IDENTIFYING HYPERMEDIA BROWSING STRATEGIES
A Statistical Approach

Kristian Kiili and Harri Ketamo
Tampere University of Technology, Pori

Abstract: Browsing strategies have been studied mostly from the browsing patterns point of view. In this paper the browsing strategies have been approached with statistical methods. As a final result, a three-cluster browsing strategy model was defined according to empirical data, collected from web-based course. This model could be used as a background for constructing adaptive navigation for web-based learning material. Also a revised framework for navigation is presented.

Key words: hypermedia, modelling, learning materials

1. INTRODUCTION

Educational hypermedia can be defined as a learning environment that relies on user-controlled choices to access information in the form of various media (Reed & Oughton, 1997). Hypermedia allows users to make decisions about what information to access and in what sequence. The user constructs the structure of a personal hyperspace, the previous knowledge about the domain controls, and redirects users' behaviour in hyperspace. According to Lawless et al. (2003) this changes the nature of learning that occurs with hypermedia systems compared to traditional learning environments by adding users responsibility to construct the structure of the hyperspace. In addition users have to be able to identify what information will further enhance their understanding and how to access this information. Thus, users' ability to structure and manage their own navigation is becoming a required skill in hypermedia based learning environments (Ford & Chen, 2000). In current research terms navigation and browsing are not well defined. These

terms are sometimes used wrongly as synonyms but also determined as sub-classes for each other.

In the literature review the terms are used as they are mentioned in the original publications. Spence (1999) has presented a semantic framework for navigation. He defines navigation as the creation and interpretation of an internal model. According to him navigation consists of browsing, modelling, interpretation and the formulation of browsing strategy. Such as Spence, we consider browsing as a part of navigation process. In discussion section we present our vision about navigation framework inspired by Spence.

2.   BROWSING STRATEGIES

Despite the fact that two users do not navigate a hypermedia space in exactly the same way, similar navigational patterns do emerge among users. At least three common navigational profiles have been identified: 1) knowledge seekers, 2) feature explorers, and 3) apathetic hypertext users. Knowledge seekers pursue screens that contain material needed to enhance comprehension of the certain content. Additionally, knowledge seekers tend to select screens in logical sequences and acquire information in a systematic manner. Feature explorers spend a disproportionate amount of time interacting with special features such us movie clips, sound and visual effects, and graphics. It seems that apathetic hypertext users do not care either to gather information or to explore its features. Their navigational paths do not reveal logical orders and the interaction with the system is characterized with short time intervals. (Lawless, 1998.)

According to McAleese (1989) browsing is using explicit or implicit associations to determine the next item to be accessed. People using hypermedia for similar purposes produce similar, but not identical browsing patterns. Browsing pattern can be defined as a series of movements through the hypermedia. Canter et al. (1985) describes four lower level browsing constructs presented in Figure 1, which combinations form more complex patterns. These higher-level browsing patterns occur as the user employs a browsing strategy. Thus, browsing strategy can be defined as a method employed by the user in order to fulfil their task while using a hypermedia system.

The use of browsing constructs in recognition of browsing strategies is challenging because it is not readily apparent where one construct ends and another construct begins. It is also possible that constructs merge into each other. The time spent in each node is also an important factor when identifying browsing patterns. It may be that user is using the node as a

stepping stone to another node or to be interested in the content of the node itself. Thus there may occur several types of browsing patterns including the same nodes. (Mullier, 1999.)
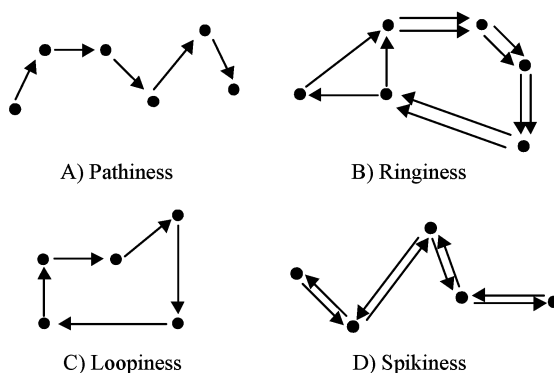

A) Pathiness   B) Ringiness

C) Loopiness   D) Spikiness

Figure 1. Browsing constructs

The study of Mullier et al. (2002) give some evidence that low-level browsing patterns are used to different degrees by different types of users. Expert users tended to use loops and rings with a high degree of path branching. They did not use spikes as much as more novice users. Mullier et al. (2002) also states that there is a finite number of browsing strategies and learning styles and therefore users with similar backgrounds may tend to behave in similar ways when using a hypermedia domain.

Canter et al. (1985) have described five browsing strategies:

1. Scanning strategy, the aim of which is to cover a large area without depth. It consists of a mixture of deep spikes and short loops.
2. Browsing strategy, in which a user just follows the links until his or her interest is caught. It consists of many loops and few large rings.
3. Searching strategy, in which a user seeks the marks of the requested issue. It consists of ever-increasing spikes with few loops.
4. Exploring strategy can be seen as finding out the extent of the information given. It consists of many different paths.
5. Wandering is purposeless and unstructured shilly-shallying that represents the lack of strategy. It consists of many medium sized rings as the user ambles along and inevitably revisits nodes.

Browsing strategies tends to be very complex because users can apply more than one strategy during a particular session and may use different strategies at different times. According to Catledge and Pitkov (1995), users tended to use different strategies, and hence produced different browsing patterns, if they had well defined goals than if they had unclear goals.

Additionally, users' goals can change at any point they browse because they can find some interesting information that differs from previous goals.

3.   RESEARCH QUESTION AND METHOD

The main goal of this study is to create methods for finding out the theoretical browsing strategies from observed user behaviour. These methods could be used as a background for future studies in adaptive navigation systems. Adaptive navigation should be designed so that 1) it is invisible for the user and 2) it could react fast to changes in users' behaviour. Therefore the system should understand the user behaviour without other human input than the user behaviour in web.

Browsing strategy is usually defined according to browsing patterns. It is also possible that the same browsing strategies m be found out from users statistical behaviour in hyperspace. If browsing strategies can be defined from statistical user behaviour, we could implement lightweight system that could both observe and analyse the browsing strategies. The benefits of this kind of statistical based user-modelling engine can be found especially when the computational speed of the server is limited and heavy pattern recognition could not be used. This could occur for example when there are large numbers of users in the server at the same time.

The empirical study was done in spring 2002 in co-operation with TietoEnator e-Learning Solutions. Approximately 400 students started the course, but finally 226 students finished the course. In this study we focus only on those users who finished the course. The total number of nodes in the learning material was 52. From a user point of view it was not necessary to read all the nodes, even though it was possible. The length of the users navigation paths varied from 30 up to 300 visited nodes. The students participated in course from their own PC-work stations. There were no lectures or other group meetings during the course. The students were aware that their behaviour could be recorded, but their identity could not be traced afterwards from these records.

The course was TietoEnator's internal web-course about business strategies. It consisted of four different views of the issue that were meant to describe different views designed for personnel in different positions. Materials were produced with Macromedia Flash and XML-features were used to communicate with the external logging system. Because of analysing consisted of constructing relation networks between log events, the user behaviour log was created directly in MySql-database.

For the statistical analysis the log files were summarized and the following variables were calculated separately for each user from raw log-data:
1. Number of visited nodes
2. Number of visited nodes in path A - D
3. Average visiting time / node
4. Average visiting time / node in path A – D
5. Standard deviation for visiting time / node
6. Standard deviation for visiting time / node in path A – D

The main variables were 1) the number of visited nodes and 2) the average visiting time per node. The detailed variables were calculated to ensure the validity of statistical analysis.

The biggest threat to validity was the possible effect of the context to users browsing strategy. To make sure that the browsing strategy was not depending on the context, the browsing strategies in different contexts (A-D) were compared. There were strong correlations (0,901 < r < 0,216: 0,000 < p < 0,003) between the main variables and the detailed variables. This proves that the general variables describe users' behaviour in all of the contexts. In other words, the context did not affect the browsing strategies.

4.   RESULTS

In this paper we focus on the main result of the study: the clusters found from the users' behaviour. These analyses were done with K-means clustering. From the literature review we can see that there were three to five browsing strategies, depending on the definitions. In our analysis we defined these strategies from our statistical data by clustering the users and then explaining the meaning of the clusters.

First we approached Canter's five- strategy model. Five clusters could be found from the user data (Figure 2). From these, the cluster that has a lot of visited nodes with short visiting times represents the wandering-strategy in Canter's model. Other strategies cannot be described in the terms of Canter, but we could build software that tries to give an explicit category for single user according to this five-cluster strategy definition. Unfortunately, this browsing profiler software could not calculate explicit browsing strategy for the user. Most users got several uncertain strategies as a result of this modelling. This indicates that this five-cluster model is not relevant enough for the basis of the statistical browsing strategy modelling.

Secondly we approached traditional three-strategy model. Three clusters were found from the user data (Figure 3). Strategy called 'apathetic hypertext users' can be seen as the same cluster as 'wandering-strategy', where user had

visited a lot of nodes with short visiting times. 'Knowledge seekers' -
category can be seen as a combination of 'searching- strategy and exploring-
strategy'. In this cluster the users search information and therefore they have
a small number of visited nodes with short average visiting times. Naturally,
when they find the information they searched, they have longer visiting time
in the specific node. 'Feature explorers' do not have explicit corresponding
categories in Canter's model. Feature explorers differ from knowledge
seekers by the average node visiting time. In this study, feature explorers are
not defined according to use of special features (video clips etc.) they just
had significantly longer average visiting times. We rather concern these
users as readers that cannot decide what information to access. Therefore
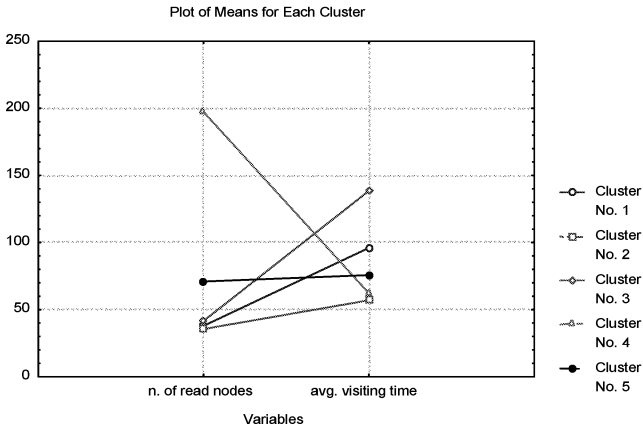they read everything in a linear way.
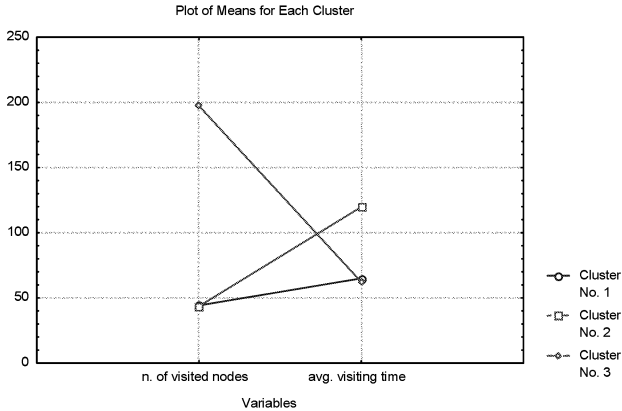

Figure 2. Five-cluster model.

Figure 3. Three-cluster model

This three-cluster browsing model could be used as a background for browsing profiler software. The browsing profiler could calculate the explicit browsing profile for over 60% of users. The clustering rules will be developed so that even more users could be profiled with browsing profiler software.

5. CONCLUSIONS

The statistical browsing strategy modelling was promising. The browsing strategies, that many researchers have defined according to browsing patterns, could be also defined by statistical methods. The strategies and definitions are not exactly the same as pattern based strategies, but they certainly describe some common features of users browsing in the learning material.

In Figure 4, a revised framework for navigation is presented. It is divided to three main components on the basis of information processing theory. The lowest level is called *action:* that describes the simple stimulus-reaction-response chain. In the second level, the user has made a selection about information that is processed in working memory. In this level the user employs a browsing strategy that is formulated on the basis of prior knowledge. Wandering represents the lack of strategy and it can be considered as transition of action and browsing. In summary browsing is regarded as a registration of content.
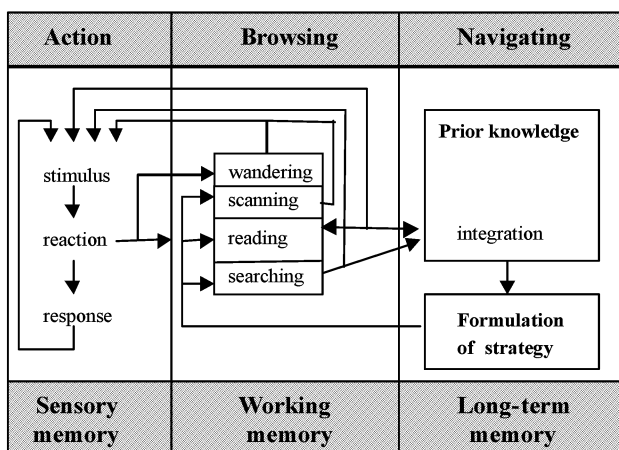

Figure 4. Framework for navigation

When the employed browsing strategy leads to integration between browsed information and prior knowledge, the user has achieved the last level, navigation. In other words navigation is concerned with learning about a space. The framework will be fully explained and discussed in forthcoming publications. In the future, the statistically defined strategies will be compared to pattern based strategies in more detail. To increase validity we will also use some qualitative methods to gather data from users' behaviour in hyperspace. The goal for this is to develop more exact and more easily traceable definitions for browsing strategies. The browsing profiler software will also be developed and used in the future studies. The main goal is to apply these methods and tools into an adaptive web-based course management system.

REFERENCES

Canter, D., Rivers, R. & Storrs, G. (1985). Characterising user navigation through complex data structures. *Behaviour and Information Technology*, *4*(2), 93-102.

Catledge, L. & Pitkov, J. (1995). Characterizing browsing strategies in the World-Wide-Web. In *Computer Networks and ISDN systems* (Vol 27), (pp.1065-1073), Darmstadt, Germany: Elsevier Science.

Lawless, K.A. & Kulikowich, J.M. (1998). Domain knowledge, interest, and hypertext navigation: A study of individual differences. *Journal of Educational Multimedia and Hypermedia*, 7(1), 51-70.

Lawless, K.A., Mills, R. & Brown, S.W. (2003). Children's Hypertext Navigation Strategies. *Journal of Research and Technology in Education*, *34*(3), 274-284.

McAleese, R. (1989). Navigation and Browsing in Hypertext. In R. McAleese (Ed.), *Hypertext theory into practice*. (pp.6-44). Oxford, UK: Intellect Limited.

Mullier, D., Hobbs, D. & Moore, D. (2002). Identifying and Using Hypermedia Browsing Patterns. *Journal of Educational Multimedia and Hypermedia*, *11*(1), 31-50.

Mullier, D. (1999) *The application of neural network and fuzzy-logic techniques to educational hypermedia.* Unpublished doctoral dissertation, Leeds Metropolitan University, United Kingdom. [online: www.lmu.ac.uk/ies/comp/staff/dmullier/].

Reed, W.M. & Oughton, J.M. (1997). Computer experience and interval based hypermedia navigation. *Journal of Research of Computing in Education*, *30*(1), 38-52.

Spence, R. (1999). A framework for navigation. *International Journal of Human-Computer Studies*. *51*, 919-945.

BIOGRAPHY

Kristian Kiili has research interests in the users' behaviour in digital learning materials. Recent research has focused on identifying browsing strategies and developing a theoretical model for navigation process. Harri Ketamo has research interests in educational technologies and user modelling. Recent research has focused on adaptive studying environments and developing network-based observation methods and tools.