A Study on Intention Network Modeling Based on User's Interest Web-Page

Taebok Yoon[1] and Jee-Hyong Lee[2]

[1] Department of Computer Software, Seoil University, Seoul, Korea
[2] Department of Computer Engineering, Sungkyunkwan University, Suwon, Korea
tbyoon@seoil.ac.kr

Abstract. There are various studies to try to provide useful information to users in the Web. Web usage mining is a method to extract meaningful patterns based on web users' log data. Most of existing approaches of web usage mining, however, had not considered users' diverse intention but created general models. Web users' search keywords can have various meaning upon their preference and context. This study is for generating the User Intent Network Model (UIN-Model) after collecting and analyzing web usage information corresponding to keywords. UIN-Model can supply web page navigation networks reflecting users' various intention information. It can also be used to recommend most proper web pages and it has been confirmed that the suggested method was useful enough.

Keywords: UIN-Model, Web Mining, User Preference.

1 Introduction

Upon development of Information Technology, information on websites are rapidly increasing so that people should spend more time and make more effort to get right information that they want. Therefore, there are various studies to attempt to reduce the burdens. The studies can be sorted into two parts: a study to understand contents and structure of web pages [1] and a study to analyze a web-user's choice on web pages [2]. The contents and structure of web pages can be understood by analyzing images, texts or video in the pages, or by analyzing the stratified structure and linking structure of websites. To analyze a web-user's choice on web pages, the web user's activity, such as visiting pages and choosing menus, are logged. There are also several researches to analyze a user's visited web pages by observing mouse drags and clicks, or keyboard typing. However, the existing analysis and estimation on web pages is limited so that it is hard to offer services that consider various users' intention. For example, let us assume there is a keyword of "soccer". What comes across in your mind first? It reminds someone soccer games like World-cup or Champion's leagues, famous soccer players like "Pelé", "Maradona" or "Ji-Sung Park", or soccer supplies like soccer balls, shoes or uniforms. A keyword, thus, can have diverse meanings for users upon their preference, context and current status. It may not be said as an

intelligent and applicable service if users have common results on their interest keywords in spite of their diverse intention. The User Intent Network Model (UIN-Model) is a navigation network that can reflect users various search intention, generated by reviewing web pages that users hit on keywords. Also, we collect the web users' web page visiting information with a keyword and create the UIN-Model as a model of navigation patterns according to search intention. The UIN-Model is useful for recommending web pages, advertisements upon keywords and grasping the meaning that a user intended to represent.

2   Related Works

Researches related to recommending web pages to supply meaningful information related to users' interest keywords are very diverse. Joh et al. [3] and Hay et al. [4] illustrated sequence alignments of web-users activities and compared the similarities among them, and Sufyan and ahmad[5] studied web search enhancement by mining users' actions to analyze their website visiting information. Also, Ahmed et al. [6] was introduced the methods to know the user's interest using Web log.

   These existing researches collect and analyze web-users log information, find out patterns and model their web using information. Those models utilize standard technology to perform services like automation, intelligence and personalization types etc. Those, however, have limited features for their extent of usage due to their models which did not regard various users' intention. It is necessary to analyze various users' intention and research to reflect them to create models.

3   Design of UIN-Model

On web environment, users can access any website to get information that they wanted through various searching engines like Google, Yahoo, Naver etc. If a user searches with a keyword and reviewed a website carefully, the information can be used as useful information to recommend web-searching.

   The user's interest keyword, user ID and user's activity information on the website can be tools to measure how visited web page was meaningful for the user. There is many users' activity information on visited web pages: user ID to distinguish users, browsed website URL by using interest keywords, starting and ending time of web pages, activities of download, copy and paste and bookmark, contents size of web pages etc. Web-page scoring must be used to value a website in number. We need to consider how each of elements affects each other for scoring. Each element can decide importance by using weight value. The numerical formula as below has been used to measure how important a web-page for a user regarding weight of each element.

$$PageWeight_j = 1 - \left[\frac{1}{\sum_{i=0}^{n}(C_i \bullet Attribute_i)} \right] \tag{1}$$

PageWeight_j means the weight of web-page **j** among various referred pages based on some keyword, and **n** means number of elements (web-users' activities like duration, book-marking etc.) used for assessing web-pages. **Attribute_i** means element **i** and **C_i** means weight (constant) of element **i**.

If **n** of users' information had been collected, cost for managing and calculating network will increased by having **n** branches. It will be helpful to understand created network if groups of web-pages referred by users based on interest keywords are not simply listed but can be implied expression of users referred to similar web-pages. We can call it intentions, collected personal web-page browsing information by key-words. For more important expression of network, similarities of intentions and relationship of inclusion are compared and unified. Unification of intentions are assorted to Unity type, Inclusion type and Partly-unity type. Partly-unity type can be decided their similarity by using formula below and possibility of unification by status of similarity.

$$Sim(X,Y) = w_S S - w_u U \tag{2}$$

From **Sim(X, Y),** **S** means the number of web-pages that two sets include commonly and **U** means the number of web pages that two sets didn't. **w_u** is weight on web-pages that two sets have commonly and **w_u** is weight on web pages that two sets don't. If threshold is more than two group's similarity, they must be unified and weight of web-page should be unified and become one weight. Combining method by analyzing web-page similarities measured similarity of two groups by multiplying number of repeated pages and weight of number of not repeated pages.

4   Experiment

The experiment used twenty keywords except games or specific searching engine among top 30 popular keywords in 2010 and 2011 on Google, Yahoo and Naver. In case of keywords to access specific sites like the lottery site, National Tax Service or games like Sudden Attack or Dungeon fighter, one-click on searching results leads
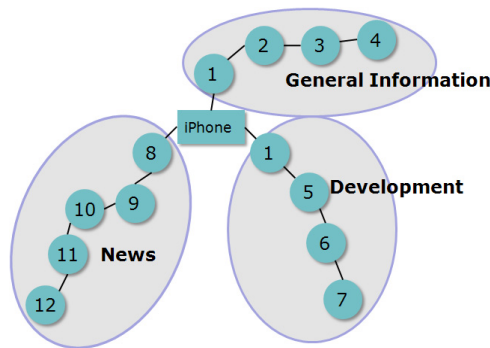

Fig. 1. UIN-Model of keyword "iPhone"

you to the right sites for a user's want. If there is only one absolute website for all users' want on some keywords, recommendation has no meaning.

Seven staffs have been selected for this experiment. They visited 823 web pages and 451 of collected data have been used for UIN-Model creation after deleting meaningless web pages. 141 groups also have been combined to 83 groups through UIN-Model. Figure 7 illustrated a keyword of iPhone network by using UIN-Model. The group with web pages 1, 2, 3 and 4 were about general information of iPhone, web pages 1, 5, 6 and 7, news of iPhone and web pages 8, 9, 10, 11 and 12, development for iPhone.

5 Conclusion

This study suggested UIN-Model as a method to include diverse propensity information on web-users' searching. It created network of web pages based on web-users' using information through keywords, measured and united similarity among users' intentions and made more meaningful network. Generated UIN-Model can be utilized for web pages recommending service and basic technology to compare, analyze and decide similarities among keywords on networks. UIN-Model based on keywords created from experiment was describing information on users' searching activities carefully so that they can be useful for recommendation service.

References

1. Eirinaki, M., Vazirgiannis, M.: Web Mining for Web Personalization. ACM Transactions on Internet Technology 3(1), 1–27 (2003)
2. Srivastava, J., Cooley, R., Deshpande, M., Tan, P.-N.: Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations 1(2), 12–23 (2000)
3. Joh, C.H., Arentze, T.A., Timmermans, H.J.P.: A position-sensitive sequence alignment method illustrated for space-time activity-diary data. Environment and Planning A 2001 33, 313–338 (2001)
4. Hay, B., Wets, G., Vanhoof, K.: Clustering Navigation Patterns on a Website using a Sequence Alignment Method. In: Proc. Intelligent Techniques for Web Personalization: 17th Int. Joint Conf. Artificial Intelligence (2000)
5. Sufyan Beg, M.M., Ahmad, N.: Web Search Enhancement by Mining User Actions. Information Sciences 177, 5203–5218 (2007)
6. Ahmed, A., Low, Y., Aly, M., Josifovski, V., Smola, A.J.: Scalable Distributed Inference of Dynamic User Interests for Behavioral Targeting. In: The 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2011, KDD 2011 (2011)