

A Web Site Navigation Engine

Mark Levene
Department of Computer Science
Birkbeck College, University of London
London WC1E 7HX, U.K.
M.Levene@dcs.bbk.ac.uk

Richard Wheeldon
Department of Computer Science
University College London
London WC1E 6BT, U.K.
R.Wheeldon@cs.ucl.ac.uk

1. INTRODUCTION

Often users navigating (or “surfing”) through a Web site “get lost in hyperspace”, when they lose the context in which they are browsing, and are unsure how to proceed in terms of satisfying their original goal [4]. The unresolved problem in Web site usability, of assisting users in finding their way, is termed the *navigation problem*. (See [3] for a survey and critique on the navigation problem.) This problem is becoming even more acute with the continuing growth of Web sites in terms of their structure, which is becoming more complex, and the vast amount of information they intend to deliver. In contrast users are not willing to invest time to learn this structure and expect delivery of the relevant content without delay.

Although global search engines such as AltaVista, Fast and Google are successful in directing users to the appropriate Web sites, finding the information on the actual sites is often very problematic due to the navigation problem discussed above. As a result, site-specific search is becoming increasingly important as evidenced by a recent Forrester survey [1], which reported that most Web site managers consider search to be a critical factor of their site’s functionality. However, as reported in [1] current site-specific search engines fail to deliver all the relevant content, return too much irrelevant content and do not rank the pages returned according to the user’s information needs. One of the reasons for the failure of current site-specific search engines is that they are not able to pick up the “scent of information” [5], that is, they fail to detect the hyperlinks that, when followed, lead to the information that the user is looking for. Moreover, Hearst [2] argues that next generation site-specific search engines should support the user’s tasks by helping to cut down the relevant routes the user should follow given an information need.

To tackle these problems we are developing a navigation system for semi-automating user navigation which builds *trails* of information, i.e. sequences of linked pages, which are relevant to the user query. (We will assume that a query is simply a set of keywords representing the user’s goal at the start of the navigation session.) The preferred trails are presented to the users in a tree-like structure which they can interact with. This is in sharp contrast to a search engine which merely outputs a list of pages which are relevant to the user query without addressing the problem of which trail the user should follow. We discuss the architecture of the navigation system and give a brief description of the navigation engine and user interface. The navigation system is being developed by NavigationZone Limited, which is a recent startup company currently lo-

cated in the Computer Science Department at Birkbeck College, University of London. We are about to release the first alpha version of the system, which is implemented to execute within Microsoft Internet Explorer.

2. ARCHITECTURE OF THE NAVIGATION ENGINE

The architecture of the navigation system is shown in Figure 1. The *user interface* executes within a Web browser such as Microsoft Internet Explorer or Netscape Navigator. It obtains the preferred trails for navigation, given a user query, from the *navigation engine* and requests pages for display from the Web site via a proxy.

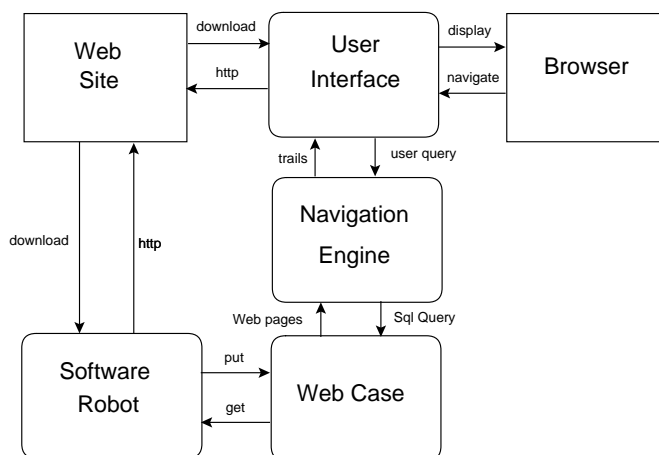


Figure 1: Navigation engine architecture

The navigation engine consists of two main modules: (i) the information retrieval module, and (ii) the best trail module. The information retrieval module does conventional information retrieval over Web pages combined with a novel page ranking algorithm that takes into account the potential of a Web page as a starting point for navigation. The best trail module computes the preferred trails for navigation given the input query. The algorithm is adaptive in the sense that it dynamically searches for the preferred trails by mimicking a user navigation session and scoring the trails as they are expanded according to the topology of the Web site. (See [3] for some detail of the algorithmic process used by the navigation engine to compute the preferred trails given a user query.)

The navigation engine interacts with a *Web case*, which is a relational database for storing the details of the pages of

the Web site in preprocessed form. The *software robot* is an offline process which is responsible for the creation of the Web case.

3. USER INTERFACE

The user interface includes two main mechanisms: (i) the *navigation tool bar* and (ii) the *navigation tree window*. The navigation tool bar contains a sequence of hyperlinks, where the hyperlink for the Web page currently being displayed is underlined. The hyperlinks to its right show the best trail found from the current page, providing a recommendation mechanism, and the hyperlinks to its left refer to the pages that the user previously visited, thus providing a history mechanism. Putting the cursor over a hyperlink causes a popup to appear which displays the title of the page, its URL, a summary of the page's content and other useful information pertaining to the page (cf. [6]). In addition, all hyperlinks are clickable and cause the navigation tool bar to be updated accordingly. The navigation tree window displays the preferred trails given the user query, organised in the form of a tree structure with the trails being ranked from the most preferred, according to their score. The user interacting with the navigation tree window can select any Web page on one of the preferred trails causing the page to be displayed. The navigation tool bar, the navigation window and the browser window are all synchronised according to the current page. The mechanisms of the user interface provide the user with guidance and context throughout a navigation session, given the input query. The user interface can be embodied in a Web site as a navigation mechanism complementing or replacing a Web site search engine.

For the alpha release of the system we have created a Web case for the Web site of University College London (UCL), whose domain name is ucl.ac.uk. Below we show two screen shots for the query "computers". The first one, shown in Figure 2, is the output from our navigation engine, and the second one, shown in Figure 3, is the output from UCL's local search engine. The results from the navigation engine help the user to find the context within the Web site of the information pertaining to "computers" by displaying the trails which are relevant to this query. For example, the first trail indicates that the Information Systems Division (ISD) at UCL provides computing services for students, while the second trail implies that ISD provides a help desk facility including Macintosh support. The search engine's results do not provide this contextual information. The best it can do for this type of query is provide a good starting point for the user to initiate a navigation session.

4. REFERENCES

- [1] P. Hagen, H. Manning, and Y. Paul. Must search stink? The Forrester report, Forrester, June 2000.
- [2] M. Hearst. Next generation web search: Setting our sites. *Bulletin of the technical Committee on Data Engineering*, 23:38-48, 2000.
- [3] M. Levene and G. Loizou. Web interaction and the navigation problem in hypertext. In A. Kent, J. Williams, and C. Hall, editors, *Encyclopedia of Microcomputers*. Marcel Dekker, New York, NY, 2001. To appear.

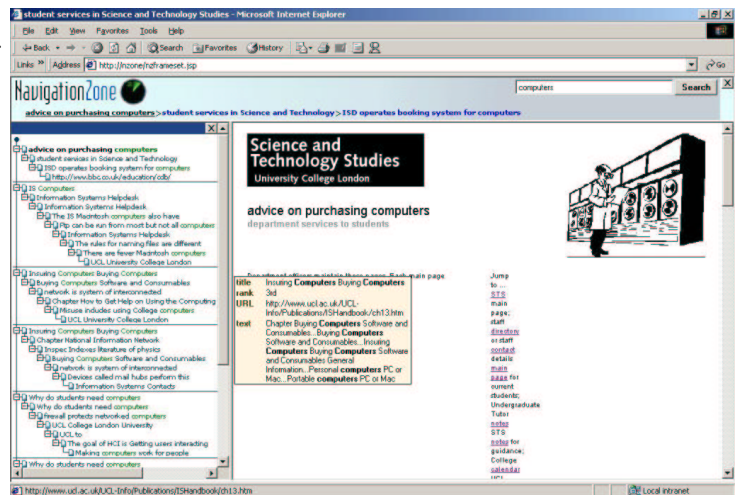


Figure 2: Navigation engine results for query "computers"

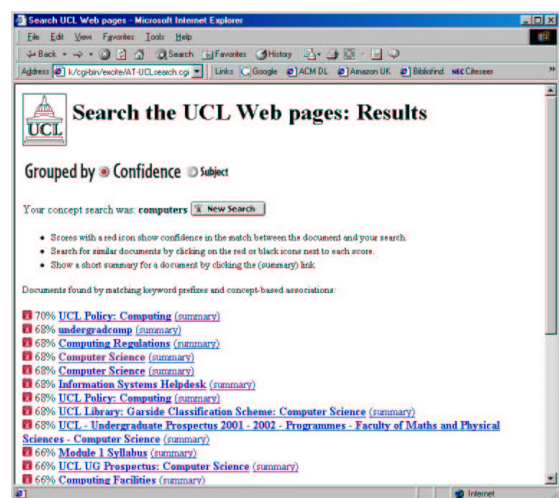


Figure 3: UCL search engine results for query "computers"

- [4] J. Nielsen. *Designing Web Usability: The Practice of Simplicity*. New Riders Publishing, Indianapolis, Indiana, 2000.
- [5] P. Pirolli. Computational models of information scent-following in a very large browsable text collection. In *Proceedings of ACM Conference on Human Factors in Computing Systems*, pages 3-10, Atlanta, Georgia, 1997.
- [6] H. Weinreich and W. Lamersdorf. Concepts for improved visualization of web link attributes. *Computer Networks and ISDN Systems*, 33:403-416, 2000.