Basic Model For Information Retrieval Based On Free-Text In Database Application

Mohd Adan Omar[1], Md Zahir Mat Cha[2], Mohd Rushdi Idrus[3]

[1,2,3] Graduate Department of Information Technology
College Arts and Sciences, Universiti Utara Malaysia, 06010, UUM Sintok, Kedah
e-mail: adan@uum.edu.my, zahir@uum.edu.my, rushdi@uum.edu.my

ABSTRACT

The purpose of this article is to design the basic model for information retrieval based on free-text domain in a particular database environment. It is an alternative method to the usage of key that is widely used in today database applications. At the end of this paper, we show the prototype that has been developed to support the free-text model verification. The result from the testing process is also given.

Keywords
Information retrieval, Free-text, Basic Model.

1.0 INTRODUCTION

Researchers in the field of information retrieval spend a lot of time in finding the technique to optimize the abilities of index and tool in retrieving information needed in the shortest period. Several techniques used such as using algorithm, key word and structured query language. Research toward information retrieval in database system focuses to the usage of the key as retrieval sources. As alternative to the key use, information retrieval using free-text domain in database environment is proposed. The literature review shows, several theories relate to the usage of free-text as information retrieval domain. However, most of the views discussed the retrieval in Internet environment.

Until the research conducted, there is no model for information retrieval based on free-text in a database application except Internet based information retrieval. So, the research focuses to the technique of data retrieval using free-text without link to any key attribute as primary key in database application environment. Most of the data search in database application individually or multi-user environment will link to the key as main field. This approach, sometime gives the implication that brings to system halt, slow and fail to search information needed in certain period.

So the objective of this research is to develop and propose the model of information retrieval based on free-text domain in database application. The model proposed will help in providing an alternative to data retrieval process using free-text without tied to certain field. Based to free-text as

alternative key, the process of information retrieval directly browses the data table without rely on unique prime key.

2.0 INFORMATION RETRIEVAL

Information retrieval is an approach to allow user makes item searching that knows already exist in a specific subject. The ability to search information toward specific topic is important part of information retrieval system (Brinkley and Burke, 1995). There are two early models of information retrieval which are query based information retrieval model and browsing information retrieval model (Cox, 1994). Information retrieval model using query is based on classical model as shows in Figure 1 (Belkin and Croft, 1987; Cox, 1994).
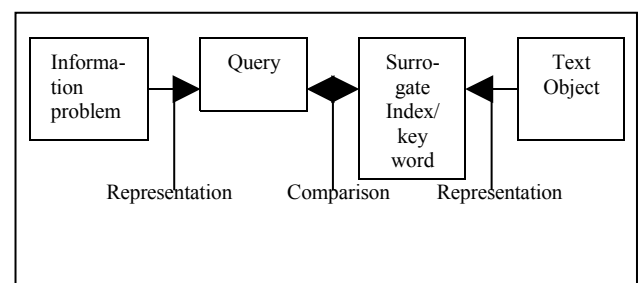

Figure 1: Query Model for Information Retrieval
Sources: (Belkin and Croft, 1987; Cox, 1994)

In this model, index and key word was used. Researches after this time use this model as reference model as Bate (1989) and Cooper (1988) concluded that, although this model work well and can be suit with current technology, but lot of problem existed in it application such as, difficult to formulate query and a variety of important query components and also term dependency. So that Gorayska and Cox (1994) added a system based on query formula contains limitation which leads to mismatch between human and information processing.

Cox (1994) and Al-Hawamdeh et al. (1991) in their study proposed browsing paradigm for various task in information retrieval because of its effectiveness and

efficiency. In browsing model, the usage of database is independent where user does not need to know the query formula and will move to whole of database. The model proposed by Cox (1994), provided the interactive searching activities and the system user determined the search direction based on system respond. However, both models show negative usage of free-text in their information retrieval system but the original concept of free-text information retrieval is suit with browsing base information retrieval.

3.0 FREE-TEXT INFORMATION RETRIEVAL

The method of information retrieval using full-text is normal several decades ago. At the beginning, the method was using key word which refers to the information wants to retrieve. The latest approach was using word indexes which record the information position in database. The list of indexes allowing retrieval process toward all data documents rely on matched query and index (Wikipedia, 2005). It is clear that, key words, indexes and queries was becoming the key component in this method. The next version of method was trying to trim the retrieval component usage such as key word, index and query. Chisholm and Koida (1999) for example proposed Vector Space Method to look the frequency of word in a query. Chávez-Aragón et. al. (2003) developed Novel Hybrid Method to strengthen the index ability until object image.

Wen (2003) said, information retrieval based on text in web environment using search engine still used key word and index as reference resource. Most of search engines in web environment are based on Boolean Retrieval Model where the user needed to determine a query as Boolean combination for search word (Jones and Willet, 1997). The major problem in this method was the difficulty in formulating the query. So researcher try to replace to new method called information retrieval based on concept (Alsaffar et al., 1999). The study about the system of information retrieval based on concept was introduced by McCune et al. (1985) through the system called Information Retrieval Based on Rule. Fanguay and Raghavan (2003) introduced information retrieval based on concept but the definition of concept was generated by decision tree and allowed user to choose the definition fixed by expert.

Searching system using free-text allows user to find and retrieve the information (document) by using word combination while the system uses key word need index to scan all document in database. In the latest system, user will retrieve the information using controlled word combination as predicted by Blair and Marron (1985). Retrieving the information using free-text means in the database, user will key in the relevant word without specifying the location of word fields (etde.org, 1999). Zimmerman (1990) concluded that, the main key of free-text in database is it ability to browse fast through many items. The information retrieval based on free-text proposed in this study is specific to database application

with the use of idea in document environment and also web search engine.

4.0 THE MODEL

The literature review shows that free-text mainly used in information retrieval to get the document such as conference proceeding, journal and article in library environment. There is no (during research period) free-text in database environment developed by tables. However, an SQL still used in this model, but combine together with algorithm to allow free-text used as searching subject.

The approach for interface is similar (in concept) with practiced in search engine where word/s can be typed in searching space and result will be displayed orderly. We adopt this approach into the process of information retrieving using free-text. A space to fill word on a form is provided and search manager will do the browsing process into database tables. If in search engine, user can choose the information needed by clicking the link, free-text environment also allows the same process.
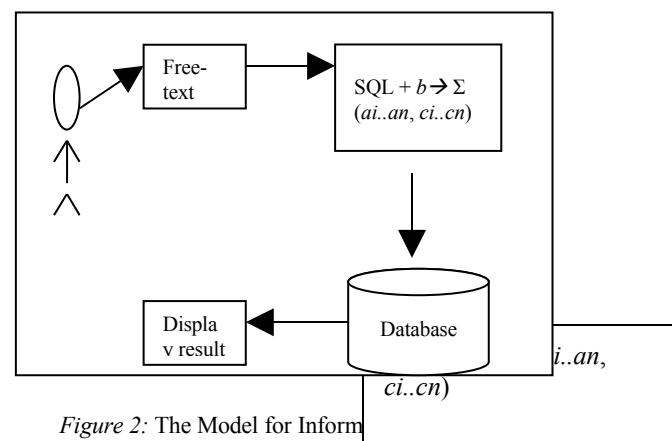

Figure 2: The Model for Inform
Based on Free-Text

Fig 2 shows the model for information retrieval based on free-text in database environment. Free-text box refer to the process text entry for searching purposes. SQL still be used together with retrieve algorithm where b is referred as text entered bye the user. Searching system will browse through tables and cells in database is showed by $ai..an, ci..cn$. Retrieval will be made if matched text occurs. The result will be displayed in order form and user can click the order for specific information.

5.0 THE PROTOTYPE

To prove the model produced in this research, we develop a prototype which focuses on free-text search module. The prototype developed together with database for Salary System. We used Microsoft Visual Basic 6.0 and Microsoft Access as development tool. Fig 3 shows the prototype's search menu module interface.
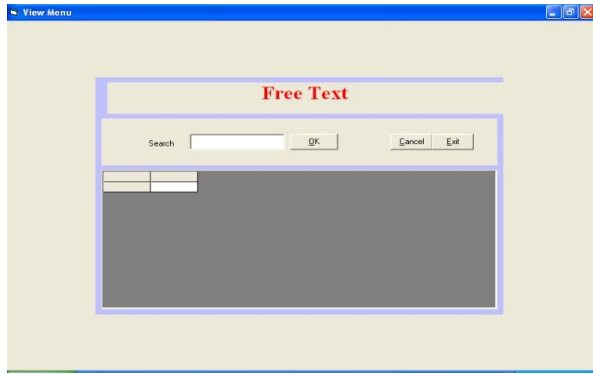
Figure 3: Search Menu Module Interface

6.0 THE TEST RESULT

The purpose of testing process is to make sure the prototype work properly and support the model of information retrieval based on free-text as proposed. The Testing is also to see whether the developed prototype is able to generate search results as describe in the model. We do experiment by entering text in free-text space on the form and record the result. The example of test experiment shows in Fig 4. The result of selected experiments is show in Table 1.
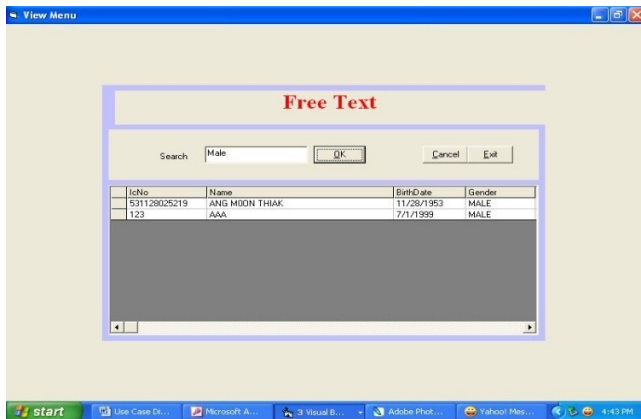
"Male" text has been entered into search space. Search result shows two rows which contain word "Male" in

Figure 4: Example of test experiment

Table 1: *Prototype Testing Experiment Result*

Testing	Expected Result	Actual Result	Comment
User click search menu	Form will be displayed	Yes	
User enters the word	Expected result will be displayed	Actual result displayed	

"Male"	Yes	Yes	
"1000"	Yes	Yes	
"100"	No	No	
"Ang"	Yes	Yes	
"Sintok"	Yes	No	No data
"Johor"	Yes	No	No data
"10"	No	No	

7.0 CONCLUSION

In general, the objective of the research is to develop the basic model for information retrieval based on free-text in database environment. After we develop the prototype to prove the model and, we do the testing experiment to validate the prototype and the model. However several aspects do not touched in this research such as the object of image instead of text. The developed model is able to improve the aspect of information retrieval and help in database system development.

REFERENCES

Al-Hawamdeh, S., de Ver, R., Smith, G. & Willet, P. (1991). Using Nearest Neighbor Searching Technique to Access Full-Text Document. *Online Review,* Vol. 15(3/4), pg. 173-191.

Alsaffar, A., Deogun, J., Raghavan, V. & Traylor, D. (1999). Enhancing Internet Search Engines to Achieve Concept Based Retrieval. *Forum 1999-Improving the visibility of R & D Information.* Oak Ridge.

Bates, M.J. (1989). The design of Browsing and Berrypicking Technique for the Online search Interface. *Online Review*, Vol. 13(5), pg 407-424.

Belkin, N.J. & Croft, W.B. (1987). Retrieval technique. *Annual Review of Information Science and Technology* (ARIST), Vol. 22, pg 109-489.

Blair, D. C. & Marron, M. E. (1985). An Evaluation of Retrieval effectiveness for a Full Text Document Retrieval System. *Communication ACM.* Available: http://portal .acm.org.

Brinkley, M. & Burke, M. (1995). Information Retrieval from Internet: An Evaluation of the tools. *Internet research: Electronic Networking Applications and Policy*, Vol 5, No 3, pg 3-10. Available: http://www.emarald.com

Chávez-Aragón, A., Ma, A. A., Starostenko, O. & Zehe, A. (2003). Novel Hybrid Method for Image Retrieval by Ontological Description of Sub-regions.

Proceeding of Artificial Intelligence and Doft Computing 2003. Acta Press.

Cooper, W.S. (1998). Getting Beyond Boole. *Information Processing and Management*, Vol. 24(3), pg. 243-248.

Cox, K. (1994). *Information Retrieval by Browsing*. City Polytechnic of Hong Kong.

Etde.org. (1999). *Free-Text of Fields Searching*. www.etde.org.

Fanguay, R. & Raghavan, V. (2003). Generating Rule-Based Trees From Decision Trees for Concept-Based Information Retrieval. *Applications, Products and services of Web*.

Gorayska, B. & Cox, K.R. (1994). *Expert System as Extensions of the Human Mind in Artificial Intelligence and Society*.

Jones, K. & Willet, P. 1997. *Introduction to Chapter Five. Reading in Information Retrieval*. Morgan Kaufmann., San Fransisco.

McCune, B., Tong, R., Dean, J. & Shapiro, D. (1985). RUBRIC: A System for Rule-Based Information Retrieval. *IEEE Transactions on Software Engineering*, Vol. 11:2, pg 939-944.

Wen, YY. (2003). *A Search Engine for Image. Project Report*. Faculty of Computer Science, University of Maine.

Wikipedia. (2005). *Keyword Searching*. www.wikipedia.org

Zimmermann, M. (1990). *Notes of Free-Text Information Retrieval*. www.his.com