Adaptation of Navigation by the Modified Results of Full Scan Algorithm in Adaptive Hypermedia Systems

Marek Bober[1], Petr Šaloun[2]

[1] Dept. of Computer Science, Faculty of Electrical Engineering and Computer Science,
VŠB-Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
marek.bober@vsb.cz
[2] Department of Informatics and Computers, Faculty of Science, University of Ostrava, 30.
dubna 22, 703 01 Ostrava, Czech Republic
petr.saloun@osu.cz

Abstract. Models of behavior mining from logs in adaptive systems are one from many methods how to obtain useful information for personalized navigation. Target of this paper is to describe modified results of full scan algorithm used to creation recommended links to concepts. For creation recommended links is used technique of traversal patterns. Established weights up to concepts and links along with the results of full scan algorithm created relevant recommended links within the scope of all subject matter and within the scope of topic in actual concept. The fundamental part is facilitating orientation for user in hyperspace.

Keywords: knowledge discovery, adaptive navigation, traversal patterns, concept recommendation, adaptive web-based educational hypermedia

1 Introduction

The content of submitted subject matter in adaptive systems can be dividing into pages, that incidents specific part of material. There are mostly chapters or parts of chapters. Sometimes is possible divide content to more detailed parts e.g. paragraphs, examples, tests etc [9]. Every chapter can be characterized as an individual page (web page) called concept.

Concept consists:
- content from author,
- list of attributes, that describe concepts,
- list of rules, (conditions) that determinate format and style of displayed parts of content concepts.

Chapters can be referenced to each other as well as concepts. References between concepts in adaptive hypermedia systems based on WWW technologies are created by

hypertext links. In adaptive systems is mostly adapted content of concept, style and navigation technique. By the references to other concepts we want to advise to user which toward a given topic is relevant other topic. Mostly, there has been related concepts with specific „relevancy" weight to understanding subject matter to study concept.

This paper describes usage of *full scan* (FS) algorithm for finding recommended links and new possibilities of modification of FS algorithm results for creating recommended links, named in adaptive hypermedia systems. The main of target is notify to user, which has the highest weight (relevancy) of link context to specific concept and made easier decision for user about relevant concept reference to visit it or not. Modified results can improve orientation in hyperspace than non-modified results of FS algorithm. For modification of results we introduce concepts and links relevancy weight, as you can see in next chapter. These values are necessary for modification of FS algorithm results.

2   Problem description

To notify user about specific link relevancy for him and his study process, is necessary to determinate not only weight of link but weight of content offered by other concept. Weight determination is not simple and all weight values should have been given to every user individually, or if we would like target group of users with the similar characteristics, for each groups individually. Every user has a different style of learning and personalization is basically difficult. The next problem is occasion of author to advise which content of text are more and which less important. One of the possibilities is non automatic weights setting of concepts by author. Non automatic weights setting are necessary for creation new concepts. The better occasion is for example automatic (semi-automatic) setting by the ontology and modifying by the neural networks. It is subject of next research.

We can determinate several *relevancy weights* (WR) for all concepts. For example we can use percentage valuation by displaying weights in ever concepts. Now is possible say to user, which is the important study content for him in light of relevancy in concrete subject, see Figure 1. The traditional case of referencing to several concepts is self WR concept poor. From the other one point of view can be one concept more relevant (it has greater valuation) than other concept for some case based on subject summary matter. There are necessary to determinate WR between the links of several concepts. If a concept has references at each other then weights can have same or different valuation in both ways of relevancy view within actual topic. Relevancy from the one side can be different than from the other side. It is necessary set WR into links every way.

Let C denotes concept. Every concept has to be denoted by unique identifier (C1, C2, C2.1 …). Concept C1 has WR 0.8 (80 %) and referencing to next two concepts (C2 and C21). Concept denoted as C21 has value 0.4 and concept C2 has value 0.8. In the case of summary topic we can determinate relevancy of referenced concepts by the value of weights into hypertext links of referenced concepts. From all of view at content relevancy in actual concept we merge relevant and local evaluation of several

links, which links starts from concept or pointed to this concept end in concept. For user, which study selected topic, is value of WR predicative about relevancy. In the case of summary subject matter is unused. For example concept C21 can be only 20 % relevant in case of topic of concept C1 but in case of summary subject matter can be 80 % relevant.
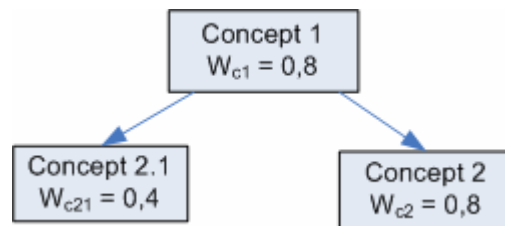

Fig.1. Weights of concepts which determining relevancy of content within the scope of summary subject matter.

By the WR links valuation we can determinate, how is close the relation between referenced concept and actual concept. Determination of WR in first phase (to create new concepts) of personalization is difficult and requires cooperation of subject material authors. In more advanced phase value of WR can be modified for example by the Radial Basis Function from neural network etc. Determination of WR into links is showed at Figure 2.
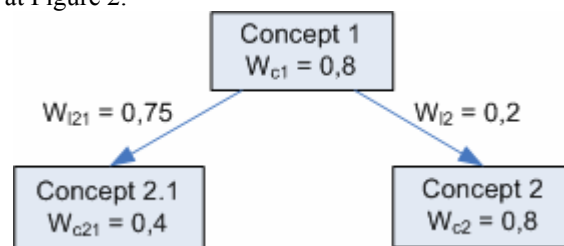

Fig. 2. Weights of links relevancy.

Values w_{c1}, w_{c2}, w_{c21} are percentage of content relevancy to other concepts in case of summary subject matter. Values w_{l2} and w_{l21} are percentage of related topic relevancy to other concepts with concept from which is linked to. In this case concepts cross references cannot use weight of link same for both ways. In opposite direction relevancy of links can be different, see Figure 3.

In line with value of WR w_{l21} topic of concept C2.1 is concerned per 75% of topic concept C1. But value of WR in the opposite direction w_{ll} determine related content C1 only per 15%. Concepts represent vertexes and hypertext links represent angles. Now they creating oriented graph with valuated angles.
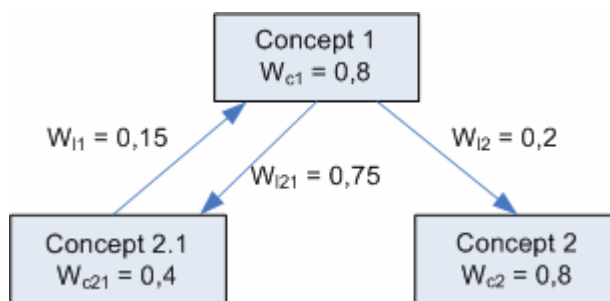
Fig. 3. Different WR for links in opposite direction.

WR of hypertext links and concepts can have relevancy restriction. WR can be determinate in case of summary subject matter, one chapter, paragraph etc. In adaptive environment system it can offer related links in ordered list by the value of weights (link recommendation), for example by the traversal pattern algorithm [8]. Recommended (related) links and the most frequently links from all users can be string together by interconnection with properties of concrete user or user group and the result is offered to concrete user or group[9].

3   Related works

Research in adaptive hypermedia systems is based on collection data, analyzing (data mining) and their consequent usage for personalization. In systems based on WWW pages the collection of data proceeds by hide form during user activity in system. Collected data represents small form patterns complex. Relevant data obtaining process and their usage for personalization can be divided into basic steps for knowledge discovery and their later usage [7, 11]. The first step is the collection of patterns that made preprocessing and cleaning. Follow choosing relevant attributes and transforming data – data mining. In last step data are analyzed, interpreted and used in work experience – personalization.

Target of summary process research is the most effective content and style adaptation for navigation to concrete user or user group in order to increase of effectiveness learning and hyperspace orientation. Typical personalization properties are links visualization, show or hide some text parts and show recommended links too [3]. Follow chapters are targeted to show recommended links by modified results of *Full Scan* (FS) algorithm of traversal patterns.

3.1   Environment for personalization

Every adaptive hypermedia system uses some technique for adaptation. There are two basic adaptive techniques – content and navigation adaptation. Term adaptation is substitute by term personalization [1, 6].

Information from web pages still increases the value of orientation in these pages

with more exacting values. Navigation in WWW pages can be shown better and allows user better orientation. Navigation and personalization target is to help user in order to better orientation in hyperspace. There are five different methods of adaptation: direct navigation, sorted links, hided links, links commentary, and map adaptation [1].

The most systems that work with adaptive hypermedia are based on model. For example adaptive hypermedia system AHA! [4, 5] is based on referenced model AHAM [5]. AHAM model define four basic models [3, 4]:

- Domain model (DM) describes structure of system content (fragments, pages and concepts).
- User model (UM) store information about users and their behaviors in system.
- Adaptation model (AM) is defined as set of rules. These rules define how adaptation has proceeded.
- Adaptive mechanism provides an adaptation (outgoing from adaptive rules) and generate pages by using different adaptive techniques.

3.2 Data for knowledge discovery

For adapting content is necessary obtaining information on which they are based adaptation proceed – is necessary observe user behaviors in system, it is means to observe his walk across system [4]. In our case of content adaptation we display recommended links within the scope of all subjects matter and within the scope of actual topic. It is necessary data stored in user log.

- Student Id – student identification
- Date and time – time stamp: date and time by the login/logout of user to/from system that represent start/end looking time over the concept.
- Type of access – attribute can have tree values: access to concept with study subject matter, access to test, and general access to system.
- Concept – Concept identification.

System Domain model contains basic information, that identify every concepts. These data and user log are used to basic adaptation mechanism.

4. Algorithms for data mining

The main source of data for obtaining knowledge is log of user activities. Log contains all information about user activities in system during current session (realized by logout a login). Our target is exploring characteristic patterns of user's navigation.

Because user activities in system looks like transactions in electronic shops, it is possible to use tree basic techniques for data mining: association rules mining, sequential patterns mining and traversal patterns mining [10].

Association rules mining. In the case of electronic shops this technique is used to search group of similar products based on user knowledge (preferences). In adaptive systems it is used to search relationships between concepts. This technique is suited to authors, which produce or modify the course. For example Apriori algorithm represents association rules mining [1].

Sequential patterns mining. Technique is resembled as association rules mining. This technique calculated only with visited concepts. Sequential patterns apply one-self to recommended concepts. From the user log we can deduce, which concept is the most visiting but it is not possible to specify, which concept will be visited in a future. Technique of sequential patterns make possibility to find concepts, which user would visit in future – recommended concepts. There is analogy with Apriori algorithm [2].

A traversal pattern mining is sometime called as a continuous sequential patterns. This technique is mostly used for web server log analysis. In adaptive systems is recommended to user relevant concepts for visiting. Algorithms that represent traversal patterns are full scan (FS) - selective scan [8].

Next parts of document apply oneself to use full scan algorithm and consecutive modification of results.

4.1   Modified algorithm results for traversal patterns

Specification of FS algorithm outgoes from DHP (direct hashing and pruning) algorithm [8]. FS algorithm was used in adaptive hypermedia systems ALEA, see [9]. Authors compared tree techniques for data mining. The result of this comparing shows recommended links for users. They start from visited to non-visited concepts because in tutorial process users very often come back to previously visited concepts.

Our main idea is integrate results of FS algorithm with WR within the scope of all subject matter, and with WR of links within the scope of actual topic of concept. As a result are recommended links to concepts in two categories: recommended links within the scope of all subject matter and recommended links within the scope of actual topic of concept. Usage of FS algorithm and modified results is shown on Figure 5.

Modified part of log in database

TID	Concept
1	C1 C3 C4
2	C2 C3 C5
3	C1 C2 C3 C5
4	C2 C5

Result of FS algorithm

{C2 C3 C5}

Weights of concepts

Concept	Weight
C2	0,75
C3	0,70
C5	0,25

Fig. 5. Results of FS algorithm

Left table in Figure 5 contains session identifier TID, and set of visited concepts. Results of FS algorithm are tree qualifications for concept recommendation C2, C3 and C5. System can display user links to concepts. But in case of concepts have WR within the scope of all subject matter; see right table in Figure 5, the result of algorithm can be modified. From the table we can see, that concepts C2 and C3 have relevancy within the scope of all subject matter "relatively big". Concept C5 has relevancy "relatively small". We determinate a minimum limit of relevancy within the scope of all subject matter, denote *L* and determinate value to 0,35. By applying limit L_A to results of recommended concepts is possible to eliminate concepts, which are not relevant for link in subject matter.

In case of results, as shown Figure 5, are relevant recommended concepts only C2 and C3. Concept denoted C5 have not to be displayed. It is possible to use another technique than displaying and hiding recommended links, for example highlighting of background color. In the case of concept C5 can be highlighted of other color.

If there are established WR for links, that represents relevancy within the scope of topic concept, it is possible to offer user relevant links with relation to given topic. Let {C2, C3, C5} is result of FS algorithm and structure of connection concepts is in agreement with Figure 6. Algorithm proceeded from root, where is user occurred in agreement with Figure 6, on concept C1.
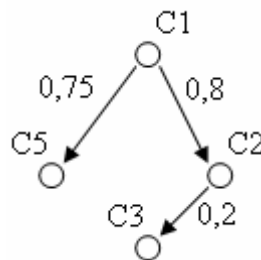

Fig. 6. Structure of connection concepts

Figure 6 show relevancy of concepts within the scope of connection. For user which occurs on concept C1 are relevant links to concept C2 and C5. In the result of FS algorithm is additional concept C3 that should have been displayed. We determinate minimal limit within the scope of relevant topic L_T and determinate value to 0,35. Link from C2 to concept C3 dispose less value than defined limit, concept is "a few relevant" within the scope of concept C2. That is why link to concept C3 not have to be displayed. In case of technique of highlighting is possible to display link of other background color. Results of obtained, modified and displayed values FS algorithm are shown on Figure 7.

Recommended links in current topic	Recommended links throughout course
Link to C2 – 80%	Link to C2 – 75%
Link to C3 – 20%	Link to C3 – 70%
Link to C5 – 75%	Link to C5 – 25%

Fig. 7. Example of display recommended links

5. Conclusions and future work

Adaptation of content to several users or user groups by the different determinate rules for all users individually is very complex problem, because several users (user groups) have different style of learning. Techniques for data mining based on user system behaviors helps to characterize specific groups with the same behaviors, and based on obtained knowledge is possible to apply adaptive mechanisms for concrete groups.

One of the eventual adaptation techniques is presented ease in hyperspace by the recommended links to relevant concepts. By using traversal patterns technique along with value of weights in concepts and links is possible offer to user relevant content for their next study of content. Modified results of FS algorithm can by tried in adaptive hypermedia system AHA! [5] and LMS systems called Barborka, that is develop on VŠB-Technical University of Ostrava, see [6]. Modified algorithm for traversal patterns can be extended by entries of ontology and information about knowledge of user obtained during their study in adaptive hypermedia system.

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke, and C. Zaniolo, editors, Proc. of 20th Int. Conf. Very Large Data Bases, VLDB, Morgan Kaufmann (1994), 487-499
2. Ayres, J., Flannick, J., Gehrke, J., Yiu, T.: Sequential pattern mining using a bitmap representation. In Proc. of the 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (2002), 429-435
3. Bieliková, M.: Creation of content in adaptive hypermedia for e-learning. Tvorba obsahu v adaptívnych hypermédiách pre e-vzdelávanie. [in Slovak] In Technologie pro e-vzdělávání, Prague (2004), 13-23. ISBN 81-01-03167-5.
4. [19] Bober, M., Šaloun, P., Velart, Z.: Interactive PDF forms for multichoice testing in AHA!. ICETA 2005, 121-124, ISBN 80-8086-016-6
5. De Bra, P., Aerts, A., Smits, D., Stash, N.: AHA! Version 2.0, More Adaptation Flexibility for Authors. In Proceedings of the AACE ELearn'2002 conference (2002), 240–246
6. Fasuga, R., Holub, L., Šarmanová, J.: Support of learning of technical subjects and their usage system Barborka. Podpora výuky odborných předmětů a jejich aplikace s použitím systému Barborka. [in Czech] In Geometry and Computer Graphics 2004, 25. konference o geometrii a počítačové grafice. Ostrava, Czech Republic (2004)
7. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. E.: Advances in Knowledge

Discovery in Data Mining. AAAI Press/MIT Press, 1996

8. Chen, M. S., Park, J. S., Yu, P. S.: Data mining for path traversal patterns in a web environment. In 16th Int. Conf. on Distributed Computing Systems, pages 385-392, 1996

9. Krištofič, A., Bieliková, M.: Improving adaptation in web-based educational hypermedia by means of knowledge discovery, In 6th ACM conference on Hypertext and hypermedia, Salzburg, Austria (2005) 184–192

10. Mobasher, B., Cooley, R., Srivastava, J.: Automatic personalization based on web usage mining. In Communications of the ACM, 142-151, 2000

11. Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.: Web usage mining as a tool for personalization: A survey. User Modeling and User-Adapted Interaction, 311-372, 2003