Sniffing Around for Providing Navigation Assistance

Eelco Herder
Group of Human Media Interaction, Department of Computer Science
University of Twente, P.O. Box 217, 7500 AE, Enschede – The Netherlands
herder@cs.utwente.nl

Abstract

In this paper we describe an approach to adaptive navigation assistance that is meant to enhance a user's information scent. The navigation assistance is composed of a combination of predictive user navigation modeling and common information retrieval methods. Besides assistance in forward browsing, the assistant helps users in deciding when to switch to searching or backtracking, while taking their navigation preferences into account.

1 Introduction

Imagine yourself looking on the Internet for information on a research topic that you are not completely familiar with. Most likely you will start with entering some keywords in a search engine, such as Google. Then you pick the apparently most promising result. The first page partially satisfies your information need, but perhaps more information is available on other pages. You select some links that might be helpful. While browsing through the site, you realize that it might be better to go back to Google and perform a more refined keyword search. Or perhaps you should dig a bit further? You cannot do both at the same time. There is even another option: you can return to a page that you visited before, which contained some pointers that in retrospective might have been more promising than what you have found thus far.

Searching and *browsing* are the two predominant patterns for finding information on the Internet [Olston & Chi, 2003]. Searching is the process of locating information by issuing queries in a search engine; browsing is the process of viewing web pages and navigating between them using hyperlinks. In addition, browsers provide means to *backtrack* to pages visited earlier. Searching is particularly useful for obtaining quick results from a broad range of sources. Browsing is more useful when it is hard to express the information need in some keywords; moreover, a great deal of information and context is obtained along the browsing path itself. Backtracking is employed for reviewing pages visited before, either for reference or as a starting point for an alternative path [Tauscher & Greenberg, 1997]. Users typically alternate between these three patterns, constantly evaluating the benefits of browsing within the context of a site, returning to the search engine and backtracking to earlier results.

Users have to actively choose between searching, browsing and backtracking, making use of incomplete cues that are different for each pattern. Navigation assistance, such as personal assistants and adaptive hypermedia techniques [Brusilovsky, 2001], typically supports the user in exploration by browsing, ignoring the options of keyword search or backtracking. In this paper we describe an approach to adaptive navigation assistance that supports the user in comparing the benefits of browsing, searching and backtracking actions by 'sniffing around a bit further and back' than users could do themselves.

The main goal of the framework is to provide the user with enhanced feedback and an integrated view on the possible navigation actions in the three patterns mentioned before; we explicitly aim to help users in locating information themselves, as suggestions 'out of the blue' – such as search results – do not provide the context that users would learn by navigating themselves [Pirolli & Fu, 2003]. We also take into account that certain categories of users are more willing and more apt to deal with context switching and backtracking [Chen & Macredie, 2002] by matching the assistance to the user's navigation preferences.

This remainder of this paper is structured as follows. We start with an introduction to a predictive model of user web navigation, the *information foraging theory* [Pirolli & Card, 1999], an overview of some information foraging-based systems and a motivation why backtracking should be taken into account when applying information foraging to web navigation. Then, we present an approach to navigation assistance that combines information foraging theory with common information retrieval techniques. Next, we discuss interface considerations that need to be taken into account. Finally, we discuss open issues and describe future work.

2 Information Foraging on the Web

In this section we introduce a predictive model of users seeking and selecting information, the theory of *information foraging*. Then we describe some systems that apply information foraging theory to user web navigation. We conclude with an argument why backtracking should be taken into account when applying information foraging to web navigation.

2.1 Information Foraging Theory

In 1999 Pirolli and Card [1999] published an influential paper on a predictive theory on how users seek and select information in a – possibly large – information domain. The name *information foraging theory* reflects the similarities in behavior of users looking for information and predators foraging for prey. In this subsection we present a brief overview of the theory, with an emphasis on notions that will be used later on in this paper.

Information relevant to a person's information needs may reside in some books in the library, or in (sections of) certain websites. In a similar way, prey animals typically live in groups; predators can choose between several hunting grounds – '*patches'*. As predators want to maximize the ratio between their take and energy spent, they will remain foraging within a patch until it might be more profitable to spend time and energy on migrating to another patch. As we have seen in the introduction, information foragers face the same trade-off; when would it be better to leave a site for a search engine, or to backtrack. Like predators, users try to maximize their *rate of gain* (R), which is the ratio between the *total amount of information gained* (G), divided by the total amount of *time spent between patches* (T_B) and *exploiting within patches* (T_W),

$$R = \frac{G}{T_B + T_W} \quad \text{Information-value-units/cost units} \qquad (1)$$

With the above formula (1) as a starting point, Pirolli and Card analyze the effects of users' *time allocation* decisions (e.g. to spend more time within a patch) and information *diet selection* on the rate of gain. As the total rate of gain is a summation of all user actions and the information acquired by these actions, the formula can also be applied to predict the rate of gain for each individual action.

Information foraging theory provides a theoretical framework and utility functions for production-systems that simulate information foragers, such as web users. These systems usually estimate the expected information value of pages at the other site of the links by analyzing the *proximal cues* (e.g. link texts, link images) [Chi et al., 2001], just like real-life web users would do. The expected rate of gain is called *information scent*, which suggests that information foragers – like predators – follow their noses. As we will see in the following subsection, information foraging has been used as a basis for adaptive hypermedia systems as well, despite its mainly descriptive nature.

2.2 Predictive Web Navigation Models

Despite its level of detail, information foraging theory does not enforce the use of specific techniques for determining the information value of a document or specific definitions of between patch-time and within patch-time, as these may vary per domain. In this subsection we describe some information foraging-based systems that model user navigation on the Internet.

SNIF-ACT [Pirolli & Fu, 2003] models users working on unfamiliar information-seeking tasks on the Internet. The information scent of link texts, given the current task, is based on word occurrences and co-occurrences in a corpus for text retrieval. The system is reported to be able to generate good predictions of user navigation actions. Moreover, the results clearly showed a decrease of a site's information scent, just before users left the site.

In [Chi et al., 2003] two computational methods are described for understanding the relationship between user needs and user actions. The first method, WUFIS, simulates user navigation actions, given a task, which is represented as a keyword vector. Standard tf.idf weighting is used for calculating the cosine similarity between link labels and the task representation [Grossman & Fieder, 1998]. The second method, IUNIS, attempts to infer the user needs from a user's actions. Both methods are reported to work reasonably well.

ScentTrails [Olston & Chi, 2003] is an adaptive hypermedia system that annotates hyperlinks with information scent, given some search keywords supplied by the user. Unlike the systems mentioned previously, ScentTrails does not compare the keywords with the *proximal cues* but with the target documents themselves. Moreover, the information scent of a page propagates to neighboring pages, with a certain decay parameter. Metaphorically speaking, the system sniffs around a bit further than a user would do and creates some high-scent trails.

2.3 Taking Backtracking Into Account

The main focus of information foraging-based systems is forward navigation. Surprisingly little attention has been given to backward navigation, such as the use of the back button and page revisits in general. In an alternative model of web navigation, CoLiDeS [Kitajima et al., 2000], backtracking is regarded as an activity that takes place when forward search fails. However, this does not correspond with how real users navigate; users often return to pages to continue along an alternative path [Herder & Van Dijk, 2004]. These pages are often *navigational hubs*, such as a site's home page or a list of search results. Effective use of these hubs is reported to be an effective navigation strategy, more effective than linear forward navigation. In a pilot study [Herder & Juvina, 2004] we identified a highly non-linear *laborious navigation style* that users successfully employed for exploring and understanding unknown site structures. Users who mainly navigated in a forward direction performed worse on the tasks that they were given. This provides evidence that backtracking to pages with high information scent is an important aspect of user navigation that should be recognized and exploited in a system that supports users in finding the information that they need.

3 Enhanced Scent-Based Navigation Assistance

In this section we describe an approach to adaptive navigation assistance that is based on concepts of the information foraging theory. The assistant assumes that the information need is known and expressed as a keyword vector, like a query in a search engine. In short, the assistant builds a predictive model of the users' navigation actions. Simultaneously, it explores the navigation options in three directions – searching, browsing and backtracking – and builds its own information scent model, enhanced with additional information from the target documents' contents. The assistant's enhanced scent model is compared with the predictive navigation model to match the assistance with the user's navigation style. How the assistance is presented to the user will be dealt with in the next section.

3.1 Modeling the User's Information Scent

As argued before, users can choose between three categories of navigation actions: browsing, search and backtracking. Within these categories, users can choose a specific link, search result or previously visited page. In this subsection we propose a predictive model of user web navigation, based on information scent and the past navigation actions, as observed from the web usage logs.

Whereas each navigation action leads to a web page, the information scent differs between categories, as the cues are different:

- for browsing actions users base their decisions on the link descriptions (e.g. link text, image);
- for searching actions users base their decisions on the result descriptions, which is typically the page title and a snippet of the page contents;
- for backtracking actions users base their decisions on what they remember of a page visited before.

For the moment we assume that the user's information need – the *query* – is known and expressed as a keyword vector, as is the case in the WUFIS algorithm (section 2.2). We define the information scent for each cue as the cosine similarity [Grossman & Fieder, 1998] between the query and *proximal cue* – link labels, search result descriptions and the contents of previously visited pages.

For backtracking actions we compensate for the recency effect: recently visited pages are more likely to be revisited than pages visited earlier. The results of [Tauscher & Greenberg, 1997] suggest a power law distribution. We estimate the recency effect with data from the web usage log: the percentage of revisits at distance *d* among all observed revisits. This situation is sketched in figure 1.

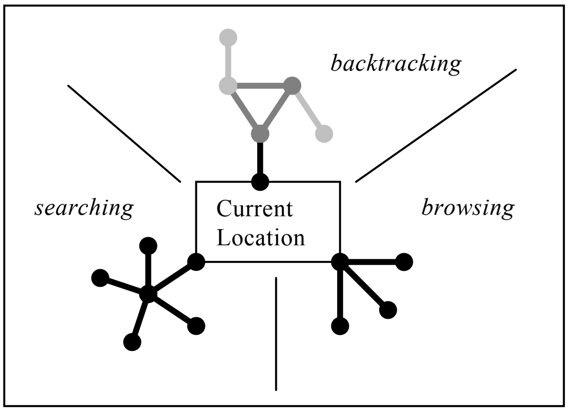

Figure 1 – The user's information scent. From the current location a user can choose between searching, browsing and backtracking. For (forward) search and browse actions, the available cues are the hyperlinks and search results; backtracking actions lead to pages visited before, of which the user might remember part of the contents.

With the measure for the expected information gain, constructed as described above, we can estimate the expected costs for each action. As users only have access to the proximal cues, they cannot guess how much time they will spend on the next page. We assume that the expected costs differ not as much between actions within one category of navigation actions as between the action categories and that they are reflected in the user's preference for each action category. Therefore, we take as the action cost the fraction of actions in this category and all navigation actions, inversed and normalized so that the mean action cost equals one.

As a result, we have a formula (2) for the expected rate of gain for each navigation action, which is the cosine similarity between query and cue, divided by the associated action cost. For browsing and search actions the *recency effec*t is set to 1.

$$R = \frac{similarity\ (query\,, cue)}{actionCost} recencyEffect \qquad (2)$$

In accordance with the information theory, users will try and optimize their information diet and time allocation to increase the rate of information gain – the information scent – and most likely choose the action with the maximum rate of gain.

3.2 Enhanced Information Scent by Sniffing Around

As might have become clear from the previous subsection, the user's information scent is far from perfect. In particular, it is hard to compare the information scent between navigation action categories. An automated navigation assistant can look ahead while a user is reading and use the documents' distal content as cues instead of the proximal cues that users have to deal with. The navigation assistant also does not suffer from the recency effect with respect to backtracking. The basic idea is that the browsing assistant acts as a *scout* that sniffs around in three directions:

- the current (browsing) local context;
- the global (search) context – e.g. Google search;
- the past (backtracking) context.

As with the user's information scent we assume that the user need is expressed as a keyword vector. We define the information scent as the cosine similarity of this vector with the document contents. As the scouting process does not require time from the user, we use an uniform cost for each action; from the perspective of the browsing assistant, the information scent is equal to the information gain – i.e. formula 2 with both actionCost and recencyEffect set to 1.

For reasons that were mentioned in the introduction, we chose primarily for an interface that does not provide the user with shortcuts to pages that are further than one click away in forward (search or browsing) direction. As the assistant can look more than one click ahead, it would be desirable to propagate the information scent of more distant pages to the pages that are directly linked from the user's current location. This can be achieved by recursively adding the summed information scent of all pages that can be reached from a page *p* to its page *p*'s information scent, with a discount parameter γ, $0 \leq \gamma \leq 1$, which accounts for the navigation effort the user has to spend [Olston & Chi, 2003][Rennie & McCallum, 1999] (formula 3)

$$R_{p\_total} = R_p + \gamma \sum_{i \in linkedFrom(p)} R_i \qquad (3)$$

In summary, the navigation assistant's enhanced information scent is calculated in a similar way as the user's proximal information scent, but with different loadings for the similarity measure, action cost and recency effect.

3.3 Comparing User's Information Scent and System Information Scent

The adaptive navigation assistance framework has two expected information scent values for each action that a user can take: one that is based on the limited proximal scent of the user and one that is based on the enhanced scent of the scouting assistant. The most straightforward way of presenting the utility of the navigation actions is simply to take the enhanced information scent. However, in certain cases it might be desirable to favor second-best options above better ones, to better match the user's navigation preferences.

If the information scent is high in both the predictive user model and the assistant's enhanced scent model, it is sufficient to confirm the user's expectations. If the assistant's information scent is higher than the user's, the following cases can be distinguished:

- if the difference is caused by a significantly lower information gain (the cosine similarity coefficient), it might be a good idea to add cues to indicate that the action is more useful than it appears to be;
- if the difference is caused by a high expected action cost, more caution is needed. The user might prefer not to switch to a different site, not to continue browsing or not to backtrack. In that particular case the user's expected action costs can be used in the assistant's model to take these preferences into account. This might lead to higher scent values for second-best options, which might be more beneficial to the user, as they better match the navigation preference.

After the procedure described above the assistant has a ranking of the navigation actions that can be presented to the user. The assistant also has an indication how this ranking matches the user's expectations. In the next section we describe alternative user interface concepts that make use of this knowledge.

4 Interface Considerations

In the previous section we assumed that the user's information need is known and expressed in the form of a set of keywords. As most users are more or less familiar with search engines, it suffices to add a query field to the browser's interface, similar to the Google Toolbar, a popular add-on to the Internet Explorer. While browsing, users can refine their query by adding, changing or deleting keywords.

As the user's information scent is based on visible cues, it is desirable to integrate the assistant's ranking into the browser window rather than to display them in a separate window. Relevancy indicators for forward browsing actions can be presented in the form of link annotations [Brusilovsky, 2001]. In the ScentTrails system [Olston & Chi, 2003] the font size of favored links was increased. It turned out that this method frequently distorted the page layout and introduced ambiguity, as users generally did not know the original font sizes. Other annotation methods that might be considered are changing the color of the text, its background or adding a border to the link. The latter method can also be used for annotating images that serve as links. Additional text, such as a snippet from the link destination, can be added as a roll-over pop-up window [Weinreich et al., 2001].

Additional navigation actions, which include links to pages that are the result of searching and backtracking, might best be presented in a separate region, directly above or below the page content. Candidates for displaying are the page titles and snippets from their contents. The background color can be used to indicate the favorability of searching or backtracking in comparison with browsing (e.g. a lighter background means more favorable).

In addition to the integrated interface, a graphical overview of the user's context can be provided. These visualizations are typically tree structures or graphs. Similar to the link annotation, the nodes that represent the pages can be color-coded according to their ranking [Herder & Van Dijk, 2004]. Separate regions for browsing, search and backtracking – as visualized in figure 1 – can be used for visually distinguishing between these options.

5 Discussion and Future Work

In this paper we described an approach to adaptive navigation assistance that integrates support for browsing, searching and backtracking actions. The assistance is based on the concept of *information foraging* and is meant to provide users with an enhanced information scent while taking their navigation preferences into account. The concepts are yet only partially implemented and there are many open issues.

The cosine similarity between the user information need, represented as a keyword vector, and the proximal cues or the documents contents, plays an important role in calculating the information scent. The calculation of this measure is highly dependent of the set of documents used as reference [Grossman & Fieder, 1998]. Thus far we have reached promising results with a set of only forty documents – which is typically the amount that will be gathered by the information sniffing scout. However, in order to reduce random effects, it might be a good idea to add a set of standard documents to the corpus.

At the moment the navigation assistant looks simply two links ahead in forward direction. We expect that a more flexible method, for example based on the proximal information scent, will lead to better results; more promising paths can be explored further and irrelevant paths can be cut after the first step. Similar approaches are being used by several web crawlers [e.g. Rennie & McCallum, 1999]. We plan to explore this by making use of web log data from earlier laboratory user studies with well-defined tasks.

The concept of combining the user's proximal information scent and the assistant's enhanced scent to match the user's navigation preferences sounds intuitive and attractive. Whether it actually increases user performance more than simply presenting the 'best' results needs to be evaluated. We plan to evaluate three versions of the navigation assistant, all with similar interfaces:

- a version that bases its advices on the user's proximal information scent;
- a version that bases its advices on the assistant's enhanced information scent;
- a version that combines both the proximal and the enhanced information scent.

Most projects on adaptive navigation assistance strive to improve collaborative filtering and page ranking algorithms. In contrast, our devised assistant is 'imperfect by design' to better match guidance to the users' natural navigation styles. However, it is up to the users themselves to decide how tactfully the assistant should behave.

Acknowledgements

The research presented in this paper takes place in the context of the PALS Anywhere Project, which is sponsored by the Dutch Innovative Research Program IOP-MMI.

References

[Brusilovsky, 2001] Brusilovsky, P. Adaptive Hypermedia. *User Modeling and User-Adapted Interaction 11*, 2001, 87-110.

[Chen & Macredie, 2002] Chen, S.Y & Macredie, R.D. Cognitive Styles and Hypermedia Navigation: Development of a Learning Model. *J. of the American Soc. For Information Science and Technology 53 (1)*, 2002, 3-15.

[Chi et al., 2001] Chi, E.H., Pirolli, P. & Pitkow, J. Using Information Scent to Model User Information Needs and Actions on the Web. *Proc. SigCHI 2001*.

[Grossman & Fieder, 1998] Grossman, D.A. & Frieder, O. *Information Retrieval – Algorithms and Heuristics*. Kluwer Academic Publishers, 1998.

[Herder & Van Dijk, 2004] Herder, E. & Van Dijk, B. Site Structure and User Navigation: Models, Measures and Methods. *Adaptable and Adaptive Hypermedia Systems*, to appear.

[Herder & Juvina, 2004] Herder, E. & Juvina, I. Discovery of Individual User Navigation Styles. *Proc. Workshop on Individual Differences in Adaptive Hypermedia Systems at AH2004*, to appear.

[Kitajima et al., 2000] Kitajima, M., Blackmon, M.H. * Polson, P.G. A Comprehension-based Model of Web Navigation and Its Applications to Web Usability Analysis. *Proc. HCI2000*, 357-373.

[Olston & Chi, 2003] Olston, C. & Chi, E.H. Scent-Trails: Integrating Browsing and Searching on the Web. *ACM Trans. On Computer-Human Interaction 10 (3)*, 2003, 177-197.

[Pirolli & Card, 1999] Pirolli, P. & Card, S.K. Information Foraging. *Psychological Review 105 (1)*, 1999, 58-82.

[Pirolli & Fu, 2003] Pirolli, P. & Fu, W.T. SNIF-ACT: A Model of Information Foraging on the World Wide Web. *Proc. User Modeling 2003*.

[Rennie & McCallum, 1999] Rennie, J. & McCallum, A. Efficient Web Spidering with Reinforcement Learning. *Proc. ICML 1999*, 1999.

[Tauscher & Greenberg, 1997] Tauscher, L. & Greenberg, S. How people revisit web pages: empirical findings and implications for the design of history systems. *Int. J. Human-Computer Studies 47*, 1997, 97-137.

[Weinreich et al., 2001] Weinreich, H., Obendorf, H. & Lamersdorf, W. The Look of the Link - Concepts for the User Interface of Extended Hyperlinks. *Proc. 12th ACM Conf. on Hypertext and Hypermedia*, 2001.