WEB PAGE RECOMMENDATION WITH KEYWORD BASED ON HIERARCHICAL MODEL USING ASSOCIATION LINK NETWORK (ALN)

Ms. Sonali Lingayat, Mr. Hirendra Hazare

MTech Student, Assit. Professor

Department of CSE, Ballarpur Institute of Technology (BIT), Ballarpur.

ABSTRACT

To establish an associated relation among various resources, an association link network is designed by extending the hyperlink network to an association-rich network. The resources are available on the Web in the form of loosely connected network of no semantics which get organized by association link network to provide web intelligent services such as web browsing, web knowledge discovery, publishing etc. This paper builds the multi-layer theory of association semantic that includes three-layer viz., Layer 1, Layer 2 and Layer 3. It uses semantic layered technology to provide theoretical support as well as to improve searching efficiency to search the related information. Experimental results show that information get searched efficiently with the recommendations for related topics as well as current topics. Association link network helps to connect the loosely connected data to an association-rich network to show the related information of a specified topic. The integration between association link network model and information resources provides a new way for organizing required information to the user with their semantics.

Keywords: Association link network, semantics, multilayer theory of association semantic, association, knowledge discovery.

I. INTRODUCTION

Web contains huge amount of information which is present in scattered form i.e. in loosely connected network. To convert loosely connected network into an association-rich network, an association link network (ALN) is designed for searching required information to the user. This paper explores the multi-layer theory of association semantic to provide theoretical support and recommend the related topic of the specified topic to the user for getting more informational by fulfilling the request of user. Semantic information includes different kinds of activity, which includes detecting different events or finding anomalous patterns related to a person or object. It also improves the searching efficiency. It recommends the user to search related topic of the specified topic as well as the current topic. Along with textual information of the topic, user can also view the image data as well as video data of that topic. The multi-layer theory of association semantic consists of three layers namely Layer 1 which contains the basic information, Layer 2 which contains the current information and Layer 3 which contains the long-term information. The experimental results show that the proposed system recommends the Webpages related with the specified keyword given by user as well as it recommends the current topic and long-term topic for the user.

II. RELATED WORK

A huge amount of information is present related with every domain or field. It is quite difficult to extract the required information among the available resources. There are various techniques available to extract the exact information from the pool of data, among them some related work is given here. S.X. Zhang, K. Lu, W. Liu, and G. Zhu[1], have proposed the multi-layer theory of association semantic based on power-law distribution of linking keywords in which power-law distribution of four types of keywords with different linking role has discovered.

Based on power-law distribution, two corollaries are presented: First are relations between ASC and exponent of power-law distribution, and the second is changing trend of ASC among any two adjacent layer. X.F. Luo, Zh. Xu, J. Yu[2] have proposed in their paper an approximate solution for adjustable average path length which can avoid the high time complexity in computing the shortest path between any two nodes in ALN.

1

This paper has demonstrated the validity of the approximate solution for adjustable average path length of ALN. X.-F. Luo and N. Fang [3] have proposed a systematic approach for automatically building Associated Semantic Overlay (ASO). ASO is a technique of semantically organizing and managing resources of a domain, which can provide associated links among web resources. It uses the matrix operations such as transpose of matrix, picking sub-matrix which facilitate for parallel computing.

S.X. Zhang, X.F. Luo, J.Y. Xuan [4] has explored the small-world properties of ALN to provide theoretical support for the association learning. In this paper initially a filtering algorithm has proposed to generate the filtered status of ALN and then investigation on the evolution of small-world properties has done on several incremental network sizes. Zheng Xu, X. Luo, W. Luo has proposed a method based on All-pairs algorithm for incremental building of ALN.

III. PROPOSED SYSTEM

1. Implementation of Proposed Model

The overall concept of proposed system architecture is given as follows. It includes the following modules:

o Collection of Webpages: Initially huge number of links are collected from different types of data sources with their content.

o Extraction of Keyword: In this module, links which contains the keyword to be searched are extracted or retrieved with their contents from the above module's webpage collection. The content of the related web pages will get stored in one file. After that pre-processing technique will be applied on that file which removes unwanted data. It means that it will remove the stop words such as "is, the, and", etc and stored only the needed data that will be used for generating the required result.

o Association: In this module, the dataset or dictionary for that keyword will be compared with the files which were generated in above module. It matches the words between dictionary of that keyword and extracted file.

Then matched words will be stored in another file.

o Count: From this file we get matched words. After getting these words, count is being calculated for each word. Then link, count or occurrence of the word in the particular link is stored in a database. Therefore, association link network helps to provide the keywords and the association relations between that keywords. In this way, the set of keywords as well as set of association relations are extracted from the given collection of web pages.

o Arranging on the basis of Priority: Now the last module provides the required data to the user on the basis of their priority. It shows the link first which contains highest count of the occurrence of the keyword, then second link contain less count as compared to the first, and so on. Along with link it also shows the content of that particular link.

o Recommendation: This module is used to recommend the user about the links which was searched currently as well as those which was searched for a long time.

o Today's Search topic: In this module, the topics will be displayed which are being searched on the same day by the same user or different user. This topic will be shown on the basis of their count i.e. how many times the same topic was searched. They will be arranged in descending order by considering count variable.
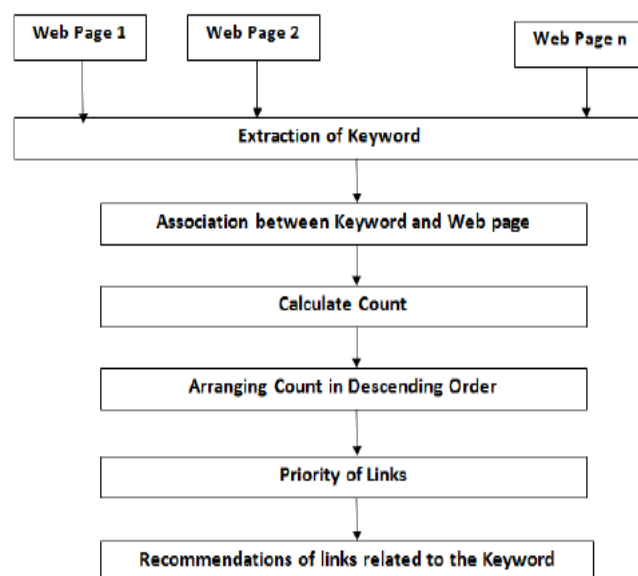
2. Workflow Diagram

2

3. Calculation of Term Frequency

Assume we have a lot of various kinds of documents among which some are interrelated with each other and some are not related. If one wish to determine a document related with "data mining". Then a user may type a query that is "data mining" in a search box. The simple way to search the documents related with the query "data mining" is to eliminate those documents which are not containing the two words such as "data" and "mining". But still there will remain many documents which does not contain these two words. To solve this problem, we will count the number of times the occurrence this keyword in each document. So, the number of times the word occurs is called the "term frequency".

To get the required documents related with the particular query, three steps should be done. In first step, all the words are separated with space. In second step, all the stop words such as "a", "the", "and" etc. will get removed. Then in third step, the count of each word will get calculated and on the basis of that count the web pages will be recommended to the user.

IV. RESULT AND DISCUSSION

1. The Experimental results are shown in the following figures. shows the Home page of this project which consists the navigations such as User, Admin, Register, Log out etc.
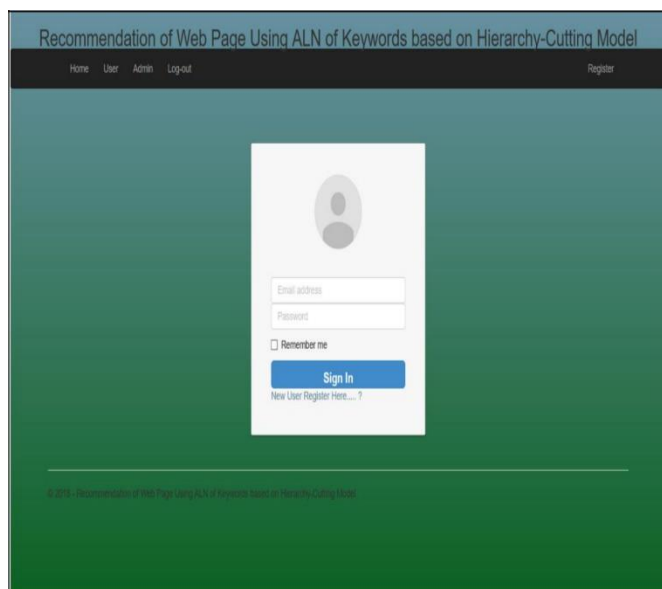
2. After clicking the Register, it will navigate to the Register page which is shown in below fig. To use these systems, first user has to register himself by creating his account.
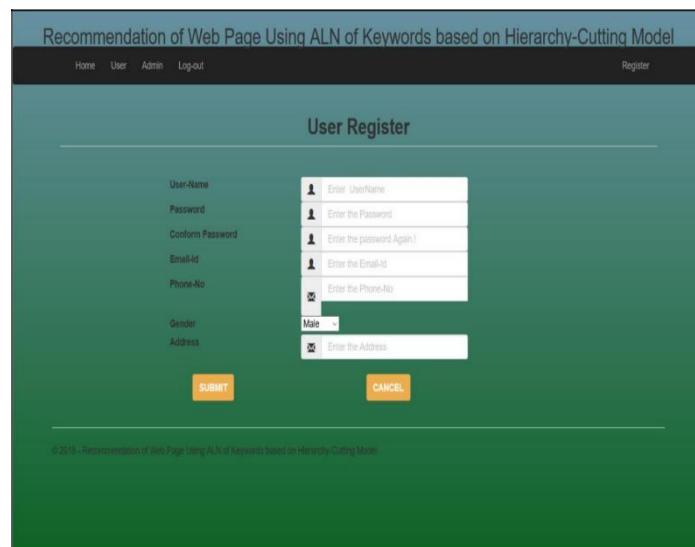
3. To search any keyword user has to first login by providing his username and password. Login page for user is shown in fig. 4.1. After that user can enter any keyword that he wants to search in a given search box and click on submit button which is shown in fig. 4.3.

Fig: Keyword Search Page

The system will show the related recommendation of web pages of specified keyword to the user. The system will also recommend for other topics which are not related but mostly search topics to the user. Along with that it will shows the today's current topic to the user. The system will recommend the web pages to the user related with specified keyword and along with that it shows the data in the form of graph. It means that it shows the count or occurrence of keyword in their related link with the help of graph. The graph is shown in below fig.
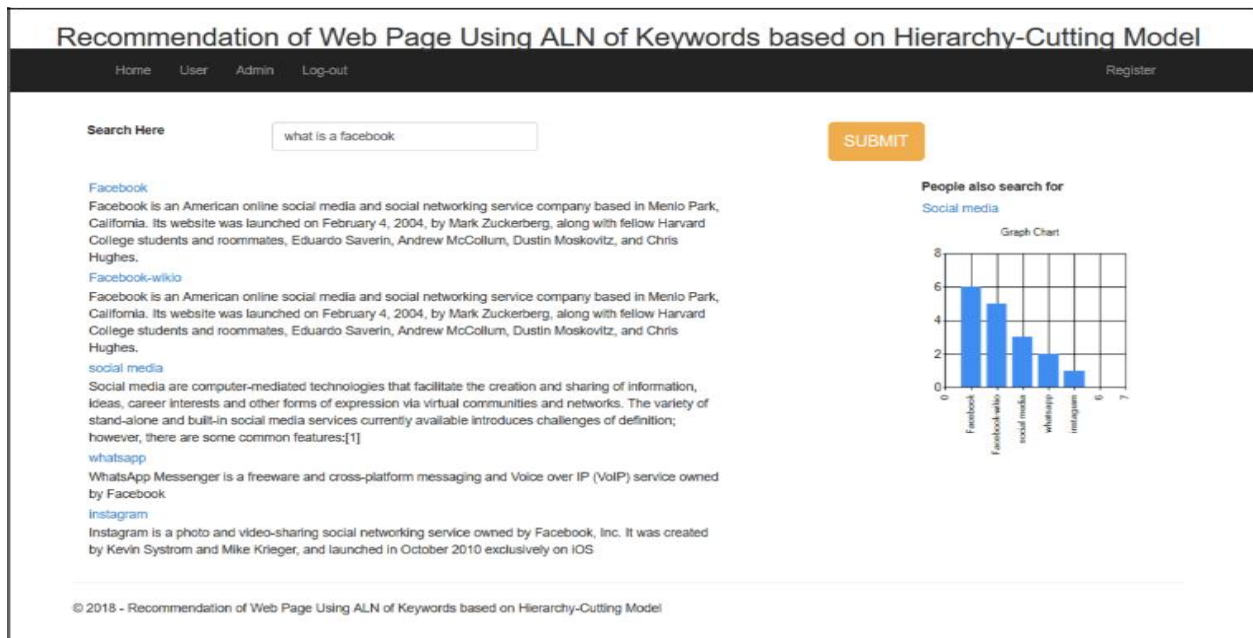
3

systems science and engineering, 26(3):153-162,2011.

Fig: Recommendation of Web Pages with Graph

V. CONCLUSION

This paper aims to provide the link of the required topic with the description to the user i.e. it provides the web page to the user which is specified by user by providing keyword to the system. The system also recommends the today's search or current topic as well as long term topic to the user. The graph is also given in this paper which shows the term frequency of the keyword to be searched in related topic. The proposed system uses the hierarchy model in which three layers i.e. Layer 1, Layer 2 and Layer 3 are used to search the keywords as well as to recommend the user regarding current as well as long term topic. Thus, the system provides required information to the user using association semantic link.

VI. REFERENCES

[1] Ziangfeng Luo, Kai Yan and Xue Chen, et al., automatic discovery of semantic relations based on association rules. Journal of Software, Vol. 3, November 2008.

[2] Z. Xu, X. Luo, and L. Wang. Incremental building association link network. Computer

[3] X.F. Luo, Zh. Xu, J.Yu, et al., Building association link network for semantic link on web resources. IEEE Trans. Autom. Sci. Eng. 8(3):482-494, 2011.

[4] Zh. Xu, et al, Semantic based representing and organizing surveillance big data using video structural description technology. The Journals of Systems and Software, 102, 217-225, 2015.

[5] S.X. Zhang, X.F. Luo, J.Y. Xuan, et al., Discovering small-world in association link networks for association learning. World Wide Web, 17(2):229-254, 2014.

[6] X. Zhang, K. Lu, W. Liu, X. Yin, and G. Zhu, generating associated knowledge flow in large-scale web pages based on user interaction Computer

Systems Science and Engineering, 30(5):377-389, 2015.

4