From Browsing Behavior to Usability Matters

Eelco Herder & Betsy van Dijk

University of Twente, The Netherlands
P.O. Box 217, 7500 AE Enschede
{e.herder, e.m.a.g.vandijk}@cs.utwente.nl

Abstract

Many features of interest can be observed from metrics of user navigation, of site structure and of paths followed through the site. These metrics can be used for recognizing and predicting usability matters. This papers discusses models and techniques needed, and how they can be used for adapting navigation support to user needs in various contexts of use

1 Introduction

Online services and information units are typically organized in a hypermedia structure, consisting of pages and links that connect those pages in a non-linear way. These links form a site's *navigation infrastructure*, which enables users to identify their current position in the site structure, to distinguish among various options for moving on from this position and to actually make their way through the web site (Thüring, Hanneman & Haake, 1995). Obviously, users cannot follow non-existing links. In a similar way, automobilists are restricted to following the existing road infrastructure.

Highways are built to have people reach their destination as fast as possible. Local roads are usually more time-consuming, but more interesting. When one is in a hurry, it is frustrating that one can only choose a scenic route. Correspondingly, when you need to carry out online banking activities in limited time, you want to have immediate access to those services, instead of having to wade through a sea full of interesting, but currently irrelevant information.

When driving an old Volkswagen Beetle with uncomfortable seats, a blurred windshield and malfunctioning handles, you might want to reach your destination as fast as possible. Ironically, using this car prevents you from doing so. In a similar way, mobile devices provide limited means of interaction, and they are typically used while on the move and having to deal with time constraints. Studies have shown that mobile users display a more conservative link exploration behavior (Buyukkokten, Garcia-Molina, Paepcke & Winograd, 2000). This behavior, though it may seem efficient at first sight, is typically less effective than regularly returning to navigational landmarks (McEneaney, 2001).

Adapting the navigation infrastructure to the user and to the context of use is seen as a proper way of providing adequate navigation support for varying user needs (Brusilovsky, 2001). However, most web applications are designed for a prototypic user, accessing the service with a browser on a high-resolution monitor, using sophisticated means of interaction, ignoring the fact that mobile internet access is increasing (Menkhaus, 2001). The only way of providing personalized navigation support for these applications is by adapting the existing link structure, without requiring or expecting an author's effort for link annotation or the like.

In this paper we present an intermediary-based link adaptation strategy, making use of site structure and navigation paths. We explain the concept of *lostness in hyperspace*, which covers many important usability matters, and how these matters can be inferred from user navigation through a web site. We conclude with a brief description of an agent that acts as an intermediary between information providers and users

2 Lostness in Hyperspace

Being lost is considered as one of the fundamental difficulties which users experience when interacting with hypermedia systems. Therefore, much hypermedia research has been devoted to this issue. (Otter & Johnson, 2001). At some point users may not know where they are, how they got there or where they should go next – or, alternatively, they know *where* they want to go, but not *how* to get there (Thüring et al., 1995). These problems may be caused by *cognitive problems* – such as low spatial ability and memory capacity (Neerincx, Lindenberg & Pemberton, 2001), by *unfamiliarity* with the conceptual structure of a site (Otter & Johnson, 2001), by problems related to the *user context* – such as device characteristics, time constraints, task-set switching, by *improper orientation clues or navigation facilities* within an information domain (Thüring et al, 1995) or by a combination of these factors.

When lost, navigation behavior usually changes. Users might arrive at a particular point and forget what was to be done there, they might neglect to return from a digression, or they might miss some pages that contain relevant information (Otter & Johnson, 2001). As a result, navigation performance and user satisfaction will drop dramatically (Smith, 1996). Therefore, although the term 'lostness' applies to the interaction *process*, it is clearly related to the interaction's *outcome* (Otter & Johnson, 2001).

The problem of lostness is even worse for mobile devices. Because of the small screen size, users can be given less orientation clues than in a desktop setting. To make things worse, information displayed on small screens needs to be split in small chunks, thus requiring the user to select and scroll more intensively, which only adds to the problem. As a result, users may not have confidence in being able to use the resources to achieve their goals if it involves tedious and error-prone navigation (Jones, Buchanan, Marsden & Pazzani, 2001).

3 Metrics of Browsing Behavior

From user navigation many interesting features can be inferred, which indicate navigation strategies chosen and usability problems encountered. However, this is not feasible without proper models of user navigation and quantitative representation of its characteristics, which are presented shortly in this paragraph. A more complete overview is given in (Herder, 2002).

The node-and-link structure of online content lends itself to be modeled as a graph – the *site graph*. User navigation can be seen as an overlay of the site graph, consisting of only the pages visited and the links followed – the *navigation graph*. This navigation graph can contain information about one user session, about one particular user or about an entire user group. Each edge/link can be assigned a weight, which indicates the distance between the source node and the destination node. Several notions of distance can be applied, such as textual similarity and transition frequency – which expresses the probability a user will follow the link (Herder, 2002). The graph structure reveals many features which have their impact on user navigation.

On a detailed level, different types of pages can be recognized by the linkage to and from a page. Pirolli (Pirolli, Pitkow & Rao, 1996) classified pages in four main categories (home pages, index pages, reference pages and content pages). Index pages, for example, are usually quite small and have a large number of outgoing links. Kleinberg (Kleinberg, 1999) deploys the link structure for recognizing so-called *authorities* and *hubs* – pages that are seen as important and pages that link to many important pages respectively. Although links within a site can hardly be seen as reliable indicators of a page's significance, they do convey semantic relationships between pages. Following such a link can be regarded as recognizing the relationship as interesting.

Seen from a higher level, pages can be grouped into *clusters*, richly interconnected parts of the site graph dedicated to a topic with but a few links to other topics (Pirolli et al., 1996). Global characteristics of each cluster, or of each site as a whole, can be calculated based on the link structure. One of the most obvious measures is the size of a cluster or a site. *Cluster size* and the *average connected distance* (Botafogo, Riflin & Shneiderman, 1992) are indicators of the average navigation effort required to reach a certain page. For each individual page its *depth*, the distance from a navigational landmark, indicates the effort required to reach this page. The average *net density*, the ratio between the number of pages and the number of links, and the *distribution* of the links indicate the amount of freedom in navigation offered. The more links, the more complex an interaction will be. Complexity can be reduced by making a site more *linear* (Botafogo et al., 1992). Besides global characteristics, interesting spots within the graph can be located as well – such as groupings of navigation landmarks, cycles and central and peripheral areas.

Users have been shown to display different navigation strategies, varying from goal-directed to explorative (Shneiderman, 1997). We expect these strategies can be recognized from global characteristics of user navigation. These characteristics include *path length*, *path linearity* and *path density* (McEneaney, 2001) – the latter two measures indicate the amount of backtracking a user performs. Global characteristics as well as other interesting phenomena – such as a user navigating in cycles – can be discovered from the navigation graph itself. On the other hand, one needs to know the infrastructure in which a user is navigating to be able to successfully interpret the navigation. Did a user return to a navigational landmark or to a content page? Furthermore, it would be interesting to know the correlations between navigation behavior and site structure. Do more densely linked sites invite less linear navigation strategies? Does site linearity shorten navigation paths? Do link junctions or dead-ends lead to usability problems? Modeling user navigation as an overlay of site structure facilitates dealing with these issues, as it allows for comparison.

4 From Metrics to Usability Matters

As pointed out earlier, lostness is seen as a fundamental usability problem in hypermedia research. However, the term covers a wide range of user problems, which are unlikely to be captured by a single measure. Otter (Otter & Johnson, 2001) suggests that we could make use of a 'battery of measures' instead, which correlate well with one another and which have been shown through empirical studies and in the literature to measure lostness to some degree.

Unfortunately, contradictory results can be seen from the literature. Smith (Smith, 1996) bases her measures for lostness on the assumption that users who cannot locate information they require and which exists in the system, can be regarded as lost. She proposes a lostness measure based on the number of revisits and the number of navigation steps compared to the optimal route. The usefulness of this measure is said to be confirmed in an experimental setting. McEneaney (McEneaney, 2001), on the other hand, directly compares patterns of navigation and search outcome measures, therewith taking into account spatial and temporal features of user paths. From his experiments it turned out that less successful users adopted a "page turning" strategy, while subjects who did well on the search task often returned to navigational landmarks, such as a table of contents. These contradictory results cannot be explained by system goals, since both authors used a system that provided teenagers study advices.

Several explanations for these contradictory results can be thought of. Most importantly, neither McEneaney nor Smith related user navigation to the site structure. As argued above, navigation infrastructure directly influences users' navigation strategies. McEneaney mentioned the presence of a reading sequence using 'next' and 'previous' buttons in his system. Further, we do not know to what types of pages users returned, or how much time they spent on each page.

The metrics described in the previous paragraph provide data on user navigation within a site structure. From these data user navigation strategies and usability problems can be inferred. As mentioned before, users display various kinds of navigation strategies, varying from goal-directed activities to explorative browsing. These various navigation types call for different types of navigation assistance. Strongly hierarchically structured sites appear to be more suitable for goal-directed activities (Modjeska & Marsh, 1997), as they provide better structured orientation clues (Thüring et al., 1995). Cross-references do increase navigational efficiency and provide more freedom for exploration, but at the cost of some disorientation (Modjeska & Marsh, 1997).

Taking the strategy employed into account, problems related to lostness – as described in the second paragraph – can be predicted from the site structure (for example, too many cross-references) or observed from the navigation path. An early experiment should give insight in correlations between user path metrics, site metrics and interaction success measures. Since we are using numerous metrics for predicting usability matters from site structure and user navigation, machine learning techniques – such as Bayesian Learning (Zukerman & Albrecht, 2001) – appear to be useful for this purpose. The actual design of the learning algorithms will depend highly on the conclusions drawn from our experiment. Once the usability matters can be inferred reliably enough, one can seek suitable adaptation strategies to solve them.

5 From Usability Matters to Adaptation Strategies

In this paper we discussed how usability matters can be inferred from metrics of user browsing behavior in a site structure. This diagnostic task is the first step towards an intermediary agent that adapts the navigation infrastructure to user needs in various user contexts. Web crawling and web logging techniques are employed to gather information on site structure and user navigation. This information is needed for building the site graph and the user navigation graph. From both graphs features of interest can be deduced and relevant metrics can be calculated. Making use of these features and metrics, usability problems related to lostness in hyperspace can be predicted or recognized.

The next step is to find suitable adaptation strategies for the usability matters found. On page level, these adaptations can be created using common adaptive hypermedia techniques, such as link ordering, link hiding, link highlighting, link annotation and direct guidance (Brusilovsky, 2001). On site level, these adaptations can be seen as a personalized sub graph imposed on the site graph, which can be evaluated by comparing user navigation and the personalized site structure once again (Herder, 2002). Adaptive navigation support is an important means for enhancing user experience in mobile contexts, which is the main issue we are addressing in the PALS project [(1)] (Lindenberg, Nagata & Neerincx, 2003).

Footnote

1. PALS stands for Personal Assistant for online Services; the project is supported by the Dutch Innovative Research Program IOP-MMI. Our research partners are TNO Human Factors and the University of Utrecht.

6 References

Botafogo, R.A., Rivlin, E. & Shneiderman, B. (1992): Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. *ACM Transactions on Information Systems 10 (2)*, 142-180

Brusilovsky, P. (2001): Adaptive Hypermedia. *User Modeling and User-Adapted Interaction 11*, 87-110

Buyukkokten, O., Garcia-Molina, Paepcke, A & Winograd, T. (2000): Power Browser: Efficient Web Browsing for PDAs. *Proceedings of the CHI 2000 Conference*, The Hague, 430-437

Herder, E. (2002): Metrics for the Adaptation of Site Structure. *Proc. of the German Workshop on Adaptivity and User Modeling in Interactive Systems, ABIS02*, Hannover, 22-26

Jones, M., Buchanan, G., Marsden, G. & Pazzani, M. (2001): Improving Mobile Internet Usability. *Proc. of the 10th Intl. WWW Conference*, Hong Kong, 673-680.

Kleinberg, J. M. (1999): Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM 46 (5)*, 604-632

Lindenberg, J., Nagata, S. & Neerincx, M. (2003): Personal Assistant for Online Services: Addressing Attention, Emotion and Navigation. *Workshop on Human Information Processing and Web Navigation, HCI 2003*, Crete (to appear)

McEneaney, J. E. (2001): Graphic and numerical methods to assess navigation in hypertext. *Intl. Journal of Human-Computer Studies 55*, 761-786

Menkhaus, G. (2001): Architecture for client-independent Web-based applications. *IEEE Proceedings of the TOOLS-Europe Conference 2001*, Zürich, 2001

Modjeska, D. & Marsh, A. (1997): Structure and Memorability of Web Sites. *Working Paper in the Department of Industrial Engineering, University of Toronto*

Neerincx, M.A., Lindenberg, J. & Pemberton, S. (2001): Support Concepts for Web Navigation: a Cognitive Approach. *Proc. Of the 10th Intl. WWW Conference*, Hong Kong, 119-128

Otter, M. & Johnson, H. (2001): Lost in hyperspace: metrics and mental models. *Interacting with Computers 13 (1)*, 1-40

Pirolli, P., Pitkow, J. & Rao, R. (1996): Silk from a Sow's Ear: Extracting Usable Structures from the Web. *Proceedings of the CHI 1996,* Vancouver, 118-125

Shahabi, C., Zarkesh, A., Adibi, J. & Shah, V (1997): Knowledge Discovery from Users Web-page Navigation. *Proc. of 7th Int. Workshop on Research Issues in Data Eng. on High Performance Database Management for Large-Scale Applications*, Birmingham, UK

Shneiderman, B. (1997): Designing Information-Abundant Websites: Issues and Recommendations. *International Journal of Human-Computer Studies 47 (1)*

Smith, P.A. (1996): Towards a practical measure of hypertext usability. *Interacting with Computers 8 (4)*, 365-381

Thüring, M., Hanneman, J. & Haake, J.M. (1995): Hypermedia and Cognition: Designing for Comprehension. *Communications of the ACM 38 (8),* 57-66

Zukerman, I. & Albrecht, D.W. (2001): Predictive Statistical Models for User Modeling. *User Modeling and User-Adapted Interaction 11*, 5-18