

SEMANTIC ANALYSIS OF WEB EVENT -REVIEW

Monika B. Bisane¹, R.V.Mante²

¹P.G. Student, ²Assistant Professor, Department of Computer Science and Engineering,
Government College of Engineering, Amravati (India)

ABSTRACT

On World Wide Web various kinds of webpages of a web event are generated in huge amount. And all of this webpages may contain certain and uncertain information. so it is difficult to understand the semantic of the web event. This paper proposes a method which do analysis of the semantic of web events which help to recommend appropriate webpage to the user. Firstly a web event considered as a system composed of different keywords Then keyword association linked network (KALN) is constructed to represent the Web event then Shannon entropy is use to identify the different semantic level of web event and construct a semantic pyramid (SP) to express the uncertainty hierarchy of a Web event. Finally, an SP-based Webpage recommendation system is developed. With these semantic analysis of web event, user could easily get the information about web event.

Keywords- Keyword association link network, web event, web mining, webpage recommendation

I. INTRODUCTION

The rapid growth of information on World Wide Web (WWW) has become huge repository of information and it keeps growing exponentially as new information or web-page added. The large amount of information is available on the World Wide Web (WWW).This information content different level of semantic uncertainties. So it becomes more difficult for users to access relevant information effectively. Hence it became more challenging to provide highly relevant information without any semantic uncertainties to users. This paper provides a new framework to identify the different levels of semantic uncertainty in terms of Web events, and then utilize these for Webpage Recommendations .The core of that web-page recommendation is recommendation of appropriate Web page of web event to the user and evaluating which web- pages that users would like to view in the future.

On the world wide web there will be huge number of webpages related to any particular event, for example if there are a events (topics) like earthquake in japan , different languages technology, or other topic that attracts broad attention on the web. There will be huge number of webpages related to that topic.

As time passes, hundreds and thousands of web pages, blogs, and post are generated. This huge amount of information makes it impossible for user to understand the event by reading that number of pages and also its takes too much time. And also that huge number of webpages contain certain and uncertain information. So to give relevant result to the user there is the need of the analysis of the semantic uncertainty. To understand, what is semantic uncertainty firstly there is a need to understand what is semantic of web event. As there are number of webpages, so in which webpage what content are there that is specify by a single meaning known as semantic

of web event. And the semantic uncertainty exists when similarities and difference between the meanings given by different people to same terms are not certain. so the goal of this paper is to analysis of the semantic of web events which help to recommend proper webpage to the user. And then if user wants to know anything on web second or third time then bases on history of records the prediction will be done and that most visited webpages recommended to the user.

In this paper we use the keyword association link network (KALN) to represent the web event. This web event is considered as it is composed of number of keywords which preserve the semantics of the web events. Then based on entropy different level of web semantics of the web event which consisted of Theme Layer, Backbone layer And Tidbit Layer are identified. And the semantic pyramid will be generated which consists of this different level of semantics. Then according to the mapping relation between webpage and keywords, the web page with different semantic uncertainties are recommended to the users.

II. RELATED WORK

The Thorsten Brants and Francine Chen [1] in 2003 proposes a model that present a new method and system for performing the New Event Detection task. New Event Detection is especially useful for the task of scanning multiple news sources for the latest news. It can be used in a categorization system to identify new categories of news stories and these stories can be used as examples of new categories. People who need to know the latest news when it happens, such as government analysts or financial analysts and stock market traders, can use New Event Detection to more quickly identify new events. But this method only focus on detecting the web event from an amount of web pages.

Zhongyun Ying, Zhurong Zhou, Fengjiao Han and Guofeng Zhu [2] proposed an improved Collaborative filtering algorithm to discover the similar users interested web page sets of the target user, based on which, a target user's collaborative filtering web page set is filtered. Collaborative filtering is successful approach for recommendation. It associates a user with a group of like-mined users based on their preferences over all the items, and recommends to the user the item enjoyed by others in the same group. But it is not practical to collect the feedback for web pages web events.

Pushpa C N, Ashvini Patil, Thriveni J, Venugopal K R and L M Patnaik[3] proposed to system uses the historical browsers data for provides users with most relevant web pages. This system propose to use the database and process rank. All the users' click-through activity such as number of times he visited, duration he spent, his mouse movements and several other variables are stored in database. The proposed system uses this database and process to rank them.

R. Forsati^{1*}, M. R. Meybodi² [4] proposed new methods for web page recommendation in Effective Page Recommendation Algorithms Based on Distributed Learning Automata and Weighted Association Rules. Firstly an algorithm is proposed based on distributed learning automata to learn the behaviour of previous users' and integrating web usage mining with link analysis techniques for assigning probabilities to the web pages based on their importance in the web site's navigational graph and makes recommendations primarily based on learned pattern and the structure of the web site. By introducing a novel Weighted Association Rule mining algorithm, second algorithm present for recommendation purpose. In which users' navigational patterns are automatically

extracted from web usage data. These navigational patterns are then used to generate recommendations based on a user's current status.

Chao Wang *, Jie Lu, Guangquan Zhang in 2007[5] proposed a method which points out the existence of key information in web pages and the significance of mining and using key information. It has also proposed KIM, a two-step method that can automatically mine key information from web pages. The method first extracts a list of candidate key information. It then uses an entropy evaluation to filter most of the noisy information in the list so that the key information can be discovered easily.

Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu in 2014 [6] proposed the two new models for representation of domain knowledge of a website. The first model is ontology based which represents domain knowledge of the website. The second model is semantic network to represent domain terms, web-pages and relation among them. This study introduced the conceptual prediction model which integrates the domain knowledge and web usage knowledge, to support effective and better web-page recommendation. This paper provides better web-page recommendation by combining domain and web-usage knowledge. But the problem is that Ontology is difficult to construct .So it would be better to incorporate more analysis of the contents of Web events.

Junyu Xuan et.al.[7], "Building hierarchical keyword level association link networks for Web events semantic analysis," proposed a method to map the web event to keyword level association link network (KALN) for deep analysis of the semantics of web events, such as the evolution semantics of web events. Firstly, the original KALN is constructed at a given time by traditional data mining technologies. Then, the hierarchical KALN, consisted of Theme Layer Network, Backbone Layer Network and Tidbit Layer Network, is built based on the original KALN by information entropy to identify the different semantic levels of the web event, including stable semantics, sub-stable semantics and unstable semantics. With the semantic analysis of hierarchical KALN, human could easily gain a thorough understanding of the web event.

Although there are many topics on webpage recommendation but the uncertainty analysis of the web event is seldom considered.

III. PROPOSED APPROACH

The main stages of the project are as follow:

- 1] Webpage collection
- 2] Keyword Extraction
- 3] Keyword Weight
- 4] Association
- 5] Entropy calculation
- 6] Pyramid Generation

The working of the proposed method are as follows:

The first step is to collect different web pages of the web event from different sources. Then that web event is represented by a keyword association link network (KALN) to preserve the semantic of web event at a given time. Web event can be consider as a system composed of different keyword. Then Keyword Extraction will be

done .By using keyword extraction algorithms (i.e term frequency (TF) and inverse DF) the nodes (keywords) of KALN from the dataset are collected. Once the node are fixed, the next step is to link these nodes by extracting the association rule.

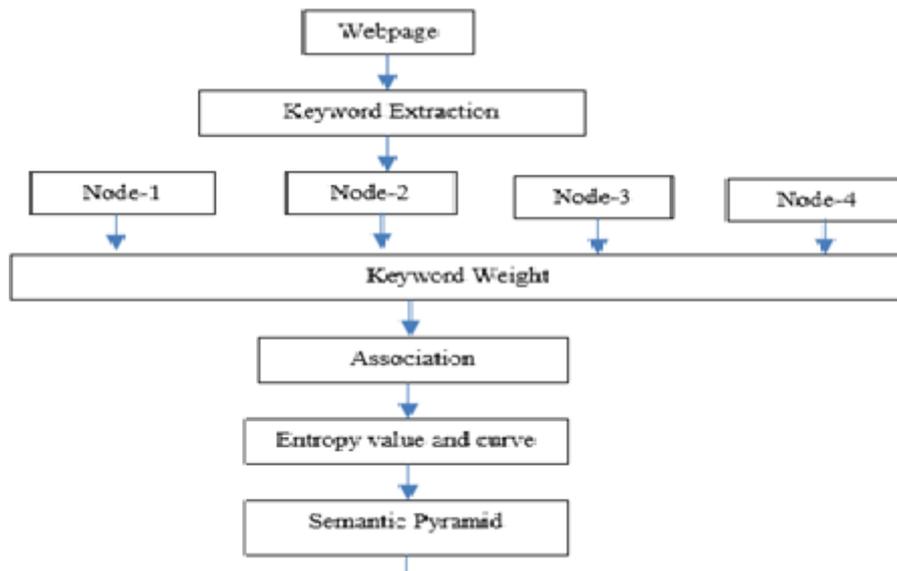


Fig. System Architecture

In step 2, based on Shannon entropy, an algorithm is proposed to identify the hierarchical uncertainty of KALN and then semantic pyramid is constructed for the uncertainty analysis. In Shannon entropy there is a classification of the keywords based on keyword weight. On the basis of that classification Semantic Pyramid is generated. This semantic pyramid consists of three layers.

- 1] Theme layer
- 2] Backbone Layer
- 3] Tidbit Layer

The explanation of this three layer is given as below:

1] Theme layer: Most certain Webpage recommendation is based on this Theme-Level KALN. This layer contain most relevant data related to web pages. This layer consists of large number of keywords which extracted from the particular web events. This will be the topmost layer. This layer is the core of the original network. It tells about the event.

2] Backbone Layer: Most uncertain Webpage Recommendation is based on this Backbone Level KALN. This layer is the Backbone of the original network. And it gives more detail information than the Theme layer Network and shows the main content of any event.

3] Tidbit Layer: Directional Webpage recommendation Based on Tidbit Level KALN.This layer is the Tidbit of the original network. It gives details of all aspect of this event.

Thus the webpage with different uncertainties are recommend to user.

IV. CONCLUSION

This paper aims to provide appropriate webpage to the user by performing semantic uncertainty analysis of the keyword system of web event. This paper proposed a content based Web event representation for preserving the semantic of web events. Then, an SP is constructed for uncertainty analysis and then by using mapping relation between Webpage and keyword, the webpage is recommend to the user. Thus different level of information is provided to people with different requirements.

REFERENCES

- [1] T. Brants, F. Chen, and A. Farahat, "A system for new event detection," in Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Toronto, ON, Canada, 2003, pp. 330–337.
- [2] Zhongyun Ying Zhurong Zhou*, Fengjiao Han and Guofeng Zhu "Research on Personalized Web Page Recommendation Algorithm Based on User Context and Collaborative Filtering", 2013 IEEE.
- [3] Pushpa C N, Ashvini Patil, Thriveni J, Venugopal K R and L M Patnaik "Web Page Recommendations using Radial Basis Neural Network Technique", 2013 IEEE.
- [4] R.Forsati1*, M. R. Meybodi2," Effective Page Recommendation Algorithms Based on Distributed Learning Automata and Weighted Association Rules".
- [5] C. Wang, J. Lu, and G. Zhang, "Mining key information of Web pages: A method and its application," Expert Syst. Appl., vol. 33, no. 2, pp. 425–433, Aug. 2007.
- [6] Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu, "Web-Page Recommendation Based on Web Usage and Domain Knowledge", IEEE Trans. On knowledge and data engineering, vol.26, no. 10, October 2014.
- [7] J. Xuan et al., "Building hierarchical keyword level association link networks for Web events semantic analysis," in Proc. IEEE 9th Int. Conf. Depend. Auton. Secure Comput., Sydney, NSW, Australia, 2011, pp. 987–994.
- [8] M. Songyun Guo et al., "Processing Combat Information with Shannon Entropy and Improved Genetic Algorithm "2015 8th International Conference on BioMedical Engineering and Informatics (BMEI 2015)