Association Sorting Algorithm Design for Error Searching System

Fan Yang, Zhenghong Dong and Mengwei Li

Academy of Equipment, Beijing, China

sarah0824@hotmail.com

Abstract. For the searching results of error searching system for Integrated Decision Information System (IDIS), this paper proposed a method to carry out association sorting according to degree of association of searched results by keywords, which can priorily recommend user interested searching results. Errors of IDIS platform are occurred very often. Because those errors belong to different stages like setup, configuration, and operation, or those errors may occurred in different services, applications, or IP ports, or may be happened in different system software, different version of software, and those errors are also can be classified into different types. As there will be many error information are searched out, it is required to sort them according to degree of association, and provide user interested results. This paper proposed a method to carry out association sorting for the searching results according to the times of searched repeated results by keywords, which effectively put results with high degree of association in the front and increase the searching efficiency.

1. Introduction

IDIS integrates a lot of sub-system from different institutes, including almost a hundred different software. Because of the differences of underlayer protocol, and the difference of standard, there are many errors occurred during the stages of setup, configuration, and operation, which seriously affects the usage. Moreover, because the errors are various, which may happen in different operation phases, stages, TCP/IP communication protocol layers, sub-system software, it is necessary to design a database system which can manage those errors. The proposed method provide a design of error searching database, which can make the correspondence among error, the reason of the error, and corresponding solution, and put them to different categories in terms of their characteristics, such that it is easy to manage, search, and use. Users do not only search the reason and solution of an error, but also can find out the accurate information of the error, such as in which stage, layer, software, or categories, etc. the error has happened.

The current error searching systems are various, including On-Board Diagnostics designed for vehicles [1], which can search vehicle errors according to vehicle OBD error code; Computer error searching system [2], which can search the hardware problems, network problems, and software problems, etc. Those error searching systems focus on errors in specific area, design different databases and the associations in between to store the errors as well as the characteristics of those errors. Because IDIS is used for communication and ordering between the superiors and the subordinates in specialized field, the design for DB tables, logical structures has to build up according the specialized characters of IDIS, and the design and definition for keywords requires personalized customization. In that case, the current error searching software cannot be applied to this platform. It is

necessary to design a specialized database system to effectively solve the various errors occurred in the IDIS.

For the searching results of error searching system for IDIS, this paper proposed a method to carry out association sorting for the searching results according to the times of searched repeated results by keywords, which effectively put results with high degree of association in the front and increase the searching efficiency.

For the arrangement of this paper, section 2 will discuss related work, analysis their advantages and disadvantages. Section 3 will introduce the proposed method, including the design on the system function, as well as the flowchart of this system. Section 4 provides the system prototype and section 5 makes the conclusion of this paper.

2. Related Work

The current sematic searching algorithms only focus on the association between information about the searching results. They only focus on the structure of information, which leads that user expected searching results have been put to the back of the results. Reference [3] has proposed a sorting algorithm based on genetic algorithm, which optimizes the sorting results.

A lot of work [4-8] focused on analyzing the data properties, such as the usage of tagging patterns, the semantics of tags, the usage of annotations, and other personalized user behavior. Reference [9] combines the contents of web pages and the social annotations as well as user's personalized preferences to search the result, which can be applied to social annotation information system. Reference [10] ranks search result based on user-label, which requires user feedbacks and user queries. However, this method requires large amount of data to make the ranking results reliable, otherwise, the ranking results are invalid if the sample data is limited. Apart from the content of search results, reference [11] also considers time dimension for ranking. Because people may not only considered about the relevance, but also would like to view the latest article. In this reference, new articles would be recommended to users. However, it is hard to define how well the time dimension would affect the ranking results.

Sematic relevance is well considered in current works for ranking search results, which is quite useful to find the relevance between user query and search results. This is the most important factor that controls the ranks of search results. So this paper focuses on sematic relevance to rank search results, which considered about the quality and quantity of user queried keywords that matches with the search results. And this way of rank is more reliable.

3. Proposed Method

3.1. System function design

Figure 1 shows the system function design. The error database system contains two modules, which are search engine and database, where database has three sub-function modules (D1, D2, and D3).

D1:   Data import/enter

This function supports two ways of importing data. One is to import by Excel file directly, and the other is to enter data by administrator.
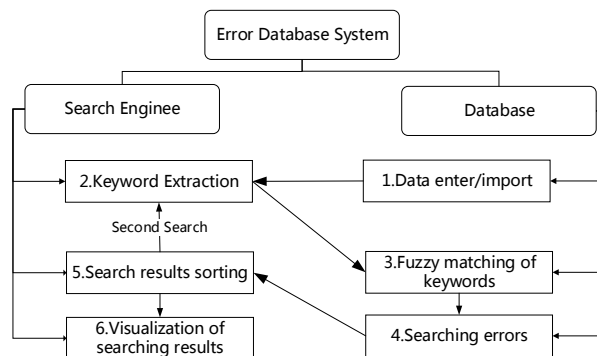
D2:   Keyword fuzzy search

Figure.1 System function diagram.

This function supports for the fuzzy matching between the extracted keywords and the keywords in keyword table.

D3:   Find out the corresponding error, reason, and solution according to the extracted keywords.

Search engine includes 4 sub-function modules (S1, S2, S3, and S4).

S1:   Keyword extraction

To extract useful keywords from the input contents.

S2:   Association Sorting for the searching results

Association Sorting Algorithm: to sort the searching results according to matching degree.

Degree of association sorting algorithm has been proposed to priorly recommend more matched results.

Considering that the researched error information contains multiple records, the proposed method priorly recommends more matched searching results according to sorting results based on the degree of association of keywords

First, to consider the errors information that contains filtering conditions (error characters), and put those errors to the first half of the searching results list. While for the searching results that are searched by keywords with priority equals to 2, we put them to the latter half of the searching results list.

Next, to sort the first half and the latter half of results list, respectively. According to the keywords to find out corresponding error ID from the table of keyword and error relation. Because an error has several keywords, so we can find out the same error ID many times according to those keywords. If an error ID has been searched out more times, then it means it contains more keywords, and it is more related to the searched error. Then we sort the searching results according to errors times that have been searched out, so that user interested results can be put to the front.

S3:   Second search

The system also supports second search from the already searched results.

S4:   Visualization of searching results

This is to display the search results visually.

3.2. Flowchart

Figure 2 gives the whole flow chart of the error searching system which contains 5 stages, which are keyword extraction, fuzzy matching, find out error ID, sorting search results, and visualization of searching results, respectively. As sorting process only evolves stage 4 and stage 5, so we explain the 2 stages in more detail as table 1.
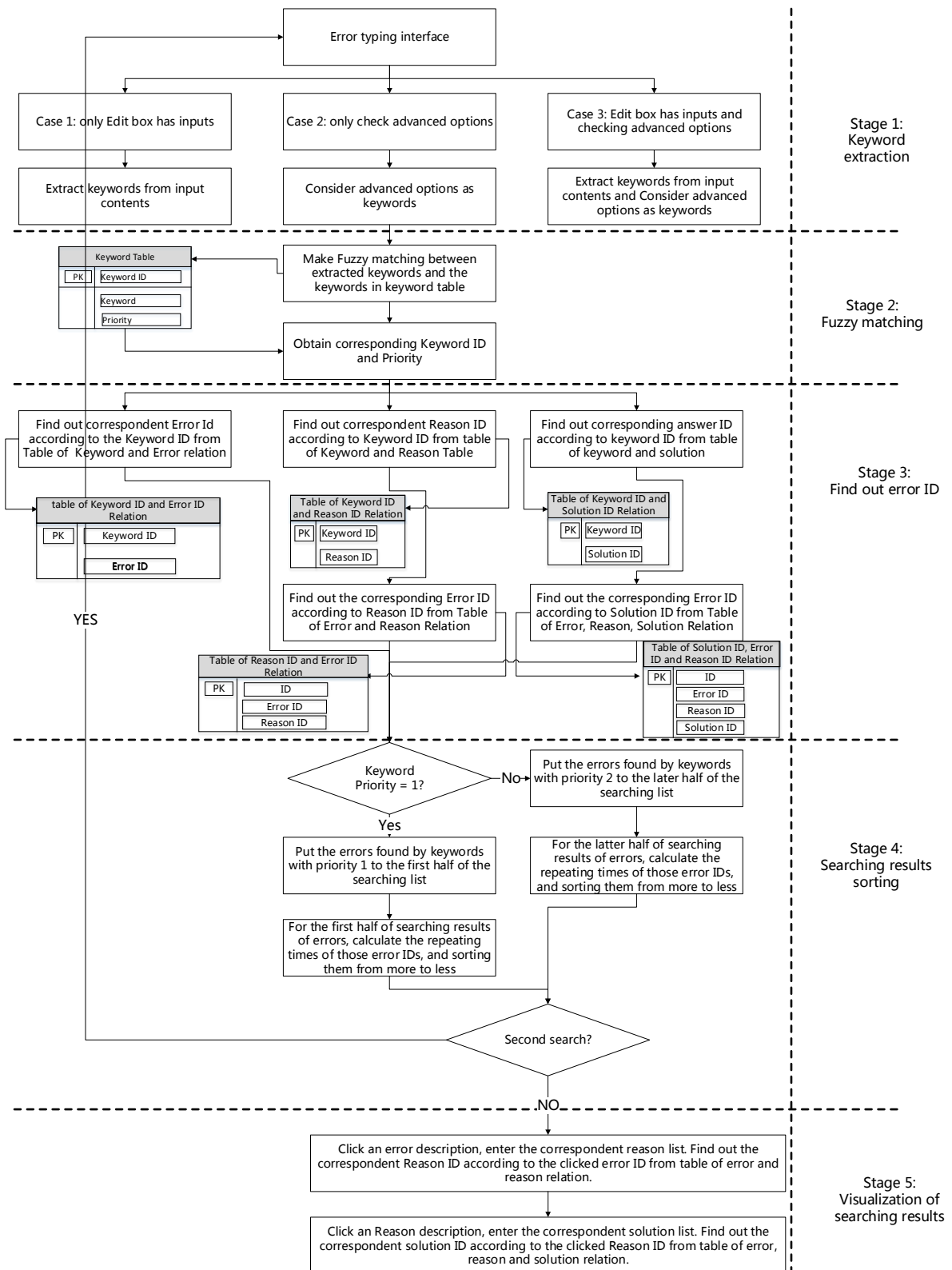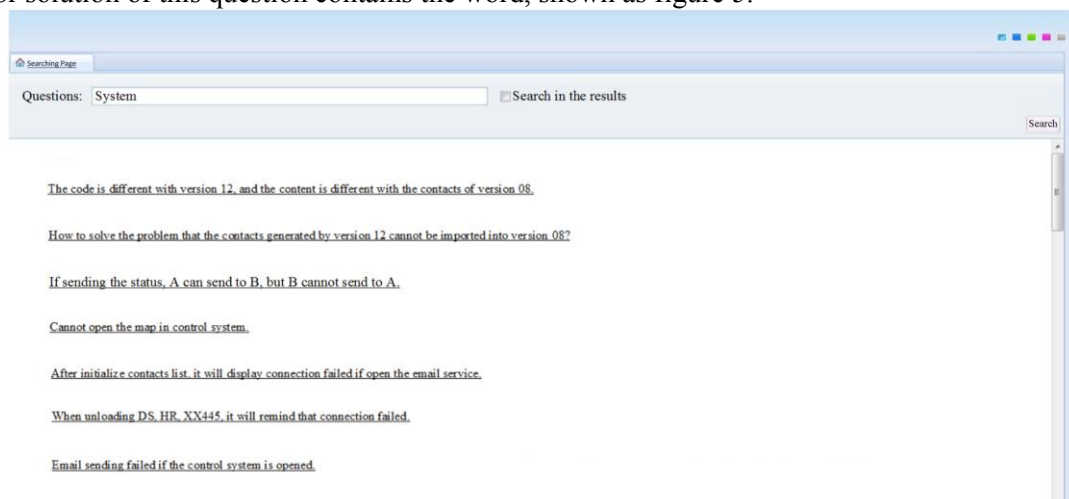
Figure 2 Flow chart of the error searching system.

Table. 1 Explanation on Database searching flowchart

Searching stages	Contents
Stage 4 Sorting up searching results	1. Find out a set of Errors by keyword with priority=1, and put them to the first half of searching result list. To compute the times that have been found out, and sort the set of errors according to the searched times from more to less, which is the first half of searching result list.
	2. Find out a set of Errors by keyword with priority=2, and put them to the latter half of searching result list. To compute the times that have been found out, and sort the set of errors according to the searched times from more to less, which is the latter half of searching result list.
	3. If there is no such result has been found according keyword database, then directly carry out fuzzy matching with the errors in the table of error.
	4. If user would like to narrow the searching scale, he can apply second search, the same searching method applies here.
Stage 5 Searching results visualization	1. After error list has been generated in terms of degree of association from high to low, according to error ID, to find out corresponding reasons from the table of error and reason relation, and generate reason list. The way of presentation is like this, click on an error description, the correspondent reason list is displayed.
	2. After reason list has been given, according to error ID and reason ID, to find out corresponding solutions from the table of error, reason, and solution relation, and generate solution list. The way of presentation is like this, click

4. Experiments and Prototype

This paper has implemented a prototype of this searching system for verifying if the sorting results are valid and useful. We have shown the prototype by figure 3. Figure 3 shows the searching results sorting page. After search for keyword "system", the searching results are list as figure 4. We can see from the figure, related items that relevant to "system" are put to the front of research results. And even if one item does not contain the exact word of "system", it still could be put to the front if the reason or solution of this question contains the word, shown as figure 5.
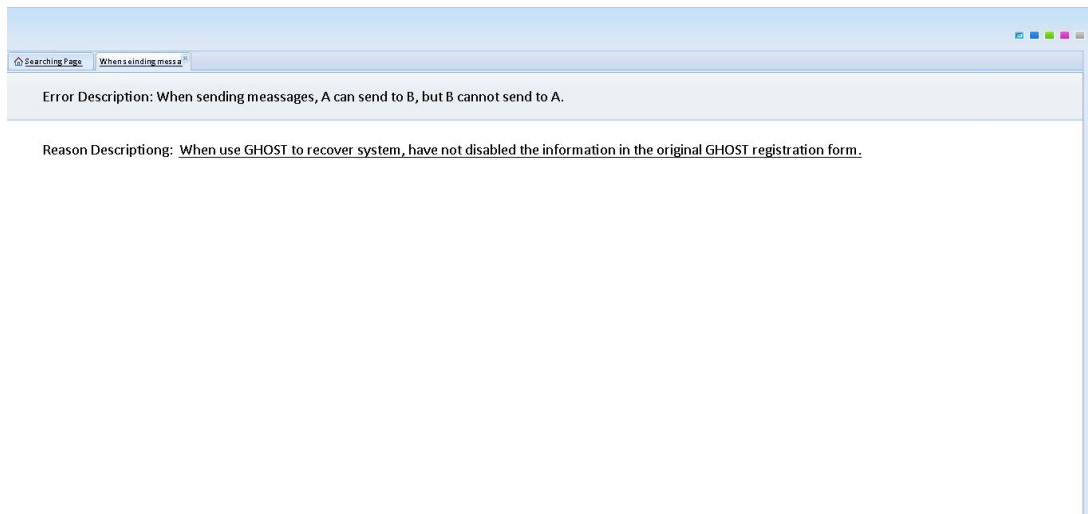

Figure 3 Search results sorting page

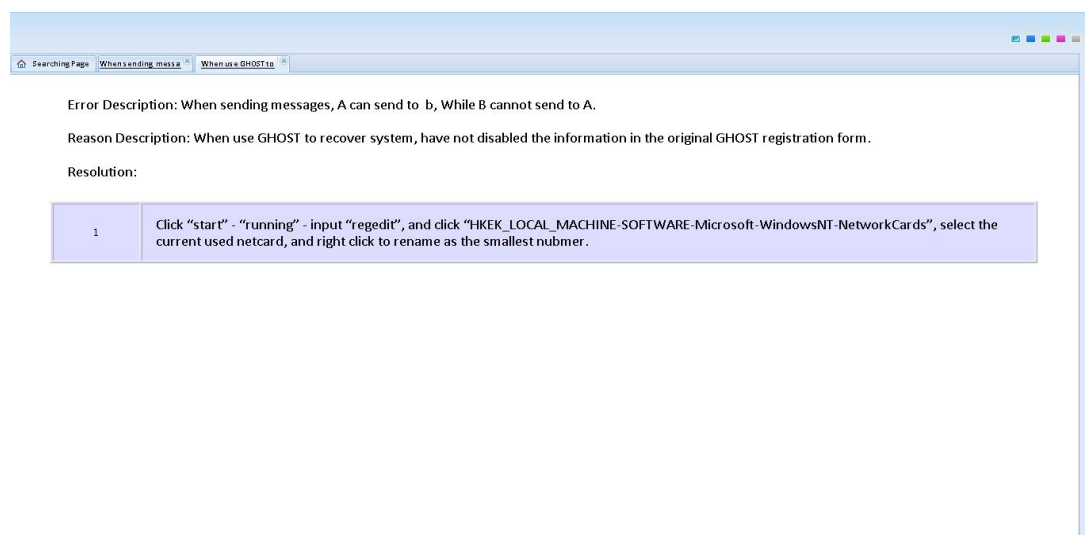**Figure 4** Page of reason that contains the word "system".

Figure 5. Page of solution that contains the word "system"

5. Conclusion

For the searching results of error searching system for IDIS, this paper proposed a method to carry out association sorting for the searching results according to the times of searched repeated results by keywords, which effectively put results with high degree of association in the front and increase the searching efficiency. We have provided a prototype to show the sorting results and explain the sorting idea.

References

[1]   Wang, J.H., Fang, M.D., Gao, J.D., Lu, H.Y. and Dai, C.B. Basic Principle and Application of On-Board Diagnostics for Gasoline Fuelled Vehicles. Automotive Engineering, 28, 491-494, 2006.

[2]   Lu, C., Yang, Y.-H. and Xu, G.-M. Exploitation of Computer Problem Repair and Require on Web System, Journal of Zhejiang Ocean University, 2008.

[3]   Yahui Wang, Sematic Searching results sorting algorithm based on Genetic

Algorithm, Journal of Jincheng Institute of Technology, 6(1), 2012.

[4]   S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. Journal of Information Science, pages 198–208, 2006.

[5]   X. Wu, L. Zhang, and Y. Yu. Exploring social annotations for the semantic web. In WWW '06: Proceedings of the 15th international conference on World Wide Web, pages 417–426, New York, NY, USA, 2006.

[6]   P. A. Dmitriev, N. Eiron, M. Fontoura, and E. Shekita. Using Annotations in Enterprise Search. In Proc. of WWW 2006, pp.811-817, 2006.

[7]   S. Xu, S. Bao, B. Fei, Z. Su and Y. Yu. Exploring Folksonomy for Personalized Search. In Proc. Of SIGIR 2008, pp. 155-162, 2008.

[8]   M. Noll and C. Meinel. Personalization 2.0: Web Search Personalization Via Social Bookmarking and Tagging. In Proc. Of ISWC 2007 + ASWC 2007, pp. 567-581, 2007.

[9]   C-C Zhao, Z-Q Zhang, H-L Li, and X-Q Xie, A search result ranking algorithm based on web pages and tags clustering, IEEE International Conference on Computer Science and Automation Engineering, 4, 2011.

[10]   Z-Q Zhang, T-T Liang, X-Q Xie, "An user-label based search result ranking algorithm", Journal of Computer research and development, 46(supply): 351-358, 2009.

[11]   Shu-Hua Chen, Han Yan, Jian-Cheng Li, Jun-Jie Wang, Yu-Ting Mao, Jun Liu, "Improvement of PageRank Algorithm: An Authoritative and Temporal Based Approach", IEEE conference on Anthology, 2013.