relevant or irrelevant to the users. Furthermore, it complicates the process of retrieving the required information. A good search engine should minimize these problems and provide the best results to the users.

In this paper, a local knowledge repository system consisting of local knowledge information in Malaysia is taken as a case study. This system includes a smartly developed searching method which returns the relevant results to the users' query by combining the search by selection and search by keywords searching methods. The results of the improved searching method will be evaluated by comparing it with the original searching methods based on the number of relevant results.

II BACKGROUND

As data are becoming larger and larger, an optimized search method is required to provide the most relevant information to the users. Therefore, a thorough review has been carried out to identify the current search method which is commonly used in retrieving information from databases.

A. Data in Local Knowledge Systems

Local knowledge information appears in many forms ranging from documents and images to videos. Local knowledge contains useful information which could be used as a reference for research or a guide for future development or advancement. A lot of studies have been carried out on local knowledge in different countries (Benferhat & Garcia, 1996; Duan, Subramanian, & Abdulrahman, 2013). Each of the collected data is managed differently depending on the structure of the data.

As the number of local knowledge data increases, the need for a proper system is required to provide a hub for knowledge sharing and dissemination of information to the users. van Splunter, van Langen, & Brazier (2005) proposed the use of templates to represent the knowledge and to use the web service technology as a technique to communicate with the web. Based on literatures, there are a lot of local knowledge data which focus largely on health and medical records. These data are further digitized to provide knowledge for the health centers such as knowledge on reducing sickness and improving a living lifestyle (Damtew, 2010).

B. Existing Search Methods

There are a lot of existing searching methods that provide the best web contents to the users (Killoran, 2013; Kawano, 2000). Different techniques were proposed to increase the search performance while providing the most accurate results for the users.

Khan, Sangroha, Ahmad, & Rahman (2014) proposed a semantic approach that is carried out by understanding the term instead of comparing the keywords. Tumer, Shah & Bitirim (2009) have investigated the semantic approach performance by comparing both the keyword-based search and the semantic search approach.

Searching methods have become an important factor especially in the medical field (Inthiran, Alhashmi, & Ahmed, 2010). The structure of medical data varies which makes it challenging to provide the most feasible search engine. The job of the search engine is to provide the most relevant results based on user queries. Lovic, Lu, & Zhang (2006) proposed the use of rules in existing search engine. It fixes the flaws of having a lot of irrelevant results while improving the performance of the searching process.

III IMPROVED SEARCHING METHOD

The searching methods deliberated in this paper are search by selection, search by keywords and the proposed improved searching method that combines both search by selection and search by keywords methods. In this paper, we respectively name the methods as follows:

- Searching Method I: Searching by selection based on the given choices.
- Searching Method II: Searching by keywords in which the users are able to enter the keywords of their choice.
- Searching Method III: Searching using a combination of the above-mentioned Searching Method I and Searching Method II

Searching Method I is a pre-defined selection which smartly extracts data from the database for the users to choose. The selection bar will not include choices which do not appear in the database. This is to keep the choices as simple as possible while retaining the most relevant results for the users.

Searching Method II is an advanced algorithm which looks for words residing in the local knowledge information. The words in the context is highlighted with a distinctive colour to denote the existence of the word in the context. The pseudocode of Searching Method II is as follows:

Seaching Method II - Keyword Search
Get keywords
Split keywords if keywords contains more than one word
Search through the context by category to determine whether the keyword exists
Replace keyword with a distinctive color style
Repeat process for the next keyword
Display results in a tabular form separated by category

Searching Method III which is combination of both Searching Method I and Searching Method II limits the search results by considering the choices chosen by the user so that the results would be more specific and relevant to the user's needs.

IV RESULTS AND DISCUSSION

An experiment was carried out to evaluate the practicality of the improved searching methods. For Searching Method I, selections from random choices were carried out to evaluate the accuracy of the results. For Searching Method II, keywords were used by entering random texts of different length. For Searching Method III, choices were selected from the given selection boxes and some keywords were entered at the same time. The final results consist of the number of relevant results over the total results. The accuracy (in percentage) is calculated using the following formula:

$$Accuracy = \frac{Relevant\ Results}{Total\ Results}\ x\ 100$$

Table 1 and Table 2 show the results for Searching Method I and Searching Method II respectively. Searching Method I is suitable for structured data while Searching Method II is suitable for unstructured data. Searching Method I is suitable for structured data because this method can easily match and retrieve data from the structured tables in the database. Searching Method II is suitable for unstructured data is due its capability which allows the search is being done by the entered keywords and match with the content in the database. By introducing Searching Method III (combining both Searching Method I and Searching Method II), the data can be searched simultaneously by considering the parameters from the structured data and, unstructured data. Table 3 shows the result for Searching Method III.

Table 1. Results of Searching Method I

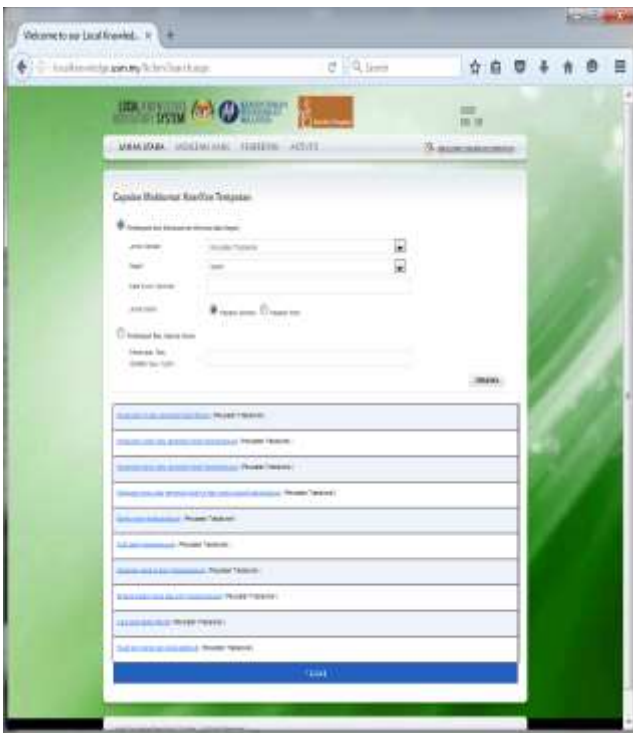| Number of selections	Relevant Results	Total Results	Accuracy
20	153	153	100%

Table 2. Results of Searching Method II

Text length	Relevant Results	Total Results	Accuracy
2 (e.g. in)	Not applicable	160	-
3 (e.g.dua)	4	8	50%
4 (e.g.ubat)	9	10	90%
5 (e.g. batik)	57	74	77%
6 (e.g. tarian)	1	1	100%

Table 3. Results of Searching Method III

Text length	Relevant Results	Total Results	Accuracy
2 (e.g. in)	0	0	100%
3 (e.g. dua)	7	8	87.5%
4 (e.g. ubat)	9	9	100%
5 (e.g. batik)	67	67	100%
6 (e.g. tarian)	1	1	100%

Based on the result in Table 1, there is no issue as far as Searching Method I is concerned. However, in Searching Method II, there are issues if short texts are used. It is found that the search engine takes into account, words which contain the search string of the keyword. Short words such as "is", "in" or "to" commonly appear in longer words, and they lead to more search results which may not be relevant to the users. Searching Method III reduces the irrelevant results as the system identifies a more specific user query from the selection menu. For example, figure 1 and figure 2 show the different result returned by the search engine. Figure 2, which is using the Searching Method III return a more specific result.

Figure 2. Search Result Using Method III

Figures 3-6 show some sample screenshots of the web application. Figure 1 shows the interface for selecting the type of search which the user desires. Figure 2 shows the search results which are displayed in the form of a map. The user is also able to display the results in tabular form as shown in Figure 3. Figure 4 shows the search results based on keywords. The keywords are highlighted to ease the user to look at the words which appears in the search results. An option to suggest keywords are also included to narrow down the search results.

Figure 1. Search Result Using Method I

Figure 3. Search Menu

Figure 4. Search Results in Map Form

Figure 5. Search Results in Tabular Form

Figure 6. Search Results with Highlighted Keywords in Tabular Form

V   CONCLUSION AND FUTURE WORK

In conclusion, the improved searching method (Searaching Method III) is able to provide an effective search feature for the users to query local knowledge information. By providing different searching methods such as search by selection method, search by keywords or a combination of both (the proposed searching method), the users are able to search the most relevant information based on their requirements.

For future work, images reflecting the local knowledge will be considered. Each image will be characterized and indexed into various categories which allows the users to search for images from the keywords. In addition, the searching method will be further enhanced by using the natural language processing technique for more accurate search results.

REFERENCES

Benferhat, S., & Garcia, L. (1996). A local approach to reasoning with conditional knowledge bases. *Proceedings Eighth IEEE International Conference on Tools with Artificial Intelligence, 1996 (pp.404-407).* France: IEEE.

Damtew, Z. (2010). Using local knowledge in information gathering process: Case study from Ethiopia. *Proceedings of the IST-Africa 2010* (pp. 1-7). Durban: IEEE.

Duan, J., Subramanian, N., & Abdulrahman, M. (2013). Knowledge management in the Chinese local beer market: A case study. *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, (pp. 58-62). Bangkok.

Inthiran, A., Alhashmi, S., & Ahmed, P. (2010). A Reflection of Current Search Engine Techniques on Medical Search Environments. *Sixth International Conference on Intelligent Environments (IE), 2010* (pp. 170-173). Kuala Lumpur: IEEE.

Kawano, H. (2000). Overview of Mondou Web search engine using text mining and information visualizing technologies. *International Conference on Digital Libraries: Research and Practice* (pp. 234-244). Kyoto: IEEE.

Khan, J., Sangroha, D., Ahmad, M., & Rahman, M. (2014). A performance evaluation of semantic based search engines and keyword based search engines. *International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)* (pp. 168-173). Greater Noida: IEEE.

Killoran, J. (2013). How to Use Search Engine Optimization Techniques to Increase Website Visibility. *IEEE Transactions on Professional Communication, 56*(1), 50-66.

Lovic, S., Lu, M., & Zhang, D. (2006). Enhancing Search Engine Performance using Expert Systems. *IEEE International Conference on Information Reuse and Integration* (pp. 567 - 572). Waikoloa Village, HI: IEEE.

Tumer, D., Shah, M., & Bitirim, Y. (2009). An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia. *Fourth*

International Conference on Internet Monitoring and Protection, 2009 (ICIMP'09) (pp. 51-55). Venice/Mestre: IEEE.

van Splunter, S., van Langen, P., & Brazier, F. (2005). The role of local knowledge in complex Web service reconfiguration. *The 2005*

IEEE/WIC/ACM International Conference on Web Intelligence (pp. 495-499). Netherlands: IEEE.