Forward, Back and Home Again

Analyzing User Behavior on the Web

Eelco Herder

Graduation Committee

Prof.dr.ir. W.H.M. Zijm (University of Twente), *chairman and secretary*
Prof.dr.ir. A. Nijholt (University of Twente), *promotor*
Dr. E.M.A.G. van Dijk, (University of Twente), *assistant promotor*
Prof.dr. P.M.E. De Bra (Eindhoven University of Technology)
Prof.dr. N. Henze (University of Hannover)
Prof.dr. T.W.C. Huibers (University of Twente)
Prof.dr. M.A. Neerincx (Delft University of Technology)
Dr. S. Weibelzahl (National College of Ireland)
Dr. J. Zwiers (University of Twente)

CTIT Ph.D. Thesis Series No. 06-83. Centre for Telematics and Information Technology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

SIKS Dissertation Series No. 2006-08. The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

The research reported in this thesis was carried out in the project Personal Assistance for onLine Services (PALS) Anywhere. This project was supported by grant MMI0122NC from the Dutch Ministry of Economic Affairs, as part of the Dutch Innovative Research Program IOP-MMI.

Herder, Eelco
Forward, Back and Home Again - Analyzing User Behavior on the Web
Ph.D. Thesis, University of Twente, 2006
Includes bibliographical references
Cover design by Eelco Herder
Printed by F&N Boekservice, Amsterdam

ISBN-10: 90-73838-73-8
ISBN-13: 978-90-73838-73-4
ISSN: 1381-3617, No. 06-83 (CTIT Ph.D. Thesis Series)

FORWARD, BACK AND HOME AGAIN

ANALYZING USER BEHAVIOR ON THE WEB

Dissertation

to obtain
the doctor's degree at the University of Twente,
on the authority of the rector magnificus,
prof.dr. W.H.M. Zijm,
on account of the decision of the graduation committee,
to be publicly defended
on Thursday April 13, 2006 at 16.45

by

Eelco Herder

born on December 14, 1977

in Leeuwarden, The Netherlands

This dissertation is approved by the promotor,
prof.dr.ir. A. Nijholt,
and by the assistant promotor,
dr. E.M.A.G. van Dijk.

Preface

Many people from my generation and younger can hardly imagine what life would be like without the World Wide Web. Currently, a large population uses the Web for keeping up-to-date, planning trips, communication, entertainment, shopping, banking, and many other activities on a daily basis. At first sight, the Web interface is extremely simple and intuitive: one types in a Web address or a search query, and there you go. Whereas this might be true for many relatively simple activities, many issues arise when working on more complex, unfamiliar, or interrelated tasks. A large number of users would still feel extremely challenged if they would have to arrange all necessary steps for booking a trip to London - flight, hotel, travel plan, theatre tickets, restaurants, attractions - online. And even though many users might not know it, they often employ coping strategies in their Web interaction.

In October 2001, I started working as a PhD student on the PALS Anywhere project, which was aimed at developing concepts for personal assistance for online services. In the initial phase of the project it became clear that existing knowledge on how users interact with online information and services on the Web, was quite limited. Even though several user studies have been carried out, and even though research in the fields of adaptive hypermedia and user modeling is specifically aimed at understanding users and their actions, no answers - or conflicting answers - could be found for even very obvious questions. To me, it was intriguing that so little was known on activities that we carry out on a daily basis. This is what sprung my interest in the analysis of user interaction with the Web.

Given the infancy of the Web, which exists only since the early 1990s, it is not surprising that many interface concepts are not yet fully developed, even more since the Web is still subject to constant change in terms of technologies, services and usage. However, the differences in how users approach similar tasks, and the differences in performance, were striking. Even more, a closer analysis of interaction patterns revealed aspects that might seem completely obvious in retrospective, but which have apparently been left unnoticed and which were

clearly not supported by any interface concept. This thesis reflects my exploration of these lesser known yet interesting things in everyday life, an exploration which I definitely could not have carried out on my own.

There are many people who I would like to thank for their direct or indirect contributions to this thesis, to the research behind the thesis, and to the life behind the research. Undoubtedly, if I would try and mention all of them, I would forget to include some names that should be included, or I might even include some names that should be forgotten. For this reason, I will refrain from mentioning names, with a few exceptions.

First of all I would like to thank my supervisors for their support, the promotion committee for their involvement, and my colleagues at the Human Media Interaction group for a large variety of reasons. Further acknowledgments go to my fellow PALS project members from the University of Twente, Utrecht University, the TNO research institute in Soesterberg, and the companies involved in the project. Interesting ideas, broader insights, and valuable feedback were given to me by various researchers that I met at conferences and workshops.

I have been fortunate to have had a number of close collaborations. In Ion Juvina from Utrecht University I found a partner for the two laboratory studies reported in this thesis. Our substantially differing backgrounds often resulted in good discussions and new insights, and his experience in statistics and experiment design proved to be invaluable and instructive. For the long-term study I found partners in Harald Weinreich, Matthias Mayer and Hartmut Obendorf from the University of Hamburg. The collaboration was intensive, at some points accompanied by frustration and arguments. For the most part, however, the cooperation was extremely pleasant and fruitful. I also would like to thank all those who participated in the user studies.

Much support and motivation has also come from my family and friends, who enriched my life behind the research. I intensely enjoyed the rich cultural life at the University Campus. I made many friends at the classical student choir Drienerloos Vocaal Ensemble, and I will sorely miss the rehearsals, concerts, trips, social events, and many other activities - some better be left unmentioned. As a cofounder of the piano club Utopiano - one of a kind in the Netherlands - I would like to thank all who organized, participated in, or facilitated its activities; the concert registrations will remain a precious keepsake. I am happy that my two closest friends, Arend de Haan and Maarten Gijselaar, will stand by me as paranymphs during the defense of this thesis. Just not to disappoint anyone, thanks to all friends yet unmentioned. And, last but not least, I would like to express my gratitude to my father, mother and brother - Oebele, Betty and Rudmer; it feels great that 'Huize Beemdgras' is still a place that I call home.

Hannover
March, 2006

Eelco Herder

Contents

Chapter 1

Introduction

Since its introduction in the early 1990s, the World Wide Web has rapidly grown, both in size and in user population; currently about one billion users go online (Nielsen 2005*b*). The main activities carried out by these users include general browsing, reading news, shopping, entertainment, online banking, travel planning, and gathering information relevant to their work, hobbies and daily life. According to a recent survey, a vast majority of the American population considers the Web as the most important source of information, in private and work situations alike (Center for the Digital Future 2005).

Even though in general users are satisfied with the functionality the Web provides, support for these activities can be greatly improved. In particular collecting information on a particular topic can be a complex task that requires reading a large amount of material, and multiple refinement steps. During this process, users need to keep track of the material collected thus far and to plan their future actions. Enhanced context information and personalization techniques are expected to provide users with tools that better support their needs. Users also often return to information or services found before. Current history mechanisms - the back button, bookmarks and the history lists - are often criticized in the literature (Cockburn & McKenzie 2001) and users often resort to alternative strategies for relocating information found before (Teevan, Alvarado, Ackerman & Karger 2004).

In order to better support the users in locating and relocating information and services on the Web, it is important to know how users currently interact with the Web; what tasks they carry out, and what tools they use. As the Web is used by a wide variety of users who carry out even a wider variety of activities - varying from goal-oriented to exploratory, leisure browsing -, it is most unlikely that all this can be covered in a general and unified theory of Web navigation (Dix, Finlay, Abowd & Beale 2004). Moreover, the Web is subject to constant change. New technologies and services have emerged that allow for dynamically updated site contents, and server-side applications that provide interactive tools and services.

Many of these changes could not be foreseen, and many of these changes have their impact on the way the Web is used. This dynamics is reflected in the constantly evolving designs of Web sites, and the often vague and conflicting guidelines that can be found in the literature (Ivory & Megraw 2005). By contrast, the user interface of current Web browsers and their integrated navigation support tools closely resemble those of browsers from the early days of the Web, even though several studies have provided pointers for improvement of these tools (Tauscher & Greenberg 1997) (Cockburn & McKenzie 2001). In addition, many support features that were common in the related field of hypermedia - different types of links, graphical overviews, trails, and annotation mechanisms, to mention a few - are not available on the Web (Bieber, Vitali, Ashman, Balasubramanian & Oinas-Kukkonen 1997) (Vitali & Bieber 1999) (Di Iorio & Vitali 2004).

Usability on the Web has been topic of a manifold of research activities. Theoretical models of mainly goal-oriented Web navigation have been developed; the effectiveness of various site designs and navigation support tools have been evaluated in laboratory studies, and user activities on the Web have been analyzed in field studies. These activities have provided us with numerous valuable, yet scattered, insights on the way the Web is used, and usability issues as experienced by the users. Particular attention has been given to the discovery and exploitation of navigation within Web sites, an activity that is commonly labeled *Web usage mining*. Client-side long-term studies on general Web usage - which typically involves more than one Web site - have been surprisingly scarce and the last reported study (Cockburn & McKenzie 2001) is already more than five years old.

Given the wide variety of users and their activities, and the lack of design guidelines, it might be hard to develop Web sites and browser navigation support that fit the needs of all users. The research field of *adaptive hypermedia* aims to find answers to this issue by developing user-adaptive support concepts; personalized interfaces that target the needs and activities of each individual user are expected to provide a more usable alternative than traditional one-size-fits-all interfaces. The current challenge is to apply this concept - which has proven to be successful in closed systems - to the open field of the Web (Brusilovsky 2004).

In order to achieve the goals of better design concepts and Web sites that automatically adapt to the user, several issues need to be addressed. These issues can be summarized by an adaptation of the challenges for adaptive Web sites, as formulated by Perkowitz & Etzioni (1997):

1. What kinds of generalizations can we draw from patterns of user interaction with the Web?

2. What individual differences in user access patterns can we find in these interaction patterns?

3. What user actions and user goals can we predict from these patterns?

4. What kinds of changes could we make by design improvements and personalization techniques?

5. Can we evaluate the impact of these techniques from changes in user interaction with the Web?

The research reported in this thesis is focused on the above questions. The approach followed to answer these questions is intentionally broad and exploratory. A wide spectrum of research from the fields of adaptive hypermedia research, as well as empirical studies and theoretical notions on user interaction with the Web, served as a reference framework during the process. Two main activities emerged to be important focal points for navigation support concepts: *information gathering* and *recurrent activities*. Two laboratory studies and a long-term study were carried out to find answers to the challenges listed above. In the analysis, the extraction of general patterns, as well as individual differences in activities - *navigation styles* -, were central topics. A further consideration is the concept of *personalization*; exciting work on understanding the users and their activities is carried out by the adaptive hypermedia community, but some 'personalizations' might better be considered design improvements.

In the following sections we introduce several topics that are relevant for the remainder of the thesis. Readers who are familiar in the field might want to skip these parts and move directly to the thesis overview in section 1.4, or to cursorily read them through as a short reminder. In the next section we sketch a brief history of the World Wide Web, which emerged as a marriage between Internet technology and hypermedia. In section two we give a bird's eye view on Web technologies, such as its communication protocols and authoring mechanisms; here we also define the terms Web page and Web site. Section three deals with usability matters on the Web, in particular matters related to finding and relocating information, and concepts to deal with these matters, mainly drawn from insights from the field of hypermedia. The introduction concludes with an overview of the next chapters.

1.1 A very brief history of the Web

The World Wide Web is a little more than a decade old. Yet, its history can be traced back to many decades before it eventually emerged as a combination of Internet technology and hypermedia. In order to understand the Web as it is today, the way it is used, and the way it is appreciated by researchers from the various fields, a brief historical perspective is in place.

1.1.1 The Internet

The World Wide Web is an online information system built on top of the infrastructure of networked computers provided by the Internet. The roots of the Internet lie in the ARPANET, an American government-initiated research project that started late 1966. The goal of the project was to design a network and communication protocols for the exchange of files and data between computers; bundling the computational power of computers all over the country would provide strategic knowledge that was invaluable during the Cold War. By the late 1970s the ARPANET covered about fifty American universities and government research centers. In this period the ARPANET was merged with several other networks of European and American networks, and a standard transfer protocol, TCP/IP, was agreed upon. 'The Internet' came to mean a global and large network using this protocol (Lee, Montgomery, Shapiro, Shah & Want 2000) (Berners-Lee 1999).

The first protocols and associated applications on the ARPANET were *telnet*, which allowed someone to login to a computer at another site, and the *file transfer protocol*, which allowed files to be transferred from one computer to another. From the early beginning, *email* was an invaluable means to send electronic messages from one person to another. To facilitate the publishing of text files that could be read by many others, discussion groups and bulletin boards were developed in several networks. Various applications, in particular WAIS and Archie, were developed to facilitate the search for resources. In the late 1980s all major American universities embraced the Internet and Campus Wide Information Systems (CWIS) became widespread. In the early 1990s a team from the University of Minnesota developed Gopher, a client-server application for the sharing of information that became the 'hottest new thing on the Internet' (Frana 2004). Due to its ease of installation and use, it gained a large user population, who did not necessarily have to understand complicated Internet protocols. Gopher provided a hierarchical, browsable structure for organizing and accessing documents and text, using a list of numbered choices. Information available on Gopher included library catalogs, government information, news articles and many other varieties of information.

1.1.2 Hypermedia

The other essential ingredient of the World Wide Web is the concept of *hypermedia*. The history of hypermedia dates back to Bush (1945), who envisioned a machine, the *memex*, which would allow users to relate information sources and their own insights in an associative manner, as an alternative for the librarian approach of indexing by alphabet, author or publication date. By consulting several sources consecutively, a user builds a *trail* of related documents, which can be labeled and annotated with notes and comments. Over time a massive selection of pertinent trails of interests will be built that mimic the user's way of thinking.

In the 1960s Bush' vision was translated into a concept of non-sequential writing. Douglas Engelbart developed his oN-Line System (NLS), which consisted of a hierarchically structured set of statements, each labeled with a unique identifier; each statement could cross-reference any other statement. Macro commands could be given to automatically find the definition of a special term in a glossary. In the same period Ted Nelson coined the term *hypertext* for these kinds of systems and started the Xanadu project (Nelson 2000). In Xanadu, authors could directly provide a *link* to an information source instead of citing it. A particular goal of the Xanadu project was to provide a universe of parallel versions of documents - revisions, summaries, translations - that could be compared next to one another. A vital feature of the Xanadu system - which was not implemented until 1988 - would be the visualization of all links between parallel or related documents, in order to understand the connections.

In the 1980's numerous hypertext systems were created and presented to the public. This growth in interest can be explained by the advent of personal computers and graphical user interfaces. The term hypertext is being replaced by the term *hypermedia*, to reflect the fact that information sources not necessarily have to be limited to text, but also could include graphics and other media. Despite the similar underlying principle of non-linear authoring, the systems differed wildly in size, focus on authoring or reading, application domain, and interface (Halasz 1988). As an example, the *NoteCards* hypermedia system (Halasz 1988) was designed to help people work with ideas. The system provides the user with a network of fixed-sized electronic notecards interconnected by typed links. Graphical overviews were given by specialized *browser* cards and *filebox* cards. The *Guide* system (Brown 1987) takes a radically different approach. In order to prevent the naive user from seeing the underlying datastructure of interlinked substructures, a hyperdocument is presented as a single - linear - scrollable document. Nonlinearity was introduced by a number of mechanisms: the *replacement* button, which causes initially hidden material to be expanded at that particular point; *pop-up* windows that contained additional notes; *reference* buttons are used to jump to a different point in the document.

Just before the introduction of the Web, the hypermedia community faced several issues. First, there was the issue of maintaining *link consistency* and *versioning* (Lee et al. 2000) (Davis, Hall, Heath & Hill 1992); what to do if a page is modified, moved or deleted; should earlier versions remain stored, should links point to the original version or the new version, and how can dangling links that point to non-existing documents be traced and treated? Further, if hypermedia techniques were to be generally used by a large user population, *rhetorics* should be established and guidelines how to write readable hypertexts (Halasz 1991). To conclude, for very large hyperspaces query-based search would be essential (Halasz 1988). But above all, there was the need for *open hypermedia* *systems* that would integrate and provide substantial large and interesting bodies of material; a *killer app* was needed to bring hypermedia to the public (Halasz 1991).

1.1.3 The World Wide Web

It was Tim Berners-Lee, an employee at the CERN research institute, who combined the networking capacities of the Internet and the concept of hypermedia in the World Wide Web. When Berners-Lee presented his WorldWideWeb browser at the Hypertext'91 conference, his demo sprung little interest. His goal was to gain attention from the research community, and to find a party willing to extend their hypermedia browser in order to exchange information via the Internet. Whereas Berners-Lee suggested in his memoirs (Berners-Lee 1999) that the hypermedia community did not recognize the great potential, the larger part of the community did so. However, the preliminary implementation of the browser did not make a convincing demo. More importantly, his system did not address any of the issues facing hypermedia systems of the time; in fact, it lacked many essential hypermedia features such as typed pages and typed links (Lee et al. 2000). One could say that the Web was hypertext stripped to its bone (Nielsen 2005*a*).

The Web has become the most predominant hypermedia system for a variety of reasons: Berners-Lee's continuing efforts, the development of the Mosaic Web browser, and the fall of the competing network information system, Gopher. In addition, the simplicity of the communication protocol and the early HyperText Markup Language (HTML) made it easy to implement and to extend (Lee et al. 2000). Ironically, Berners-Lee never wanted actual users to write markup code; in his opinion, editing facilities were essential for collaboration via the Web (Berners-Lee 1999). However, with the exception of some early Web browsers, no browser provided combined browsing and editing facilities. As a result, the Web has mainly become a hypermedia system for information delivery rather than a system for collaboration.

Since the popularization of the World Wide Web research on hypermedia systems has moved from closed systems to the Internet. The distributed architecture of the Web posed a number of limitations. In particular, the one-way links - which were only defined at the source page - made it impossible to check which links lead to a particular page. This leads to the risk of dangling links once documents are moved or deleted. Further, the number of different Web servers and the flexible design led to non-uniformity in interfaces. Native search facilities, global orientation and navigation aids, and multi-language support were not available. In order to overcome many of these issues, the *Hyper-G* system was developed as a client-server application on top of the World Wide Web (Andrews, Kappe & Maurer 1995). Its structuring features, external linkbase, and attribute and content search did provide functionality not available on the Web. Whereas

Hyper-G still exists as a commercial product, HyperWave, it never became the second generation network information system it intended to be.

The World Wide Web as we know it today is radically different from the Web in the early days. In the 1990s the Web was dominated by male academics, who accessed the Web from university and corporate sites. Web sites were mainly static and manually crafted; the focus lay on information and content delivery. With the growing number of home connections the user population became wildly diverse. The late 1990s saw the rise of search engines like Google and Yahoo, which provided the much-needed query-based access to web content. New technologies emerged, including content management systems that allow the reuse of layout information and the dynamic creation of documents; information can be updated as frequently as required for news pages and online shops. Functionality that used to reside on desktops is made accessible through Web interfaces, covering an application range from email, chat, and bulletin boards to complex applications like travel agencies, libraries and shops. These developments changed the Web into a dynamic, interactive system that is used on a daily basis for a broad scope of activities: news and entertainment, online commerce, travel planning, research, communication, and collaboration, to name a few (Hawkey & Kellar 2004) (Cockburn & McKenzie 2001) (Dix et al. 2004) (Network Working Group 2005).

1.2 A bird's eye view on Web technologies

For most readers of this thesis, the content presented in this section will not be new. For those who are not familiar with the technologies behind the Web - networking protocols, addressing mechanisms, and markup languages - the most important concepts are introduced in this section. Please note that due to its brevity, this overview is far from complete, and many details have been left unmentioned or have been generalized.

Network communication in the World Wide Web is based on the Hypertext Transfer Protocol (HTTP), built on the TCP/IP protocol. Upon receiving a HTTP request, a Web server executes the associated action, in most cases this means that it sends a Web document or a file to the browser client that sent the request. HTTP is a *stateless* protocol: unless it is specifically contained in the request headers, no context information on the request is available. This implies that it is hard to determine whether a requested resource is a Web page on itself or an embedded file that is used within the page (Baldi, Frasconi & Smyth 2003).

Web pages are identified by a Uniform Resource Identifier (URI), also known as a Uniform Resource Locator (URL) or simply a Web address. Its general structure is as follows:

http://host:port/path?query#fragment

The *host* part is either a numerical Internet (IP) address or a registered, logical name like 'www.foo.com' that is resolved to an IP address by a domain name server; it identifies the Web server to which a connection should be made; optionally a *port* can be specified through which the connection should take place. The *path* specifies the location of the file on the web server; this may be a directory path and a filename, but many dynamic web servers use alternative database or script commands that are not human legible. The *query* part is used for non-hierarchical allocation of the resource, and to send additional parameters to the Web server. The *fragment* specifies the view on the Web page requested, typically a marker to which the Web browser should automatically scroll (Baldi et al. 2003) (Network Working Group 2005).

Web pages are written in the Hypertext Markup Language (HTML). Within running text, markup *tags* are used for structuring and laying out text and images, inserting links and scripts, and other non-textual items. The actual contents of a Web page are preceded by a *header*, which provides meta-information, such as the page title, author, language, and the date of creation (Baldi et al. 2003). Current web sites often make use of external *cascading style sheets*, which specify the manner in which all tags should be rendered; this improves page layout consistency (Ivory & Megraw 2005). Interactive screen elements can be integrated using JavaScript and applets, code that is executed on the client side.

A note should be made on the terms *Web page* and *Web site*. It may well be that what a user sees as one page is built of more than one file; frames and inclusions allow authors to reuse certain elements, such as menus and advertisements. In this thesis we adopt the following, somehow simplified but user-centered, definition of a Web page: 'everything that is shown in a browser window as a result of a page request initiated by the user'. Further, it is not immediately clear in what manner a Web site should be defined. In some cases a Web site is defined as all pages that reside on one Web server. However, these pages may be owned by several individuals or institutions; analogously, a large company may have multiple Web servers. As the registered, logical host name is the common way in which users distinguish Web sites from one another, we define a Web site as 'all pages that share the host part of the URI'. We use this definition loosely, as in many cases it is desirable to group several prefixes, as they refer to separate parts of one institution's Web site - for example www.foo.com, members.foo.com, www.cs.foo.com. Analogously, in several cases different host names are registered that lead to the same site or to mirrors, exact duplicates of the site.

1.3 Web usability and personalization

The non-linear structure of hypermedia systems allows and encourages authors to divide pieces of information in units that can be related in multiple manners. In contrast to traditional flat text, there is not one story line that needs to be followed; hypermedia does not force a strict decision about whether any given idea is either within the flow of a document or outside of it - like footnotes and references. This writing style provides the opportunity to cater different reader interests, as readers are free to pave their own way; they can read the document in the order they like, and can choose to skip or zoom into specific parts of the document. In addition, users may choose to follow an interesting side path and later return to the main topic. This freedom comes at a price, though, as users must invest effort to keep track of their locations. This will cause few problems in sites that a user is already familiar with, and in many cases users will find their ways in unfamiliar sites as well, making use of their previous browsing experiences and knowledge about the domain.

If users fail to understand the way a site is structured, they will most likely not succeed in finding the things they are interested in. Most people will end up getting lost when dropped in a big forest with poorly indicated trails, especially if they do not walk forests on a regular basis. Similarly, if people want to return to a place visited before, much efficiency can be gained from the trails followed earlier, in particular if these are marked by pebbles or breadcrumbs. Interesting places can be easier recognized and seen from a distance if they are marked by some sign. The structure provided by the forest's guided tours, available maps, and marks that have been left intentionally and implicitly (footprints, for example) enable visitors to easier orient themselves, which facilitates both finding of new, interesting places and returning to places of interest.

Web navigation support may be considered the electronic counterparts of the forest navigation support. A large amount of research on Web usability and personalization is dedicated to providing users with better tools for finding, organizing and relocating information. From the physical metaphor of navigation several points of focus can be distinguished: *landmarks* provide initial points of reference; gradually, users develop *route knowledge* - they learn paths that lead to the desired location; at some point, users can abstract from the routes and develop *survey knowledge*, a mental model of the environment that allows users to confidently use alternative routes and shortcuts (Dillon & Vaughan 1997). While a useful metaphor, one should keep in mind that physical navigation does not cover all that is relevant to finding and relocating information on the Web; as an example, keyword search does not have a physical equivalent. Further, in general users are not as much interested in finding a page location, but rather in the information and services associated with the page. With these limitations in mind, in this section we describe the 'lost in hyperspace' issue, and navigation support concepts that target this issue.

1.3.1 Lost in hyperspace

The *disorientation problem*, more colloquially called the 'lostness in hyperspace' issue was first mentioned by (Conklin 1987). The concept of lostness has been inspiration for a large amount of research on navigation support tools and personalization. The root of the problem is the additional *cognitive overhead* needed for users to (Thüring, Hanneman & Haake 1995):

- identify their current position in a hyperdocument;
- reconstruct the way that led to this position;
- distinguish among different options for moving on from this position.

When reading a hypermedia document, users continuously need to choose 'which links to follow' and 'which links to leave alone'. They have to create their own trail across the document. Moreover, the associative linking invites them to engage in side paths on related topics, to return to the main topic at a later point. When users fail to keep track of this task scheduling process, they might arrive at a particular page and forget what was to be done there, they might neglect to return from interesting side-tracks or they might fail to find some pages that contain relevant information (Otter & Johnson 2000).

Disorientation typically occurs in large, densely linked hyperdocuments, in particular if they are incoherent and do not sufficiently differentiate among pages and links (Conklin 1987). Users who are disoriented tend to return to a focal point in the document, a *landmark* - in the context of the web this is typically a site's home page (Pilgrim & Leung 1999) - or give up (Smith 1996).

Whereas disorientation also may occur due to unfamiliarity with the task domain, there is empirical evidence that disorientation is mainly caused by unfamiliarity with the structure of a hyperdocument. Although users with more hypermedia experience tend to better construct a mental model of a document structure, the unpredictable character of associative links seems to prevent even experienced users to make effective use of it (Otter & Johnson 2000). Individual differences cause some users to be more likely to experience disorientation. *Field-dependent* users strongly rely on the visually presented structure, while field-independent tend to adopt a more active analytical approach - they build and make use of their own frame of reference (Chen & Macredie 2002); consequently, field-dependent users are more vulnerable to experience disorientation in unfamiliar document structures. Spatial ability - the ability to understand and use spatial and multidimensional patterns - and associative memory also play a major role in navigation performance (Chen 2000). Hence, hypermedia documents should provide visual elements for orientation support.

In linear texts, context is provided by rhetoric elements such as introductions, summaries, overviews, and division of the text in units as chapters, sections and paragraphs. These elements add *coherence* to the text and ensure that readers can keep track of the argumentation. In a similar manner, rhetoric elements are needed to provide structure to the hyperdocument (Thüring et al. 1995).

1.3.2 Coherence in the Web

Pre-Web hypermedia systems mainly contained *associative links*. Structure was provided by local and global overview maps. However, in particular for larger Web sites these overview maps would quickly become too complicated to be useful. For this reason, it has become good practice to provide alternative means for adding coherence to a Web site; in particular by imposing one or more *hierarchies* on the document by *structural links* - grouped in menus and other kinds of navigation bars. The user's current location in the hierarchy is often indicated by *breadcrumbs* that show the shortest path from the site's home page to the user's current page. The context information provided by these structural links is important for effective navigation, as each navigation process is inextricably tied to the structure of the document. In addition to structural context, *temporal context* is provided by marking links to pages that the user has already visited. Browser history mechanisms - the back button, bookmarks and the history list - can be used to reconstruct the navigation path that led to the current position (Park & Kim 2000) (Nielsen 2005*a*).

A continuing discussion is whether broad, shallow hierarchies and associated menu structures should be preferred to narrow and deep menus. There is a trade-off between the smoother, less cluttered look of narrow menus - which provide fewer options at a time -, and limitation of the number of navigation steps needed to locate a specific item. There is growing evidence that broad, shallow menus are often better (Dix 2005); they demand less from the user's short-term memory, and often provide better cues which direction to go for finding the desired information (Katz & Byrne 2003).

In addition to the context provided by navigation support, the page content should be considered. If there is no specific order for reading assumed, an author does not know what content visitors have seen before. This implies that any piece of information should be self-contained to a large extent, or provide pointers to the pages that contain the information that the user is assumed to be familiar with (Dix 2005).

1.3.3 Missing hypermedia features on the Web

From the early inception of the World Wide Web until now, several authors mentioned the lack of many hypermedia features which were available in pre-Web closed hypermedia systems, or envisaged as needed by the hypermedia community. Several research efforts, in particular the Semantic Web movement (Van Ossenbruggen, Hardman & Rutledge 2002) have been initiated to overcome the issues. It is beyond the scope of this thesis to describe them all in their historical context. In this subsection a selection of frequently mentioned issues is discussed, heavily drawn from the following publications (Halasz 1988) (Halasz 1991) (Vitali & Bieber 1999) (Andrews et al. 1995) (Bieber et al. 1997) (Nelson 2000) (Van Ossenbruggen et al. 2002) (Nielsen 2005*d*).

Typed, or annotated, links and pages provide context to users, and are considered indispensable for providing coherence on the Web. Whereas HTML supports typed links, few browsers take them into account. Cascading Style Sheets (World Wide Web Consortium 2005*a*) would enable authors to specify different visualizations for different link types, but this option is hardly used. The current standard still is blue underlined links, with purple underlined links for pages visited in the recent past. The most obvious distinction would be links that point to pages within the same site and external links. A more semantic distinction would be to differentiate between relations between two pages: for example 'glossary', 'annotation', 'structural' and 'associative' links. More domain-specific differentiations would enable an insurance company to distinguish between policies, claims, accident descriptions, police reports and complaints (Bieber et al. 1997).

Links function as proxies for the information behind the link. However, current Web links do not provide any more cues than a small number of words or an image that the link is associated with. They do not convey the page title, its author, the date it was created, or the language it is written in. Even though the browser marks links that lead to recently visited pages, it does not show how often the page has been visited, or when it was last visited. Also, some links lead to unexpected actions, such as the opening of a new window. Many of these aspects can be recognized without additional effort from author or reader, and conveying these aspects would arguably be as useful as manually typed links. Weinreich & Lamersdorf (2000) implemented a prototypic system that shows several of these aspects in a pop-up window. Unfortunately, the system has never been formally evaluated, which implies that at the moment we can only recognize the likely benefits of these additional annotations.

Hypermedia systems and the Web have translated Bush's original concept of user-defined *trails* into the concept of non-linear writing. Although several hypermedia systems allowed authors to superimpose linear paths on top of the hypermedia structure, author-defined *guided tours* have never become commonplace. An obvious mechanism would be to let users store their navigation paths as *trails* that can be retraced at a later point, or shared with others (Greer & Philip 1997). Trails are one type of *history mechanisms*. Current browsers provide three types of history mechanisms: the back button, a history list, and manually defined bookmarks. Each of these tools have been shown to have their own limitations (Cockburn & Jones 1996) (Kaasten & Greenberg 2001) and various alternative history mechanisms have been proposed (Tauscher & Greenberg 1997) (Mayer 2000).

A concept related to history mechanisms, which was very common in hypermedia systems, but not present on the Web, is the option to *annotate* pages with comments or links to other pages - reminders, comments, funny associations or summaries. Whereas some projects have attempted to enhance browsers with this feature (Karger, Katz, Lin & Quann 2003), it is a feature not available on the Web. Annotation is a form of *tailoring* the environment to individual needs; hypermedia systems that allow users to add notes, to change the visual appearance or the structure of a document, or to perform any other kind of modification, are called *adaptable*. Research on *adaptive* hypermedia systems aims to automatically perform these modifications, based on inferred profile information on the users and their tasks.

From the above discussion four dimensions emerge that provide context to the user: the document structure, the document content, the interaction history, the user (group) profile, and the user task. These contexts may be provided by various interface elements: link annotation, menus and lists, graphical overviews, and additional notes and comments. There are several reasons why these concepts have not been integrated in the Web interface, either as part of a Web site or common browser functionality. First, whereas some pieces of context are readily available or easily retrievable, other pieces of context would have to be manually provided, or inferred from other sources. Research on these topics is mainly carried out in the fields of *information retrieval* and *user modeling*. Further, given the share amount of additional context provided, it is not clear *which* elements have most impact in what user contexts; clearly, not everything can be presented at once, as this will result in a highly cluttered interface. Also, there are several methods to translate the context information into user interface elements, and it is not apparent which interface design would work best. Research in the general field of human-computer interaction, and in particular in the field of *adaptive hypermedia*, is carried out to provide knowledge on this matter.

1.4 Research approach and thesis overview

As stated in the beginning of this chapter, this thesis is focused on the analysis of user navigation on the Web. We aim to find general patterns of user interaction with the Web, and individual differences in these patterns, and to interpret these patterns in terms of user tasks, navigation strategies, and usability matters. A further goal is to investigate what design improvements and personalization techniques can be used to support users in navigating the Web, and how these techniques can be evaluated. The research approach used is broad and exploratory, in which we drew heavily on existing bodies of knowledge, in order to integrate, update and extend current insights in Web navigation.

In the next chapter we give an overview of the field of adaptive hypermedia, a research area that is particularly involved in modeling and understanding the user and the user context, and in providing personalized support. The activities in the field, and the adaptation techniques used, give valuable insights on how navigation support can better address the users and their tasks. The concept of layered evaluation, and theory assessment measures, provide the framework used for the design of our user studies.

How users interact with the Web has been topic of a large variety of empirical studies and theoretical research. Most user studies were focused on one particular aspect of Web navigation - such as the usage of browser tools, how users gather information, web search, and recurrent behavior. Theoretical models of web navigation mainly focus on information gathering activities, with a strong bias toward browsing. In chapter 3 we summarize, discuss and relate earlier work, in order to come to a more integrated view on Web navigation. The results provide several points of reference, but also reveal that many aspects of Web navigation are still not known, mainly because of lack of empirical data. We conclude the chapter with an outlook how the laboratory studies and the long-term study that we carried out, aim to fill these gaps.

In chapter 4 the techniques for data collection, preparation, and analysis used in our studies are discussed. First matter of concern is the collection of data. We used and extended the proxy-based Scone framework (Weinreich, Buchman & Lamersdorf 2003) for this purpose. We describe several preprocessing steps that were needed to remove artifacts from the data - in particular embedded advertisements have become a nuisance for client-side or proxy-based Web usage mining; further, several annotations to support the mining process are described. We continue with an overview of the graph-based framework and associated measures that we used to identify navigation styles. The chapter concludes with a description of the Navigation Visualizer, a tool that we developed to support the interpretation of the statistical results.

In chapter 5 we present the results of two laboratory studies on individual differences and the impact of adaptive navigation support in information gathering activities. The first laboratory study was carried out to provide answers to challenges 2 and 3. Conflicting results exist in the literature on how users behave when they fail to understand and make use of the structure of a site; some models and empirical findings suggest that extensive backtracking indicates these kinds of problems (Pirolli & Card 1999) (Smith 1996), where other sources associate shallow search and backtracking activities with users making effective use of the site structure (McEneaney 2001) (Teevan et al. 2004). The aim of the study was two-fold: we wanted to find measures to *automatically* identify these situations, and we wanted to *explain* the differences found in terms of individual differences. Based on the results of the first laboratory study, we conducted a follow-up study to investigate the impact of one particular personalization technique, task-sensitive link suggestions. Again, the aim of the study was two-fold. First, we wanted to verify whether our approach of identifying navigation styles from a large number of measures would provide effective *theory assessment measures* for evaluating design improvements and personalization techniques (challenge 5). Further, we wanted to investigate whether users would benefit from additional personalized context information. The study was purposely kept limited to the evaluation of one particular technique, to ensure that we did not merely test one system against another.

In chapter 6 we present the results of a long-term study on Web navigation, based on a pool of 25 participants whose Web activity was logged for an average period of three months. The last reported long-term study is based on a data set collected in late 1999, and was mainly focused on how often users revisit pages and the use of bookmarks. Statistics on the use of browser interface elements date back to 1997. Moreover, the results were mainly overall statistics, generalized over the whole participant pool, and individual navigation sessions were not taken into account. Further, search activities - which are an important and integrated part of Web navigation - have only been researched in separate studies. The long-term study provides us with detailed data on current user interaction with the Web. The analysis showed significant differences in current Web usage compared to the earlier studies. We identified various general patterns in user interaction with the Web (challenge 1), and found a wide range of individual differences between users and user navigation sessions (challenge 2). From the results several implications and directions for design improvement and user-context aware support that directly addresses current issues on the Web were distilled (challenge 4).

In the concluding chapter 7 we return to the challenges that we aim to answer in this thesis. We discuss several insights gained on models of Web navigation, user navigation styles, and client-side Web usage analysis. We present several design implications, and directions for enhanced, user and context aware navigation support. The thesis will end with the usual concluding remarks.

Chapter 2

Adaptive hypermedia and personalization on the Web

2.1 Introduction

Web sites, and the World Wide Web in general, can be regarded as a sub domain of hypermedia documents. Similarly, *adaptive hypermedia* can be seen as a generalization of personalized Web systems (Anderson 2002). Adaptive hypermedia is a research field that emerged in the early 1990s from its two main parent areas, *hypertext* and *user modeling*, which aims to provide alternatives for the traditional 'one-size-fits-all' approach of traditional hypermedia systems (Brusilovsky 2001). Adaptive hypermedia systems were expected to be useful 'in any application area where the system is expected to be used by people with different goals and knowledge and where the hyperspace is reasonably big' (Brusilovsky 1996).

Various working definitions of adaptive hypermedia and Web personalization can be found in the literature - e.g. (Brusilovsky 1996) (Brusilovsky 2001) (Eirinaki & Vazargiannis 2003) (Perkowitz & Etzioni 1997). In the context of this thesis we adopt the most general working definition, as used in one of the first overviews on the field, published by Brusilovsky (1996) - italics are added for clarity:

> By adaptive hypermedia we mean all hypertext and hypermedia systems which reflect some *features of the user* in a *user model* and apply this model to *adapt various visible aspects of the system* to the user.

A multitude of adaptation techniques have been proposed, implemented or evaluated in the past. Traditionally, these techniques are roughly distinguished as *adaptive presentation* and *adaptive navigation support*. Adaptive presentation techniques adapt the content of pages; adaptive navigation support techniques adapt a hyperdocument's link structure.

Whereas the World Wide Web is by far the largest existing hypermedia system, only a limited scope of user-adaptive features can be found on the Web. This is partially due to the fact that most adaptive hypermedia techniques, in particular adaptive navigation support, require a well-structured set of documents enhanced by metadata annotations (Brusilovsky 2004). This is feasible in a relatively small closed corpus, but not in the large open corpus of the Web. In addition, many of the classic hypermedia's features for structuring, navigation and annotation are not supported in the World Wide Web (Vitali & Bieber 1999).

This chapter provides an overview to the related fields of adaptive hypermedia and user modeling. Emphasis is given on the similarities and differences in developments and approaches between the general field of adaptive hypermedia and its subdomain of Web personalization. The upcoming section is an introduction to adaptive hypermedia. We discuss the goals of incorporating adaptivity into the system and compare the benefits of incorporating adaptivity into the system with overall design improvement. Further, we describe its various application areas, the features to what a hypermedia system can be adapted, and a number of adaptation techniques. In section three we concentrate on the various aspects related to user modeling: the acquisition of user and usage data, and the inference of knowledge from this data. Two approaches can be distinguished: hand-crafted explicit representations of the users and their contexts, and implicit user models that are built using statistical inferences and machine learning techniques. In the final section of this chapter we provide a high level model of the adaptation process, and discuss the concept of *layered* evaluation.

2.2 Adaptive hypermedia

Hyperdocuments allow users to freely pave their paths through a document structure. However, this freedom comes with a cost, as users need to keep track of their current position, the way that led to this position and the different options for moving on from his position (Thüring et al. 1995). In the early 1990s researchers recognized the usability problems that might be caused by these extra efforts, especially in larger hypermedia structures intended for larger audiences (Brusilovsky 1996). In response, various navigation support technologies were proposed, all sharing the same core idea: adapt the hypermedia system to the goals, knowledge and preferences of the individual user (Brusilovsky 2004). Initially, the most popular application area for adaptive hypermedia research was educational hypermedia, as it provides naturally restricted domains and can work with relatively simple user models that reflect a learner's knowledge on the domain. With the gaining popularity of the World Wide Web, most adaptive systems became Web-based and the focus shifted toward information retrieval in unrestricted document spaces.

In this section an overview of adaptive hypermedia techniques and applications is given. We start with a discussion of the goals to be served by incorporating adaptivity into the system. In the second part of this section the concept of personalization is compared with overall design improvement. We continue with a bird's eye view on the various application areas of adaptive hypermedia. The second-last part deals with the features of the users and the user context that a system can adapt to. The section concludes with a categorization of personalization techniques.

2.2.1 Goals of adaptive hypermedia

According to Kay (1993), the general goals of adaptivity are to improve the efficiency and effectiveness of the interaction. This means that adaptive systems aim to make complex systems more usable, present the users with what they want to see, as well as speed up and simplify the interaction.

Jameson (2003) separated several adaptation goals in the context of information acquisition in a large repository. One type of adaptivity is *helping users to find information* by support for browsing, and support for query-based search or filtering. In addition to user-responsive link annotation or annotation of search results, Jameson (2003) specifically mentioned systems that spontaneous provide information by anticipating user needs based on user actions. As an example, the browsing assistant Letizia (Lieberman 1995) builds an interest model from the documents that a user is reading and uses the 'idle' time in which a user is reading to perform a breadth-first search for documents in order to save the user some exploration.

A second adaptivity type is *tailoring information presentation* to the user, the user context or the device. As an example, the Power Browser (Buyukkokten, Garcia-Molina, Paepcke & Winograd 2000) is a proxy-based system that translates regular Web pages to a format more suitable for handheld devices and low bandwidth by collapsing text, reducing or removing images and link ordering.

The third adaptivity type mentioned by Jameson (2003) is perhaps the most practical: *recommending products*. This type of adaptivity can be found in many commercial Web sites. Product recommendations may be beneficial for both the vendor - whose sales might increase - and for the user - who might find new relevant products. In most e-commerce sites, the product recommendations are presented *in addition to* the existing navigation infrastructure, which is an unobtrusive and quite accepted manner. Other recommender systems, such as MovieLens (McNee, Lam, Konstan & Riedl 2003), have recommendations as their main navigation infrastructure.

Other adaptivity goals mentioned by (Jameson 2003) are *supporting collaboration* and *supporting learning*. As an example, the ActiveMath system (Ullrich 2005) dynamically generates course structures from various student characteristics using hierarchical network planning.

An adaptation goal that is not specific to supporting users in finding new information, is *taking over parts of routine tasks*. The aim is to speed up the interaction and to save the user from repeatedly performing tasks that may place heavy demands on a users' time, but typically not on their intelligence, creativity or knowledge. This type of assistance might vary from automatically filling out fields in a form to negotiating in Web auctions in behalf of the user.

Various other adaptation goals may be thought of, many of them serving commercial goals rather than enhancing usability. E-commerce site might want to provide discounts, promotions and more flexible paying options to loyal customers. Various simple adaptations with relatively strong impact on customer loyalty include user salutation and memorization of the user's last actions.

2.2.2 Personalization versus overall design improvement

The ambition of adaptivity is that not only 'everyone should be computer literate', but also that 'computers should be user literate' (Browne, Totterdell & Norman 1990). This rather futuristic statement might as well be stated in a more mundane manner: personalization is a designer's approach to achieve harmony between users, tasks, environments and the system (Benyon 1993). Adaptation of hypermedia systems, such as the Web, can be applied to the needs of each individual user, of certain user groups, or of the whole user population. In the latter case it might be more appropriate to call the adaptation a 'transformation' (Perkowitz & Etzioni 1997) or 'design improvement' rather than a 'personalization'. The distinction between personalization and transformation is not always easy to make. As an example, would one consider annotating Web hyperlinks with information on the user's page visit history as a Web personalization or as a design improvement? In his two-weekly Alertbox on Web usability, Nielsen (1998) claims that personalization is an overrated concept. Rather than investing time and energy on trying to predict individual users' needs it would be better to enhance the overall site design.

Nielsen's claim clearly challenges the view of the adaptive hypermedia community that 'one size does not fit all' (Brusilovsky 2001). However, the claimed successes of adaptive hypermedia systems might also be due to the relative infancy of hypermedia systems, and the Web in particular. According to Shneiderman (1997) the Web is still 'in the Model T stage of development'. Strategies for effective use of various multimedia resources and overall site navigation designs are still rapidly evolving and changing. As a consequence, and due to the rapid growth of the Web and its use, the way people use the Web is subject to change as well (Cockburn & McKenzie 2001). In Nielsen's view, 'Web personalization is mainly used as a poor excuse for not designing a navigable website'. As almost always, the truth might be somewhere in the middle. Addressing individual user needs and peculiarities separately might be more effective than addressing the widely diverging peculiarities of a user population as a whole, especially when knowledge on effective design approaches is limited (Shneiderman 1997). Incorporating user-adaptivity into the system is a design approach that should seriously be considered when dealing with a high degree of variety, as is the case with 'reasonably large hyperspaces that are used by people with different goals and knowledge'(Brusilovsky 1996).

The proof of the pudding is in the eating. Bringing adaptivity into the system might not only solve problems, but might also introduce a number of other problems. As Nielsen argued, it costs time and effort to incorporate adaptivity into the system design. Further, there is the problem of *hunting* (Benyon 1993): a system adaptation might alter a user's behavior, as a result of which the system adapts to this altered user behavior, thus forming an endless loop. An inherent problem of adaptivity is the uncertainty that is associated with drawing inferences from user behavior and making adaptation decisions (Herder 2003). As in regular usability engineering, empirical studies should help in estimating the benefits and drawbacks of incorporating adaptivity into the system. However, publications on user modeling systems and adaptive hypermedia rarely contain empirical studies (Weibelzahl 2005). At the end of this chapter the issue of evaluation will be dealt with in more detail.

2.2.3 Where can it be used

In principle, adaptive hypermedia techniques can be applied to any non-linear information domain and the range of applications is quite broad. In his 2001 review of adaptive hypermedia systems, Brusilovsky (2001) identified the following main application areas:

- educational hypermedia;
- online information systems;
- hypermedia for information retrieval.

The Web has become the main development platform for adaptive hypermedia systems. As a result, the interest in application areas that typically make use of stand-alone systems - online help systems and institutional information systems - has faded.

Traditionally, educational hypermedia has been the most popular application area for the adaptive hypermedia and user modeling community, as the domain of online courses is naturally restricted, the structure of the domain is relatively clear and the user needs are well-identified: acquiring knowledge on a certain topic. Furthermore, the user characteristics taken into account - typically the student's current knowledge level, level of interest and the student's learning style - are generally well-researched (De Bra, Aerts, Smits & Stash 2002) (Chen & Macredie 2002).

With the growth in usage of the Web, *online information systems* and *hypermedia for information retrieval* have become the most intensively used hypermedia systems. Examples of information systems mentioned by Brusilovsky include electronic encyclopedias, information kiosks, virtual museums, handheld guides and e-commerce systems. In the context of the Web, information systems might better be simply be referred to as Web sites. Adaptive navigation support within these sites is typically browsing-oriented and provides support for local navigation through link following. As popular Web sites are visited by a large amount of users, a relatively new form of within-site navigation support has emerged: automated recommendations of products or other items has become a popular technique, mainly in e-commerce sites (Baldi et al. 2003).

Since the very beginning of the Web, search engines, meta-search engines and directories have been available for locating relevant information in the ever growing collection of Web sites. Although these tools for *Web information retrieval* are widely used, there are several issues that hinder their performance: low accuracy, incompleteness and low coverage, low timeliness and bad ranking of results (Brusilovsky & Tasso 2004). Although users are reported to have a firm mental model of how search engines work , many of the issues mentioned are likely to be caused by the fact that queries are typically very short and therefore ambiguous (Nielsen 2005*a*) (Lau & Horvitz 1999). Various research projects have focused on filtering, sorting, categorizing or annotating search results based on a user model that typically represents a user's long-term interests.

An application area not specifically mentioned by Brusilovsky is general, application-independent personalization, of the user interface. The Web in particular has been shown to be used in a recurrent manner (Tauscher & Greenberg 1997). As described in chapter 1, early hypermedia systems provided local and global overviews of the hypermedia structure, user trails through the structure and various means for marking places already visited (Bieber et al. 1997). Various research efforts have been taken to introduce these features in Web browser interfaces as well. Whereas these representations of a user's past interaction arguably might be called a design improvement rather than a personalization, they do fit Brusilovsky's definition of 'adaptation of noticeable aspects of the system based on some features of the user'. Moreover, as will be discussed later on in this chapter, many adaptive systems do make use of a model of a user's past interaction with the system. A plausible reason for not specifically including navigation support based on a user's interaction history is the fact that most research on his topic is carried out in the more general field of human computer interaction.

In the research community, adaptation to the mobile context has become an important field of research. Mobile device interfaces are known to pose additional challenges to the user compared to their desktop counterparts: due to limited screen real estate, less context can be provided, context that is needed for maintaining a sense of location in the information environment (Thüring et al. 1995). In addition, as mobile devices need to 'fit into one's pocket', widgets are usually limited both in number and in size. Moreover, as mobile devices are designed to be used while on the move, users typically are more constrained in time and need to divide attention between the mobile device and their environment.

Several lines of research are specifically aimed at reducing navigation effort, or at constructing portal sites with instant access to information and services most likely to be relevant to the user (Buyukkokten et al. 2000) (Smyth & Cotter 2002) (Anderson & Horvitz 2002). Obviously, mobile devices can also serve as a means for information or amusement that relate to the user's current location. Several projects are devoted to offering users personalized guided museum tours, city tours and location-based information. Albeit exciting, research in these directions still appears to be carried out in relatively limited, academic environments.

To summarize, adaptive features can be used in any non-linear information domain. However, the Web has become the main arena in which adaptive hypermedia techniques are being researched. The fact that there is a large group of people who are familiar with and who use Web interfaces, provides two additional explanations. First, there is the relevancy: the Web is more than an academic playground and provides concrete issues to be solved. Second, the large pool of users provides excellent means for implicit user modeling and collaborative filtering; adaptive hypermedia techniques can be tested on experienced Web users. The unclosed hyperspace of the Web does pose the researcher with several challenges as well: often little to no semantic information is available and - except in laboratory studies - it is hard to infer what tasks users are working on. In chapter four we deal with the issues of Web mining and Web usage mining in more detail.

In the context of the Web, three main application areas of adaptive hypermedia techniques can be distinguished: personalization of Web sites, personalization of tools for Web information retrieval and personalization of browser interfaces. However, the three domains do overlap: for example, many Web browsers offer toolbars that provide instant access to search engines. Where and how the personalization actually takes place, is a mere technological issue and might be subject to change. More important are the questions what can be adapted, based on what and for what purpose.

2.2.4 To what can it be adapted

As is the case with general user interface design, any feature that is likely to influence the interaction with the system might be used as a source of adaptation. Likewise, these features can be identified using methods like GOMS, cognitive models and task analysis (Dix et al. 2004). Every system encapsulates an implicit model of the intended users and the intended use (Browne et al. 1990). However, in adaptive system design the focus is on features that are likely to vary between users or over time. During the interaction an adaptive system keeps track of these features in what is usually conveniently called a *user model*.

Irrespective of what the user model looks like or what information the model represents, a general observation is that the user model is used to reason about the most suitable individualized interface (Browne et al. 1990). From this perspective we could define a user model as an intermediary representation that is used for storing and translating concrete observations into interface adaptation decisions. In some cases a one-to-one relation between concrete observations and adaptations can be identified, such as a system that corrects commonly misspelled words. In many cases it is impossible for designers to relate observations to adaptations without the use of intermediary concepts (Browne et al. 1990). Moreover, these intermediary concepts are often used as a starting point for adaptive hypermedia systems, as adaptivity is aimed to be the answer to diversity in users needs. For instance, the observation that users might experience lostness in large hypermedia documents (Thüring et al. 1995) has been the starting point for research on how to recognize problems related to lostness from interaction behavior and on how to solve these problems (Smith 1996) (McEneaney 2001) (Herder & Juvina 2004) (Park & Kim 2000).

Traditionally, adaptive hypermedia systems base their adaptation decisions on various characteristics of the users (Brusilovsky 2001). Currently, additional input sources for data are used as well. In (Kobsa, Koenemann & Pohl 2001) these sources are categorized as *user data*, *usage data* and *environment data*. User data denotes information on the personal characteristics of a user; usage data denotes a user's past interaction with the system; environment data denotes all relevant aspects that are neither part of the user nor of the system, like the hardware and software platform used, available bandwidth, the user's current location and environment conditions - such as background noise, light sources and possible distractions.

Kobsa et al. (2001) identifies several categories of user data that have been used in adaptive systems. Perhaps the most elementary category is the *demographic data*, which includes information on a user's location, age, gender, education and profession. In combination with high-quality statistical data these simple demographics can be used for a rough initial fine-tuning of the interface. In particular the extraction of default values for dialog responses can save users some work or provide an example of the expected answer (Nielsen 2005*c*). *User background knowledge* is typically used for determining which concepts a user is already familiar with and which concepts might need additional explanation. User knowledge is an import characteristic for educational hypermedia systems. In the AHA! system (De Bra, Aerts, Smits & Stash 2002) the user knowledge level is directly employed for determining which parts of a course are suitable, too basic or too advanced for a student. *User skills and capabilities* represent the user's familiarity with the system and practical knowledge on how to interact with the system. For instance, first-time users may have different needs than experienced users. The category may also represent a user's physical challenges, such as limited vision or motor skills. *User interests and preferences* are used for determining the information, services or products that users are most likely to appreciate. In particular with the advent of Web information retrieval systems and e-commerce applications this category has received a great deal of attention from the adaptive hypermedia and Web personalization communities (Brusilovsky 2001), both due to real-world needs or economic benefits, and the availability of a large amount of data that can be used as an input for information extraction techniques and collaborative filtering (Baldi et al. 2003). *User goals* and *user plans* were considered important mainly in non-Web based hypermedia systems that attempted to satisfy user needs as effectively and efficiently as possible, such as help systems and educational hypermedia. As the World Wide Web is used for exploration and entertainment to a large extent, this category of user characteristics is less represented in Web personalization.

As a final category of user characteristics - not mentioned by Kobsa et al. (2001) - a *user's individual traits* is a group name for user features that 'together define a user as an individual' (Brusilovsky 2001). Individual traits involve relatively stable features, such as personality factors, cognitive factors and learning styles. Whereas of great interest to psychologists and system designers who want to understand, model and exploit user interaction with a system (e.g. (Chen & Macredie 2002) (Juvina & Van Oostendorp 2004)), it has been of little practical value in actual hypermedia systems (Browne et al. 1990) (Brusilovsky 2001). A plausible explanation is that it may take lengthy questionnaires or tests to reveal these factors. However, in controlled experiments this category of user characteristics has been shown to be valuable for explaining differences in measurable aspects of user behavior and usability issues, as experienced by the user (Chen & Macredie 2002) (Herder & Juvina 2004).

Usage data consists of the observable aspects of user interaction with the system. Initially, user errors were considered the most prominent candidate for automatic adaptations (Oppermann 1994). Brusilovsky referred to usage data as 'data about user interaction with the systems that cannot be resolved to user characteristics, but still can be used to make adaptation decisions' (Brusilovsky 2001). This definition implies that usage data should ideally be used as a source for inferring higher level user characteristics. To a certain extent, this is true mainly for early adaptive systems, which make use of hand-crafted student models or logic-based user models. However, statistical user models are becoming a main-stream approach to user modeling, especially in the context of the Web (Zukerman & Albrecht 2001). Perkowitz & Etzioni (1997) defined personalization as a challenge for artificial intelligence for finding mappings between usage observations to effective adaptations. Whereas these statistical approaches have as a drawback that user models might become uninterpretable for humans, its potential has been shown in the field of Web usage mining. In chapter four, which deals with Web usage mining, many potentially interesting observable interactions are described, most importantly click-through data that represents when, how often, how long and by what means accessed information items. Currently, the main application of usage data involves automatic discovery of user interests (Claypool, Le, Waseda & Brown 2001) and recurrent behavior (Tauscher & Greenberg 1997) (Cockburn & McKenzie 2001) that can be exploited for better access to information and services of interest to the user.

From the above discussion it can be observed that a virtually unlimited number of features can be used as an input for adaptations. During a recent workshop on personalized information access (Brusilovsky, Callaway & Nürnberger 2005) a trend was recognized to store and exploit as much of the user interaction with the system as feasible. The popularity of desktop search applications (Jones, Bruce & Dumais 2001) possibly inspired this trend and provides means to support this approach in real-world situations. However, there are risks involved in trying to make a user model as 'complete as possible', as it is known from the field of machine learning that irrelevant data may decrease performance or even lead to incorrect results (Mitchell 1997). Evaluation of the various steps taken by adaptive systems is gaining attention from the community as an indispensable step in the design process (Weibelzahl 2005) to overcome pitfalls and to recognize opportunities. At the end of this chapter this matter will be discussed in more detail.

2.2.5 What can be adapted

A large hypermedia structure of richly interlinked documents allows for a virtually endless variety of adaptations. Obviously, some adaptations will be useful in general usage situations, whereas other adaptations will be applicable only for very specific systems, user groups or usage contexts. As the variety of adaptations appears to be only restricted by the designer's creativity and experience, and the system's support for these adaptations - both at interface level and at the implementation level - it is nearly impossible to provide a complete taxonomy of adaptive hypermedia techniques. Instead, in this section we provide a rough categorization, present examples of techniques used in earlier systems and deal with several issues that come with these techniques. As a starting point we take the two main categories identified by (Brusilovsky 2001): adaptive presentation and adaptive navigation support.

Adaptive presentation techniques works on the page level. Items that may be adapted include text, layout, graphics or any other form of multimedia. A significant amount of research has been carried out on *text adaptation*. The general goal of text adaptation is to hide from the user some parts of information that are deemed not to be relevant for the user (Brusilovsky 1996) for a variety of possible reasons: the user may already be familiar with the information, the information may be beyond the user's knowledge level, the user might not be interested in the information or the user or the text might consume too much screen real estate of the user's device. The most popular form of text adaptation is *canned text adaptation*: conditioning the view on certain ready-made text fragments. Text fragments may be removed from the running text or hidden until the user clicks a certain *hotword*; one of several alternative fragments can be chosen to be displayed; text may be annotated with relevancy indicators or any other kind of marking.

Adaptive navigation support works on a hyperdocument's link structure. Hypermedia links may or may not be embedded in the content regions. Links within the text are usually associative links, which interlink semantically related concepts. Links that are not embedded in the text - such as menus, indexes, and site maps - expose a document's primary structure and therewith function as contextual navigation aids (Miles-Board, Carr & Hall 2002) (Park & Kim 2000). Both types of links can be adapted to the user needs. Disabling, removing or annotating associative links (Brusilovsky 2001) can help users to find relevant items more easily, but does not provide the context information needed to prevent disorientation or cognitive overload. Therefore, many research projects have focused on adaptive contextual navigation aids. These projects can be categorized as either focused on the user's local or global spatial context or on the user's temporal context.

Strategies found in the former category include personalizing or adding textual or graphical views of relevant parts of the document structure (Brusilovsky 2001). Site maps, contextual menus, direct guiding and recommendations help users to decide where to go. Most strategies in the latter category concentrate on various visualizations of previously visited pages. Several examples are mentioned in (Tauscher & Greenberg 1997). As a large number of user navigation involves revisits (Tauscher & Greenberg 1997) (Cockburn & McKenzie 2001), users highly profit from reminders where they have been.

Each form of navigation support addresses one or more user needs with respect to ease of navigation. Unfortunately, one cannot provide users with all contextual information at once: even if sufficient screen space is available, users will most likely be overwhelmed by the quantity of navigation suggestions (Park & Kim 2000). Therefore, the best approach is to provide users with only those associative, spatial or temporal navigation aids that match their navigation strategies and address the problems they are experiencing.

Many adaptive hypermedia techniques on the Web might be regarded as 'fixes' for the poor context information provided by current Web browsers. The implementation of the Web does not support many of hypermedia's rich structuring, navigation and annotation features (Vitali & Bieber 1999). In particular, the lack of link annotations makes it hard to identify the type of relation between two pages. Link types such as 'explanation', 'further details', 'contrasting argument' convey the relation between a link's source and destination. More general link categories used in hypertext systems are 'glossary', 'annotation', 'structural' and 'associative' (Bieber et al. 1997). Moreover, current browsers do not convey essential information on the link behavior and the user's link history, which includes information on whether the link leads to a page on the same or on a different site, when the link was last followed, how often the link has been followed, the destination page title and whether the page will open in the same or in a new window (Weinreich & Lamersdorf 2000).

2.3 User modeling

In the previous section it was mentioned that personalization is based on a user model, which contains a selection of data on the user characteristics, past interaction with the system and environment data. In this section we explain how these user models can be built and what they look like.

Three tasks can be separated in the user modeling process: acquisition of the data, inference of knowledge from the data and representation of the user model (Pohl 1999). As discussed in the previous section, adaptive hypermedia systems traditionally aim to adapt the interface to better fit the user's knowledge, skills, interests and goals. As a consequence, the inference techniques and representations used might be called *mentalistic* (Pohl 1999); the user model explicitly represents the relevant aspects of the user's 'mental state' as closely as possible, and the model is built from human-like inferences. This so-called *knowledge representation* approach has as an advantage that the process is intuitive and the models are interpretable and reproducible; however, it is known to have limitations in scalability and extendability (Pohl 1999) (Zukerman & Albrecht 2001). As an alternative approach, statistical models and machine learning technique have become popular, in particular in the context of the Web. The implicit, statistical approaches are more flexible and better suitable for dealing with huge quantities of data, at the cost of having an implicit representation of the user model, often directly stated as adaptation decisions instead of assumptions on the user characteristics.

In this section we provide an overview of techniques for data acquisition, knowledge inference and user model representation as used in various systems. As data acquisition is the first step for both knowledge representation techniques and statistical modeling, this issue will will be dealt with in the next subsection. We continue with two subsections on the two respective approaches to the inference and representation of knowledge.

2.3.1 Data acquisition

In order to provide relevant adaptations, the system must acquire some information on the user and the user context. If the information is not readily available, the most direct way to do this is to ask the user (Rashid, Albert, Cosley, Lam, McNee, Konstan & Riedl 2002). In general, users are the ones who know best about their interests, preferences and current goals (Nielsen 1998). Further, there is evidence that involving the user in the process creates focus and engagement with the system (McNee et al. 2003); users who spent the effort of filling a user model with explicit ratings rated the system understanding of their tastes higher than users who did not. User input is often gathered upon the first use of a system using forms or questionnaires. User input is also commonly gathered while the user interacts with the system. One possible option is that users make adaptations themselves by ordering lists, enabling or disabling options, dragging interface elements or by any other specific interaction with the system (Brusilovsky 1996). Another option is that the user gives relevance feedback. In recommendation systems such as Movielens (McNee et al. 2003) feedback is an essential part of the system, as the recommendation process mainly relies on user ratings and reviews of movies.

However, in many cases it is not possible or desirable to ask the user the information needed to incorporate into the model. Whereas users will be willing to spend time on creating a profile for certain systems, in many cases users just want to start working on their tasks without first reading manuals, following an introductory tour or filling out forms, even though there are apparent benefits. This is particularly the case with systems that are used as a matter of daily life, like Web-based systems. Further, providing feedback will probably interfere with the task at hand, which might disrupt the normal user behavior that the system wants to adapt and annoy the user (Kellar, Watters, Duffy & Shepherd 2004). Moreover, many characteristics cannot be provided by the users themselves and need to be inferred from potentially long questionnaires or user tests. For example, individual differences such as learning ability, cognitive strategies and cognitive abilities are aspects that can mainly be measured with lengthy psychological tests (Browne et al. 1990). These questionnaires and tests may be very helpful in laboratory studies (e.g. (Ahuja & Webster 2001) (Herder & Juvina 2004)), they are not suitable as a part of an adaptive system's data acquisition process. In addition to the issues mentioned above, it is hard to design effective interfaces for soliciting user feedback (Waern 2002) (McNee et al. 2003).

For the above reasons many adaptive systems attempt to infer knowledge directly by unobtrusively monitoring the user interactions with the system, as an alternative for or in addition to explicitly provided information. Many interactions contain meaning in themselves, such as page visits, bookmarking or saving actions, queries issued by the user and items inspected or bought from an e-commerce Web site (Claypool et al. 2001) (Kobsa et al. 2001). Other interactions need to be combined or interpreted in order to become meaningful, such as key strokes, mouse clicks and eye gaze behavior. Several research projects have been carried out to relate observable actions to higher level user characteristics such as the user's vulnerability to getting lost (Smith 1996) (McEneaney 2001) (Otter & Johnson 2000) (Herder & Juvina 2004). Results from empirical studies and models of user behavior are useful frameworks for interpreting user actions and the design of inference mechanisms. In the next chapter we will deal with these frameworks in more detail and give an overview of results from field studies and theoretical work.

2.3.2 Knowledge representation

Knowledge on the user may be represented in many different formats. Which format to use highly depends on the goals that the user model serves and the size and complexity of the model. The most basic model is a simple collection of variables and associated values (Kay 1993). These variables can represent a variety of independent user characteristics, such as the user's demographics, the liking of certain interface elements and knowledge on certain topics. These variables may be combined at will for adaptation decisions in the form of basic rules. An example rule might indicate that if a user's age is lower than eighteen and the user is female a selection of news items interesting to young females should be made. Due to the flatness of the model, it is hard to make more complex deductions.

A *hierarchical representation* allows some aspects of the user model to be regarded as higher level and more general than others. In contrast to the flat model, hierarchical structures represent user characteristics and relations between these user characteristics. A common hierarchical structure is a tree or a directed acyclic graph (Rich 1979). The hierarchies are typically hand-crafted based on the domain knowledge of the designer.

A popular user modeling paradigm that often makes use of hierarchical structures is *stereotyping*. As explained by Rich (1979), people often make assumptions of other people, often based on fairly simple observations - or *triggers*. For example, if one knows that someone is a judge, it is probable that the judge is over forty, well-educated, pro-establishment, fairly affluent, honest and well-respected. While not all of the *default assumptions* may be true for this particular judge, the assumptions tend to work fairly well in every-day life and can be *overridden* when shown otherwise. Stereotyping is particularly useful when a solid amount of statistical data of user groups is available. In figure 2.1 a part of a sample stereotype network is shown that may be used for creating a personalized information package for university members. Suppose we have evidence that Bob is a first-year computer science student who has made use of the University's piano studios. Logically, we would select information relevant for first-year students, on computer science and on music events on the Campus. In addition, it would be natural to assume that Bob would appreciate information on culture in general as well. However, in a welcome questionnaire Bob has indicated that he is not interested in theater at all - this exception is included in his stereotype model and we won't bother him with the long list of theater performances planned this year.

A related hierarchical model is based on the domain structure. The user model can be regarded as an overlay of the domain structure. For each item in the domain overlay model certain attributes can be set representing the user's knowledge of, interest in or any other relation between the user and the item (De Bra, Aerts & Rousseau 2002). In the most simple case, the items directly

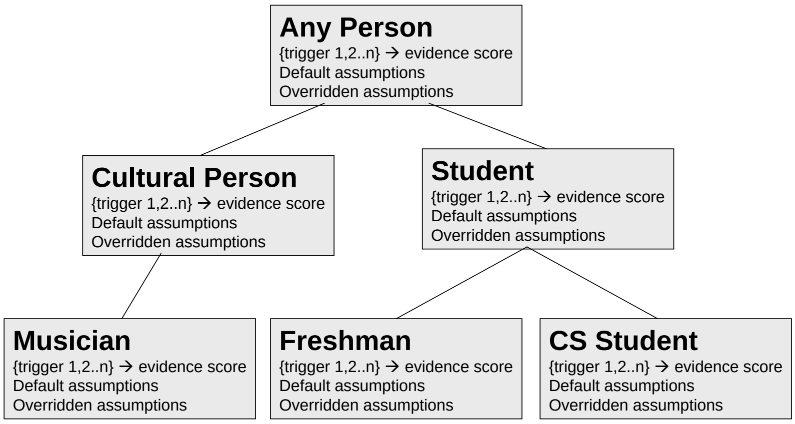

Figure 2.1: *Sample stereotype hierarchy, adapted from (Rich 1979)*

refer to corresponding hypertext nodes. Alternatively, the items refer to more abstract concepts that may relate to more than one hypertext node. A system that makes sophisticated use of a concept hierarchy is the AHA! system, developed by the group of Paul De Bra (De Bra, Aerts, Smits & Stash 2002). An example application is depicted in figure 2.2. For each concept the user knowledge is stored. Knowledge on certain concepts may propagate to knowledge on other, more generalized concepts. For certain concepts prerequisite knowledge on other concepts is required. Knowledge propagation and prerequisite knowledge relations are depicted in two different graphs. Behind the screens the node attributes and relations are coded as rules.

A more sophisticated yet more complicated approach is the use of logic-based representation and reasoning. A comprehensive overview is given in (Pohl 1999). Rules as described above can well be represented in first order predicate logic (FOPC). In addition, in FOPC it is possible to state generalized rules and attributes that always hold for all concepts in a certain category or hold for at least one concept. Further, special predicates and modal logic operators can be used for expressing the difference between observations and inferred assumptions and uncertainty of inferences. Dedicated logic programming languages such as Prolog can be used for reasoning. Forward reasoning is used for acquiring primary and secondary assumptions from observed data - like the user's possible membership of a certain stereotype, the user's knowledge on or interest in certain topics or other higher level user characteristics. Analogously, backward reasoning can be used for querying for implicit knowledge contained in the model. Although backward reasoning combined with limited forward reasoning is reported to be more goal-directed and therefore more efficient in number of inferences, most systems mainly rely on forward reasoning. This may be because forward reasoning has one important advantage: computation can be done off-line, which provides a far

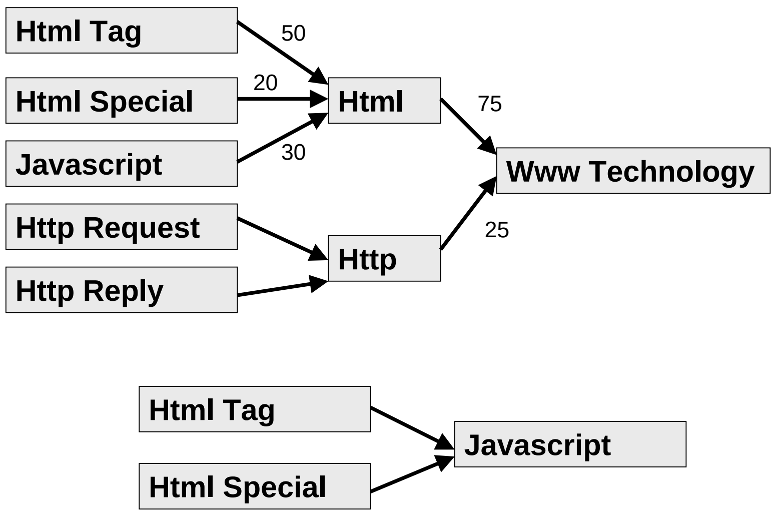

Figure 2.2: *Example AHA! domain overlay model (De Bra, Aerts & Rousseau 2002). The upper graph depicts how knowledge is propagated from more detailed to more general concepts. Edge labels indicate the proportion to which knowledge on the source item contributes to knowledge on the target item. The lower graph depicts prerequisite relationships between the concepts.*

better response time for adaptive systems than having to wait for query results.

From the above it can be observed that knowledge representation and reasoning mechanisms are quite interrelated concepts. User models may contain both concrete *observations* or *data* and inferred knowledge, which are generally called *assumptions*. Whereas the former category contains information that may be held as 'true', the information contained by assumptions is uncertain. This uncertainty may be inherent to the reasoning process, but it may also well be that the design and underlying choices of reasoning mechanisms are questionable. In general, any choice of user characteristics to take into account and any reasoning introduces uncertainty in the user model that needs to be evaluated. This matter will be dealt with at the end of this chapter.

In order to save designers from having to implement user modeling mechanisms themselves, a number of generic user modeling systems have been developed (Kobsa et al. 2001) (Kay, Kummerfeld & Lauder 2002) (Kobsa & Pohl 1995). However, these systems have rarely been used outside of the institution where they were developed (Kobsa et al. 2001).

2.3.3 Statistical methods

As mentioned before, from the last section it can be observed that knowledge representation techniques mainly adhere to a mentalistic paradigm. In other words, they try to represent the user and the user context as realistically as possible. User behavior is treated as an information source and not regarded as a phenomenon that should be analyzed and modeled *per se* (Kobsa et al. 2001). Kay (1993) recognized the need for developing *pragmatic* light-weight models and modeling tools, developed with the actual goal in mind. According to Orwant (1995) this was the aim of his DOPPELGÄNGER system: ‘useful inferences about a user’s cognitive state are made without resorting to a formal cognitive model’. As a further motivation for using implicit user models based on statistical modeling of user behavior, Webb, Pazzani & Billsus (2001) described the knowledge-based approach as ‘creating a new model by modifying an initial model’, which suggests that the approach is severely influenced by the designer’s way of modeling the domain, the user and the user context.

Statistical modeling and machine learning techniques are focused on directly exploiting user behavior rather than forming assumptions on the user. Similar to the knowledge representation approach, the underlying assumption is that user interaction with a system is predictable to a certain extent. In contrast, machine learning techniques make less assumptions of the initial model. In general three approaches can be identified (Zukerman & Albrecht 2001):

- detecting patterns in user behavior;
- matching user behavior with the behavior of other users;
- classifying users or hypermedia content based on user behavior.

The first approach is useful when the aim of the adaptive system is to respond to recurrent behavior or to infer items that may be of the user’s interest. The second approach is useful when a user behaves in a similar way to other users and is typically used for making recommendations involving items not seen before. Common applications of the third approach include stereotyping and the modeling of user interests. In the remainder of this section several example adaptive systems are mentioned and the methods used by these systems.

Smyth & Cotter (2002) used Markov models for creating adaptive menus for mobile Web portals. From the main page the menu subitems were ordered according to the probability that a user will visit them, regardless the distance in clicks. Pages at a lower layer were ’promoted’ to a higher layer if the probability of being visited exceeded a higher layer item’s probability; the latter got ‘demoted’. The Lumière project employed Markov models for temporal modeling of user goals in adjacent time periods (Horvitz, Breese, Heckerman, Hovel & Rommelse 1998). By constructing aggregated trails, the WUM system (Spiliopoulou, Faulstich & Winkler 1999) is able to answer questions like ‘what path do most users follow from page *a*’ or ‘how would page *b* most likely be reached from this point’.

Content-based filtering is used by Syskill & Webert (Pazzani, J. & Billsus 1996), an interface agent that helps users distinguishing interesting pages on a topic from uninteresting ones. For each topic a list of *hot* and *cold* words is learnt by the user explicitly rating pages as interesting or uninteresting; hot words are words that appeared frequently on pages marked interesting but infrequently on pages marked uninteresting. A Naive Bayesian Classifier is used for predicting whether a candidate page is interesting or not based on the presence or absence of a limited set of these hot and cold words and the evidence for user interest provided by these words. After some experimental runs it turned out that a combination of user-provided words and inferred hot and cold words performed best. WebMontage (Anderson & Horvitz 2002) assembled a personalized portal page, containing pages most likely to be visited based on evidence from the overall popularity, popularity within a three-hour time span and the predominant topic of the pages viewed during the last four-hour block of time. PageGather (Perkowitz & Etzioni 2000) is a cluster mining technique for finding collections of related pages at a Web site based on the co-occurrence frequencies between pages.

2.4 Structure and evaluation of adaptive systems

Evaluation plays a major role in the software engineering cycle in order to verify that a system meets user expectations on functionality, reliability and performance. Likewise, the goal of user interface evaluation is to find out whether an interface provides the functionality that the user needs, to investigate the learnability and usability of the interface and to identify specific problems with the design (Dix et al. 2004). One of the most important challenges for evaluation of adaptive user interfaces is to show that adaptive behavior does *improve* the interaction with the user (Höök 2000). In addition there are more theoretical issues that are of interest (Weibelzahl & Weber 2002):

- what will be the effect of a change in the adaptation model on its performance;
- what will be the effect of a change in the adaptation goals on the adaptation model's performance;
- what will be the effect of a change in the application domain on the model's performance;

In other words, one of the major concerns of adaptive systems to be addressed is the fact that an adaptive system changes its behavior based on assumptions on the users and their user contexts. Empirical evaluation should answer the question whether the assumptions hold, whether the characteristics taken into account are relevant for adaptation purposes and whether the adaptations address the issues as represented by the assumptions (Weibelzahl 2005). Adaptation is sometimes characterized as *deferred design*; instead of searching for a general design that is suitable for most users, adaptive interface designers build in mechanisms able to determine a user's particular needs and automatically adapt the interface accordingly (Browne et al. 1990). Despite the potential advantages in terms of the system's general applicability and life-span, there are risks involved in not knowing exactly how the interface will behave. As an example, the Lumiere Project (Horvitz et al. 1998) designed a Bayesian Network for inferring user goals and needs in Microsoft Office applications in order to provide users with suggestions relevant to their tasks. The commercial Microsoft Office Assistant that was designed based on the results of the Lumiere Project used slightly less advanced reasoning mechanisms, which made the assistant's behavior less intelligent but at least more predictable.

Extensive evaluation of adaptive systems is rarely reported in publications on user modeling systems and adaptive systems (Weibelzahl 2005). According to Weibelzahl, most adaptive hypermedia researchers have no experience with typical procedures and methods that are required to conduct an empirical study. However, there are additional problems associated with evaluating the benefits of adaptation. A common practice is to compare the performance of the adaptive system with the performance of its non-adaptive counterpart. If adaptivity is an integral part of the (deferred) system design, the non-adaptive system is handicapped and therefore not suitable as a baseline for comparison (Höök 2000). Further, in order to gain more insight in the adaptive system's behavior and uncertainty introduced by this behavior in different usage contexts, insight is needed in *how* the adaptation process works.

To address these issues, the concept of *layered evaluation frameworks* has become a popular paradigm (Paramythis, Totter & Stephanidis 2001) (Weibelzahl 2005) (Brusilovsky, Karagianidis & Sampson 2001). Layered evaluation breaks the adaptation down into its constituents and each of these constituents should be evaluated separately where necessary and feasible. The main idea behind the approach can be traced back to Browne et al. (1990), who propose a number of measures related to different components of a logical model of adaptive user interfaces.

2.4.1 Layered Evaluation

Recently, several frameworks for layered evaluation of adaptive applications and services were proposed by a number of researchers (Brusilovsky et al. 2001) (Karagiannidis & Sampson 2000) (Weibelzahl & Weber 2002). Although these frameworks are described at different levels of granularity, in essence they separate the process in the evaluation of the interaction assessment phase and the evaluation of the adaptation decision making phase (Karagiannidis & Sampson 2000). The basic intuition behind this approach is that unsuccessful adaptations might be due to incorrect assessment results, or to improper adaptations based on a correct

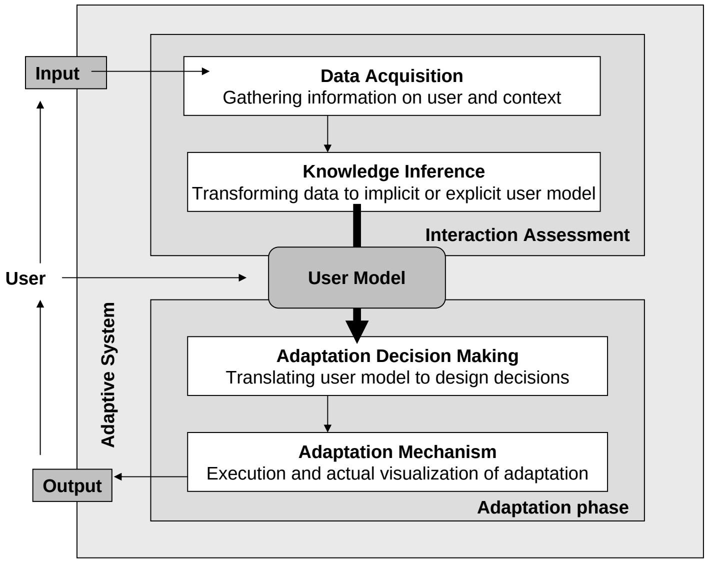

Figure 2.3: *High-level model of an adaptive system*

assessments. Layered evaluation of adaptive systems appears to be a promising approach, as shown in a case study described in (Brusilovsky et al. 2001).

In (Herder 2003) we proposed an utility-based approach to the layered evaluation process. The basic idea is that the added value of an adaptive system can be expressed by a utility function U (Russell & Norvig 1995) that maps selected, measurable criteria with respect to the performance of the adaptive system to a quantitative representation. If one would compare an adaptive system with its non-adaptive counterpart, the value of adaptation is the difference in utility between the two systems. The main advantage of layered evaluation methods is that it separates the utility function in several functions. In the previous section the basic intuition behind this approach is explained. The undesirable effects of correct adaptations based on incorrect assumptions can be modeled by the utility function as well. Suppose there is a utility function U_1 that maps the interaction assessment and the resulting user model to a real number that represents its correctness. Suppose there is also a utility function U_2 that maps a system, given some user model, to a real number that represents user satisfaction or performance. We can then express the whole utility function as $U = U_1 U_2$. It is clear that the latter utility function better indicates the usability of an adaptive system. As one would expect, an adaptive system that coincidentally makes correct decisions based on wrong assumptions will be rated poorly by this function (Russell & Norvig 1995).

In figure 2.3 a high-level model of an adaptive system is depicted, which represents the different steps of the adaptation process. The first phase involves the interaction assessment. Accuracy measures of interaction monitoring and inference methods are needed for this purpose. At the point of *interaction monitoring*, the input data does not necessarily carry any meaningful information (Weibelzahl 2005). The main concern at this point is that the data is *reliable* and *sufficient* for the purpose of the system. As mentioned earlier, user data may be explicitly provided by the user or unobtrusively gathered by a monitoring mechanism. In the former case, it would be worthwhile to evaluate the feedback mechanisms used (Waern 2002). In the latter case, it is important to investigate the noise in the data. In real-world situations the user environment might introduce a large amount of variability, for example due to distraction by the presence of coworkers. More importantly, as we will deal with in more detail in chapter 4, the captured usage data might be incomplete or noisy, due to the way it is captured.

The *knowledge inference* phase is the stage of adding 'meaning' or 'relevance' to the data (Weibelzahl, Paramythis & Totter 2003). The interpretation may be straightforward when there exists a direct one-to-one mapping between the raw input data and their semantically meaningful counterparts. This is the case with data explicitly provided by the user; user actions may be meaningful on themselves, such as a user's location or the number of a user's undo-actions. Uncertainty is introduced when the data needs to be interpreted, either with hand-crafted knowledge representation techniques or with machine learning techniques. In the field of information retrieval, the concepts of *precision* and *recall* are used to evaluate the correctness of predictions based on a user's detected patterns, deduced similarity to a group or interest in items. For explicit user models that contain assumptions on user characteristics, expert-based evaluation and elicited user feedback would be needed in the early design and evaluation stages (Paramythis et al. 2001). Background knowledge on user behavior, derived from a cognitive framework or predictive framework on user behavior appears to be indispensable at this point.

An intermediate point between the interaction assessment and the adaptation decision phase is the *user model*. As explained before, the model may represent assumptions on user characteristics, or predictions on user actions, or interests that are closely related to adaptation decisions. Besides the validity of the deductions, as dealt with in the previous paragraph, *comprehensiveness* and *accuracy* are two other criterions (Weibelzahl 2005). Comprehensiveness is a two-fold criterion. First, the designer needs to understand what is in the model - in other words, to be able to relate input data to the knowledge in the model and to understand how the model can be used for adaptation decisions. Second, it might be desirable that the user understands the model in terms of its attributes and why the model is in its current state (Paramythis et al. 2001). Accuracy refers to the representation of the items in the model: is the chosen representation sufficiently expressive? As the user modeling process necessarily is associated with uncertainty, it would be desirable to have uncertainty indicators or confidence intervals where feasible and desirable. Further, the stability of the model might be an issue.

Based on the user model, the *adaptation design decision* is made. In some cases there is a one-to-one relation between an item in the user model and the corresponding adaptation decision. In other cases some form of reasoning is used. Interestingly, the reasoning used by most systems in this phase is fairly straightforward in the form of simple if-then rules. Evaluation of this phase is mainly geared to appropriateness, necessity and user acceptance. Finally, the *adaptation mechanism* should be evaluated for its timeliness, obtrusiveness and means for user control. A rather extensive overview of these matters can be found in (Weibelzahl 2005).

The above steps of layered evaluation provide evidence to what extent the adaptive system parts fulfill their goals in terms of reliability, understandability, completeness, correctness and uncertainty introduced. However, there is still the need to get the big picture: does the adaptive system serve the goals it was intended to do. If an objective measure is available, it is a matter of conducting an empirical study - taking precautions, problems and pitfalls as mentioned in (Weibelzahl et al. 2003) into account. If no objective measure is available, one has to resort to *theory assessment measures* (Browne et al. 1990), of which the means for determining the objective measure may rest on an untested theory. In particular in the field of adaptive Web-based systems no integrative theory on Web navigation and associated issues is available; rather, there are several cognitive models and empirical data from laboratory studies and field studies that can be used as a baseline. In addition, attempts have been made to relate user Web navigation to perceived usability problems in terms of *behavioral complexity* (Weibelzahl 2003). If adaptive systems are aimed at improving usability in general, more insight is needed in how people actually navigate the Web and what goals they have.

From the above it can be concluded that many different evaluation criteria can be thought of. In order to compare systems or approaches, one needs to decide on sets of criteria to be used in several domains. These domains can range from broad (e.g. hypermedia in general) to more narrow (e.g. educational hypermedia). In this process, previous work on general usability matters, as carried out by e.g. the W3C, as well as overviews of the current state of the art in adaptive hypermedia - e.g. (Brusilovsky 2001) - should be taken into account. Once such common criteria are established, researchers will be able to employ them to guide them in their research and to compare results.

2.4.2 Separation and reintegration of concerns

From the above separation of concerns in adaptive hypermedia systems, it can be observed that from a computer science point of view the problem of discovering user needs and representing them in a suitable format is the most challenging part of the adaptation process. The actual adaptation decision may be regarded as a mere user interface design issue. A complicating matter is that both processes are closely related in a feedback loop. In order to evaluate the interaction assessment or learnt user needs, a proper interface is required; in order to evaluate an adaptive interface, a proper user model is needed. There are some ways to escape from this loop. In chapter 5 we describe an experiment that evaluates the impact of link annotation on user impressions and navigation behavior by using a fixed user model (Juvina & Herder 2005). Although it might seem less challenging than building and evaluating an entire adaptive system, we feel that more of these purposely limited studies are needed to fill gaps in our knowledge on user behavior on the Web and the impact of adaptive interface elements.

In the last section we expressed the layered evaluation process in terms of utility functions that should be built from assessments of the system components. This fairly simple observation provides a mechanism for reintegrating the evaluations of the components to an integrative evaluation measure (Weibelzahl 2005). These sort of mechanisms are needed if we want to move beyond one-shot learning algorithms to adaptive hypermedia that continually improves with experience (Perkowitz & Etzioni 1997). Adaptive systems without a feedback loop can be regarded as reflex response systems that react to the environment as built in by the designer. (Browne et al. 1990) envisaged more advanced systems that are able to learn from the effects of the adaptations and to learn more effective adaptive responses by trial-and-error, self-evaluation and generalization. In order to design systems that can reason about user actions, we still first have to address the first challenge (Perkowitz & Etzioni 1997): 'Can we formalize user navigation of the Web as a planning process that is amenable to goal recognition? Do user actions on the Web carry enough evidence of their purpose'? The remainder of this thesis will concentrate on this challenge.

Chapter 3

Empirical and theoretical models of Web navigation

3.1 Introduction

The task of user interface designers is to understand the users and their tasks, and how to translate this knowledge into a working system. Whereas this might seem a fairly straightforward process, it turns out that it is not at all intuitive or easy to design consistent interfaces that fit all user needs. Like in regular engineering tasks, previous experiences and theory on human computer interaction can help designers in the process. However, at the moment there is no general and unified theory of human computer interaction, nor is it certain whether such a theory can be derived (Dix et al. 2004). In particular the design of Web-based interfaces is still in an early stage of development. Incorporating user-adaptivity into the system can be considered as a strategy for simplifying the task of designing an interface that fits *all* users into the task of designing interfaces that can adapt to the needs of *individual* users. Yet, this still requires the interface to be 'user literate'; the program needs to keep track of the current user activities and to translate them into adequate interface solutions.

Designers have a certain goal in mind when incorporating adaptivity into the system. In more restricted domains, such as e-learning, the goals are relatively well-defined and it can be evaluated whether the system reaches these goals. For adaptive systems that aim to support general activities as Web navigation, it is harder to state the goals in such a manner that they can be evaluated. Knowledge is needed on the activities that the Web is used for, on how users navigate the Web, and on strategies that users employ for locating and relocating information and services.

Empirical data is a valuable source of knowledge on how users interact with Web-based systems. In the next two sections we provide an overview and discussion of several laboratory studies and field studies that aim to provide these empirical data. Laboratory studies are typically short-term and focused on one particular aspect of Web navigation or interaction tools. Whereas less focused, field studies - in particular longer-term studies - provide insight on how the Web is used in daily life, the tasks that users carry out and the usability issues that users perceive. Two categories of field studies can be distinguished: *observational studies*, which provide data on the users and their tasks as collected through video cameras, diaries or questionnaires, and *click-through studies*, which provide quantitative data on the actual user navigation behavior. A matter of concern for interpreting the outcomes of field studies is that they might be biased by the participant pool and the method of observation (Hawkey & Kellar 2004). Furthermore, although empirical data provides qualitative and quantitative data, they are merely descriptive and may provide little insight on the design implications - what tool would support these activities and how can we evaluate whether this tool actually works. Therefore, the data needs to be interpreted and modeled in a theoretical framework that explains the underlying processes of observed patterns. A specific flavor of theoretical models, *cognitive architectures*, aim to explain the observations in terms of cognitive processes found in studies and in the literature. Cognitive architectures are more behavioristic: they aim to simulate user navigation by a production system. In the fourth section of this chapter we introduce two well-known theories on Web navigation - the theory on *information foraging* and the *CoLiDeS* model - as well as a number of other more limited models that provide complementary insights. In the last section of this chapter we summarize, discuss and elaborate on the issues dealt with in this chapter.

3.2 Laboratory studies

In laboratory studies, users are taken out of their normal work environment to take part in controlled tests. As the experimenter can control and manipulate the conditions in which the study takes place, laboratory studies allow for analysis of one particular topic - such as the influence of individual differences or the impact of novel tools. In addition, many usability laboratories are equipped with sophisticated recording and analysis facilities that cannot be used in the work environment. However, taking away the users from their natural work environments may result in recording situations that never arise in the real world; users are likely to act differently when they know that they are being observed, and users may act differently by lack of their usual context, such as calendars, books or interruptions (Dix et al. 2004).

In particular, evaluation of specific interface features calls for controlled conditions, which can only be created in laboratory settings. Several studies have been conducted to investigate the impact of different kinds of navigation support on user navigation. The results of these studies are discussed below.

Park & Kim (2000) studied the impact of contextual navigation aids on task performance and user perception on two different Web sites, a well-structured e-commerce site and an ill-structured content dissemination site. Both sites provided a global context in the form of a menu bar. Whereas the e-commerce site had a well-balanced product hierarchy, the content dissemination site's structure was more random, due to the domain covered. On top of the global contextual navigation support, two different kinds of context were provided: the local *structural context*, in the form of links that allowed users to move up or down in the hierarchy, and the *temporal context*, a list of the most recently visited pages. Forty undergraduate students were asked to carry out several tasks in one of the four conditions: control group, structural context, temporal context, both contexts. Results indicated that in both sites either form of added contextual navigation support increased both task performance and user perception. However, if the two kinds of support were combined, the added value of the temporal context was only observed in the content dissemination site. The authors hypothesized that this may be caused by the less well-defined structure of the latter site; due to the less regular site structure, which makes the temporal navigation support more effective in certain situations. We believe that the nature of the tasks might have been of influence as well; information finding is generally less well-structured than product finding. As will be discussed later on in this chapter, backtracking activities are a common and effective strategy in information finding tasks, activities which are supported by temporal navigation aids. Whereas an excellent study, a drawback is that only contextual navigation support is taken into account; associative links - which reside in the running text - were not taken into account.

Katz & Byrne (2003) investigated the usability of several alternative menu structures of fictitious e-commerce sites. They hypothesized that if a menu structure does not match the user needs, they would search rather than follow links. Thirty-two participants were presented the portal sites of four 'competing' fashion stores that differed in *breadth* - the number of menu items - and *specificity* of the link labels - specific link labels included items such as 'men's socks', unspecific link labels included items such as 'relaxed wear'. Thirty-two undergraduate students were presented four alternative sites - small breadth, unspecific labels; small breadth, specific labels; high breadth, unspecific labels; high breadth, specific labels. The participants were asked to indicate whether they would prefer to search or to select a menu item. Results indicate both broader menus and more specific labels resulted in less search activities. However, the breadth of the menus had a larger effect than the specificity of the links. This suggests that the number of available items is more important than the actual labels of these items. The authors hypothesize that, although some users have a stronger preference for searching than others, this bias is strongly influenced by the navigation options at hand. Broad menus with specific link labels carry a high *information scent*, which makes users decide to follow links rather than having to formulate a search query. We think that the higher impact of the menu breadth might be inherent to the domain of e-commerce sites; in less goal-directed information finding tasks a higher number of navigation options might lead to an increase in cognitive load and a smaller number of link options might be desirable in this situation. Further, only two specific menu breadths were taken into account; it is very likely that at some point broadening the menu might cause more problems rather than enhancing the information scent. We will come back to the influence of the breadth effect in the theoretical section on *information foraging*.

Obendorf & Weinreich (2003) conducted two studies on the visual aspects of associative links - links that appear in the running text. Their study emerged from the observation that associative links are relatively rare on the Web. An explanation for this effect is that Web authors judge highly linked text as unfavorable, because the visual appearance of links - blue underlined text - might interfere with readability. In the first study they compared performance of twelve students and academic staff on a number of text reading tasks. In one condition, ordinary underlined links were present in the text; in the second condition, the link visualization was replaced by a changed background color; in the control condition, no links were present in the text. Text reading performance was best in the control condition, followed by the alternative link visualization condition; text reading performance was significantly worse in the underlined link condition. In a follow-up study task performance in a standard Web hypertext was compared to task performance on a Web hypertext with the links only visible on-demand. Results indicate decreased performance in the links-on-demand condition. The authors conclude that the non-linear character of the Web asks for instantly recognizable links; however, more attention needs to be given to the actual visualization of the link anchors.

Cockburn & Jones (1996) investigated the user's mental model of the behavior of the back button's stack-based history. Upon each page visit, the page visited is put on top of the stack. Using the *back* and *forward* buttons, the user can move down and up the stack. However, if a user follows a link or types in an url, the pages above the current position are popped off the stack. Any user whose model of the back button is temporal rather than stack-based will find the back button's behavior unpredictable. Eleven participants, all academic computer science staff, were asked to describe the back button's data structure. Only one of the participants correctly identified the structure as a stack. Afterward the users were asked to follow two sequences of link traversals and backtracking activities. Having completed these actions the participants were asked to state whether the *first* page of the sequences could be revisited using the back button. For both sequences, only two participants confidently responded correctly. The other participants were either not sure or responded incorrectly. The authors conclude that, whereas users are surprisingly successful in working with incorrect mental models, more effort should be spent on explaining the user the internal working of interface elements such as the back button.

From the studies described in this subsection several observations can be made. First of all, they show interesting, isolated aspects of Web navigation that raise important questions on the design of contextual navigation support, breadth and specificity of menu structures, visualization of link anchors, and tools for back-tracking. In that respect, the studies are extremely valuable. On the other hand, most of the results are merely confirmations of the authors' initial hypotheses, which were already supported by anecdotal evidences. Further, as might have become apparent from the descriptions of the studies, all participants were experienced internet users with an academic background; there is no evidence that the results can be generalized to all users, or to different Web sites or user contexts.

3.3 Field studies

Field studies take the experimenter out into the user's work environment. As the context is retained, user behavior is recorded in their 'natural environments'. In addition, long-term effects may be recorded - which is next to impossible in laboratory studies. Unfortunately, also in field observations the participants are likely to be influenced by the presence of the experimenter or logging equipment. The presence of noise caused by interruptions can be seen as an advantage, as these interruptions will invoke behavior that occurs in daily life; evaluation of specific interface features would largely be hindered by the noise. Field studies are considered to be preferred to laboratory studies, as it allows to study the interaction as it occurs in actual use (Dix et al. 2004).

In *observational* field studies user actions are recorded through video cameras, diaries, interviews and questionnaires. Participants are actively followed and often they play an active role in the recording process. Observational studies are particularly useful for obtaining qualitative data on usability matters - although protocol analysis also allows for quantitative analysis. *Click-through* studies make use of unobtrusive data gathering techniques that register low-level events such as key strokes and Web activity. The unobtrusiveness of the technique makes it suitable for long-term studies. A disadvantage of click-through studies is that the low-level data does only indirectly indicate why certain actions were performed and how they are structured. On the other hand, observational studies may rely heavily on the interpretation by the experimenter, which might introduce a certain bias (Dix et al. 2004).

3.3.1 Observational studies

Perhaps the most influential observational study is carried out by Byrne, John, Wehrle & Crow (1999). In this study, the Web activities of ten participants, all with an academic background, were observed by a video camera during a normal work day. From the video transcripts a taxonomy of user tasks was constructed. As there often was the need to violate the hierarchical structure - some events

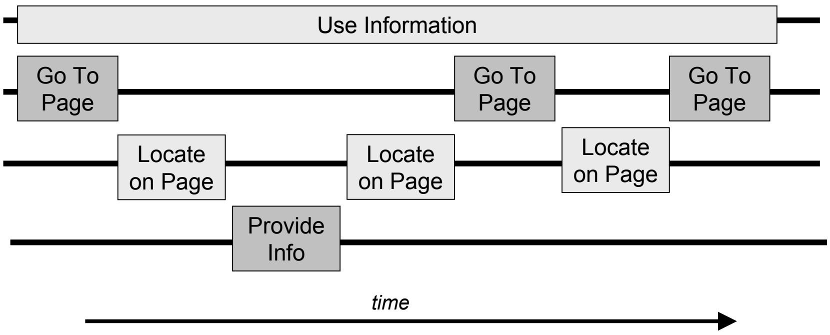

Figure 3.1: *Timeline for nested subtasks with time represented on the horizontal axis*

may occur as a sub event of another event or merely on itself - a less strict categorization was made, playfully named 'taskonomy'.

- *use information:* reading or viewing the document content, download or duplicate;
- *locate on page:* locate 'something interesting', related concepts, a specific string or image;
- *go to page* using hyperlink, back button, bookmark, history list or address bar;
- *provide information:* for example a search string, a shipping address or a survey response;
- *configure browser:* adding a bookmark, change the cache size, window management - including scrolling;
- *react to environment:* respond to a dialog, respond to dialog change, page reload, etc.

Whereas the taskonomy might be a good start for annotating browsing behavior, it is clearly not complete or definite; as an example, the prominence of search actions might require a separate subcategory; further, page scrolling is a 'browser configuration' action that naturally takes place during Web navigation, whereas changing the cache size is an action that is expected to occur independently of Web navigation activity.

From the results it became clear that whereas in actual numbers interaction with widgets - such as page scrolling - were the most predominant category of activities, actually most time is spent on using information. The next most time-consuming activity was locating information, which includes a number of page scrolling activities. Locating 'interesting' or 'related' information was more time-consuming than looking for a specific string or image. 'Go to page' activities were particularly time-consuming due to the time users had to wait until the page was fully loaded; following a hyperlink was the most common method to go to a page, followed by the back button. The authors conclude with several design implications. First, in terms of time efficiency, it might be more worthwhile to spent effort on supporting readers to use information than on the design of interface widgets. Second, page design might need to be more targeted to 'scannability' than to 'readability' and the need for scroll activities might be lessened. Third, page load times need to be significantly improved. The latter implication might already been solved, as the speed of the Web has significantly increased since the time the study was carried out.

Whereas Byrne's study aimed to decompose general Web tasks into its constituent parts, Sellen, Murphy and Shaw (Sellen, Murphy & Shaw 2002) aimed to separate various kinds of activities carried out on the Web. Different kinds of activities might call for different kinds of support. Interestingly, whereas the title of the publication suggests a very specific participant pool of 'knowledge workers', the twenty-four participants had quite different backgrounds, including consultants, journalists, engineers and academics. All participants were interviewed about their Web activity during the past two days. They were asked to describe the various events that showed up in their browser's history list and two answer a series of questions for each activity. Six clusters of activities emerged from the data:

- *information gathering (35%):* gathering information in order to compare, choose, decide about or use for some purpose;
- *finding (24%):* find some specific, well-defined information on the Web;
- *leisure browsing (27%):* visiting personal or work-related sites with no specific goal in mind, but rather to be informed, stay up to date or be entertained;
- *transacting (5%):* using the Web to execute transactions, such as online banking or ordering a product;
- *housekeeping (5%):* check or maintain the accuracy and functionality of Web resources;
- *communicating (4%):* participate in chatrooms or discussion groups.

Most Web activities involved *information gathering*. Many of these activities were aimed to carry out a *set* of related tasks, which were often hard to articulate. Information gathering activities were also quite time-consuming and generated complex navigation patterns; users were often skimreading large amounts of information to assess its relevance. Besides search engines, users often consulted known, trusted resources. Trusted resources were also important in *finding* activities. Most finding activities were self-contained and relatively short; however, information finding often involved multiple keywords, following many links, and frequent scanning. *Browsing* activities were generally considered less important than the former categories, although several participants talked of the importance of keeping up with events at home or at work. A large amount of browsing activities were carried out on a daily basis. An important observation from the above results is that information gathering is more prominent on the Web than specific information finding, both in number of events and in time spent. Support for these activities should take the vagueness of the expressed information need into account and should provide evidence for the authority of sources.

Information finding and gathering activities were further investigated by Teevan et al. (2004). They conducted 151 semi-structured interviews with fifteen participants, all graduate students in computer science. From the results it became clear that, even for very specific information searches, keyword search was used to a surprisingly small extent - 42% of all specific information searches involved keyword search. In the majority of cases the participants followed a trail of links from a known starting point. Even for relocating information found before, this strategy was used. The authors called this strategy *orienteering*. Orienteering involves using both prior and contextual information to narrow in on the actual information target, often in a series of steps, without specifying the information need up front. The other extreme, directly jumping to the information target - for example by using keyword search - is called *teleporting*. Whereas at first sight orienteering might seem an ineffective strategy for locating information, the authors listed three significant benefits of orienteering over teleporting. First, orienteering lessens the *cognitive burden*. Instead of having to specifically articulating their information need, users can make use of cues provided by the interface, to gradually narrow down the search. Users associate information with an *information source*, such as a file, that needs to be located. Locating or relocating information by orienteering involves finding a starting point to navigate to the information source. The second benefit of orienteering is that users maintain *sense of location*. Users are generally bad at understanding how search engines work. By contrast, orienteering provides users with points visited before and with which they are already familiar. The third benefit is related to the second: while orienteering users gather information that can be used as *context* for *understanding the answer*. The context provides evidence that the information they found is the information that they were looking for, and provides evidence for judging the trustworthiness of the information found. We believe that the concept of orienteering is a very important result from this study; it shows that navigation in hypermedia does not just serve to arrive at a certain destination, as is often the case in physical navigation (Dillon & Vaughan 1997). By merely teleporting users would not be able to completely appreciate the information found, as it is part of a larger context. We believe that orienteering and keyword search are not activities that exclude one another. Orienteering from an initial query provides information that can be used for further specifying the information need and the associated query. From this perspective, orienteering can be seen as a means for transforming an initially vague information gathering task into more specific information finding tasks.

Capra & Pérez-Quiñones (2003) carried out a laboratory study to further investigate orienteering behavior while re-finding found things. Their participants consisted of six couples of graduate students. In the first session, one member of each couple - the *User* - was instructed to find information on five different tasks that were selected to provide a variety of directed and freeform information finding activities. During the information finding session, the participants were allowed to make as many or few annotations as they wished. In a second session, which was scheduled one week after the first session, the User and the other half of the couple - the *Retriever* were seated in two different rooms. The User was asked to provide the Retriever directions to relocate the information found before. This experimental procedure was meant to observe the step-by-step process followed in re-finding from the collaborative dialogs. The results confirmed that re-finding was mainly conducted as an iterative process to find specific information targets, with initially underspecified goal statements and additional information provided at the point where needed. The Users heavily relied on *waypoints* - a small number of important sites on the path to their goal, of which they remembered the address, the title or salient details. From the dialogs it was observed that it may not be the actual *path* that lead to the information which is remembered, but rather the *process* of how to reach the information.

3.3.2 Long-term studies

As mentioned earlier, click-through studies typically make use of unobtrusive data gathering techniques that gather low-level events such as key strokes and Web activity. Whereas the data merely reflects usage patterns and not qualitative information on why certain activities took place, in particular long-term studies provide interesting information on real life Web usage behavior. Several techniques for collecting the data are dealt with in the next chapter on Web usage mining. In this subsection we discuss the results of a number of influential long-term studies that are frequently cited in articles on user navigation support.

The first category of long-term studies focuses on *browsing behavior* and recurrent activities on the Web. These studies are actually the studies that qualify for the term *click-through* studies, as they make use of the complete log of user browsing activities. Surprisingly, the number of reported long-term click-through studies is quite low and the last study reports on data collected in 2000, more than five years ago at the moment of writing this thesis. This might be explained by the fact that it is hard to find a sufficiently large pool of participants that are willing to share their Web activities for a longer period. In chapter seven we elaborate on and extend these long-term studies, and report on the issues we experienced in the preparation phase of the study. The second category of long-term studies focuses on user *search behavior*. Most of these studies make use of server logs of major search engines. In a sense, this data might be considered incomplete, as no information is available on the navigation activities carried out on other sites between searches. Nevertheless, interesting results can be found in the reports.

Click-through studies

The first long-term click-through study was carried out in late 1994 by Catledge & Pitkow (1995). In this period the Web was relatively new; the typical user is a 31 year-old educated male, Mosaic the Web browser of choice and the number of Web sites about 10.000 (Hawkey & Kellar 2004). During a period of three weeks data the Web activity of 107 participants was collected using an instrumented Mosaic browser at university computers of the Georgia Institute of Technology. Following hyperlinks (51.9%) and the browser's back button (40.6%) were by far the most common mechanisms for accessing a Web page. The list of most popular sites is dominated by research institutes, including CERN, the institute from which the World Wide Web was initiated by Tim-Berners Lee. As a comparison, more recent lists of most visited sites are dominated by search engines, entertainment and commercial Web sites (comScore Media Metrix 2004), which indicates that the findings of this study might not be valid for today's Web browsing anymore. The average length of a site visit was 12 pages, and users tended to operate in one small area of these sites; the navigation structure of these site visits resembles a hub and spoke structure due to the frequent use of backtracking. Based on these results the authors conclude that 'must see' information should be no further than two clicks away from the site's home page.

Between the first and the second study the Web continued to grow, both in size and in usage. Early adopters got home internet connections via commercial providers (Hawkey & Kellar 2004). Tauscher & Greenberg (1997), who conducted their study on Web usage behavior in late 1995, recognized the fact that Web users often carried out recurrent tasks on the Web. Their study aimed to characterize people's revisitation patterns and to evaluate the history systems employed by Web browsers, which were quite similar to the history systems of current browsers. 28 unpaid volunteers, mainly academics and engineers with at least one year of internet experience, participated in the study. The participants made use of an instrumented Mosaic browser for a period of 5-6 weeks. The results confirmed Catledge and Pitkow's data that link following and back button usage were the most frequently used methods for accessing a Web page; bookmarks and the temporally ordered history list were rarely used. They used the *recurrence rate* to calculate the probability that any page visit is a repeat of a previous visit, expressed as a percentage. An average overall recurrence rate of 58% was found for their participants; reanalysis of the data from the Catledge and Pitkow study yielded a recurrence rate or 61%. Based on these results, the authors concluded that Web browsing activities are a *recurrent system*, 'where users predominately repeat activities they had invoked before, while still adding new actions to the repertoire'. In a post-study questionnaire the most common reasons for revisiting a page were collected:

- The information contained by them changes;
- They wish to explore the page further;
- The page has a special purpose - e.g. search engine, home page, index page;
- They are authoring a page;
- The page is on a path to another revisited page.

The authors made some further characterizations of page revisits. It was found that the ratio between the relation between the number of page requests and the number of unique pages visited thus far is roughly linear; the *url vocabulary* grows linear with the number of page requests. Changes in the slope indicate periods in which users are mainly revisiting pages or looking for new information. Two important characteristics of revisited pages were described: first, most page revisits were visits to pages very *recently*; the probability for a page to be revisited decreases steeply with the number of page visits since the last visit - see figure 3.2. Second, there is a small number of highly *popular* pages that are visited very frequently; the probability for a page to be revisited decreases steeply with its popularity ranking. The most frequently accessed pages were the user's personal or organization pages, the browser's home pages, search engines and pages that the users were authoring themselves. The authors further hypothesized that *locality patterns* could be found in the data. Locality considers recurrences in terms of periods of time where repeated references are made solely to a small and related group of items, called a locality set. However, it turned out that the locality sets found were small both in terms of pages covered and length of periods; moreover, few locality sets were repeated, which makes the concept not very useful for history support. Similarly, longest repeated sequences of page visits tended to be short and infrequent. Based on the findings, the theoretical support for revisits was calculated for the sets of most frequently and most recently visited pages. A list of the ten most recently visited pages would cover 74% of all page revisits, and 81% if duplicates are removed. The stack-based back button performed similar or slightly better on short distances, but performance got much worse for longer lists. The list of ten most popular pages would cover 46% of all page revisits, significantly worse than the recency-ordered list. While the characterizations of

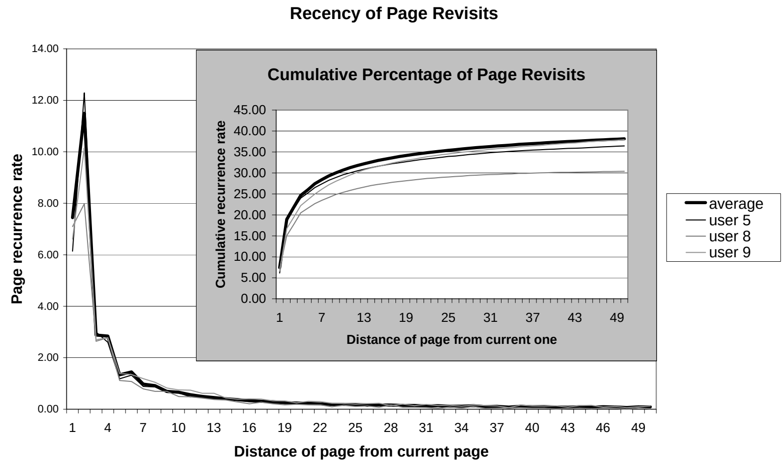

Figure 3.2: *Percentage of page revisits as a function of distance from the current page. The popularity distribution shows a similar pattern.*

page revisits and the implications for history support as presented by Tauscher and Greenberg were an important progress, we think the presented results only sketch a far from complete picture of recurrent behavior in Web browsing. In chapter seven we describe the above results in more detail and separate different categories of revisits: revisits that take place in one navigation session and revisits to pages visited in earlier sessions. Further, we will look into distributions of page revisits over time and investigate which browser tools are used for initiating page revisits.

The third long-term click-through study was carried out by Cockburn & McKenzie (2001) late 1999. The authors believed that an update on the earlier studies was required, as the Web has grown wildly in size and usage, new browsers such as Netscape and Internet Explorer were dominating the market, and the usage of search engines has increased dramatically. Moreover, the earlier studies covered a relatively small period and participants were not using their browser of preference. The 17 participants in this study were mainly computer staff and students. They were asked permission to retrieve and analyze the backups of their browser's history and bookmark files from the previous four months. By asking permission after the logging period, the 'Hawthorne Effect' was avoided - modifications to participant behavior due to their awareness that their actions were being logged. Several normalizations were made in the data analysis program, such as the truncation of the *query part* of the Web address. Whereas the authors stressed that this particular step did not distort their results, in chapter seven we will see that it had an high impact on the recurrence rate. Cockburn and McKenzie found a recurrence rate of 81% and concluded that recurrent activities have become even more prominent in Web usage; we believe that this effect is caused by differences in data preprocessing compared to the earlier studies rather than by a dramatic change in Web usage. The recency and popularity distributions of page revisits found by Tauscher and Greenberg were confirmed by this study. An important result from the Cockburn and McKenzie study is the observation that browsing is a *rapidly interactive* activity; the most frequently occurring time gap between subsequent page visits was approximately one second and gaps of more than ten second are relatively rare. This implies that *response time* is an important issue for Web navigation; in addition, pages should be kept short and scannable. Analysis from the bookmark files revealed dramatic differences between users; one participant did not use them at all and another participant had 587 bookmarks. On average, users had 184 bookmarked pages, stored in 18 different folders. In the course of time, far more bookmarks were added than deleted, which implies that users have or will have problems managing the size and organization of their bookmark collections. As a final result, we mention that there was almost no overlap between the pages visited by different users, even though they worked in the same institution: only 9.2% of all pages were visited by more than one participant and no page had been visited by all participants.

Web search studies

The click-through studies presented thus far were primarily focused on navigation behavior on the Web. However, none of the studies report on search activities, whereas in 2004 search engines such as Google and Yahoo rank in the top 10 of most visited Web sites (comScore Media Metrix 2004). Even though search engines were far less common in the mid-nineties, they were commonly used during the time of the Cockburn and McKenzie study. Although no integrated click-through data on search and navigation behavior is reported in the literature, several studies on Web search activities have been conducted. These studies typically make use of Web logs of major search engines. Below several characterizations of Web search behavior are presented and discussed.

Broder (2002) treats search as a classical information retrieval task, which involves translating an information need to a query posed to a search engine. Query refinements might be needed to refine the results. He distinguished three classes of Web queries, related to the finding, gathering and transaction Web activities identified by (Sellen et al. 2002) and discussed earlier in this chapter:

- *Navigational queries:* The immediate intent is to reach a particular site. A typical navigational query is ‘Greyhound bus’, which is presumably meant to reach the site www.greyhound.com. Navigational queries are also called ‘known item’ search in classical information retrieval;

- *Informational queries:* The intent is to acquire some information assumed to be present on one or more Web pages; information queries are closest to classic information retrieval tasks, except that many informational queries on the Web are extremely wide, such as 'cars' or 'San Francisco';

- *Transactional queries:* The intent is to perform some Web-mediated activity. This type of queries is hard to recognize and needs to be manually tagged.

The prevalence of these query types was determined both with a user survey and analysis of query logs from the AltaVista search engine. Randomly selected users were presented a pop-up window asking them to indicate the specific goal of their search activity. 3190 responses were received, a response rate of about 10%. In addition, a random set of 400 English-language queries, with adult content removed, was analyzed. Results indicate that informational search is the most common search activity, accounting for about 50% of all queries; transactional queries accounted for about 30% of all queries; 20% of all queries involved navigational queries. The authors indicate that the results are 'very soft', as inferring the user intent from the query is 'at best inexact science'.

Lau & Horvitz (1999) analyzed 4690 queries from the Excite search engine, collected from the server log over a twenty-four hour time period. The goal of their study was to analyze and model the process of Web query refinement. Search actions were grouped into a set of mutually exclusive refinement classes:

- *New:* a query for a topic not previously searched for by the user within the scope of the dataset;
- *Generalization:* A query on the same topic as the previous query, but seeking more general information than the previous query; one or more keywords have been deleted from the query;
- *Specialization:* A query on the same topic as the previous query, but seeking more specific information than the previous query; one or more keywords have been added to the query;
- *Reformulation:* A query on the same topic that can be viewed as neither a generalization nor a specialization;
- *Interruption:* A query on a topic searched on earlier by a user; one or more keywords appeared in earlier queries, but not in the last query;
- *Request for additional results*

From the results it became clear that new queries and requests for additional results were by far the most common interactions with the search engine. Relatively few queries were generalized, specialized or reformulated. However, during the first twenty minutes the probability of a query being modified is higher than the probability of the user issuing a new query. Specialization and reformulation are more common activities than generalizations. We think that this relates to the *orienteering* approach of gradually refining an underspecified information need. The average query length was 2.30 words, which also reflects the users tendency of not fully specifying their goal. Due to their nature, query specializations were considerably longer - 3.40 words on average.

Spink, Wolfram, Jansen & Saracevic (2001) analyzed over one million queries by users of the Excite search engine. Like in most search engines, Excite searches are based on the exact terms in a query. The use of logical operators, such as AND and OR, is allowed but not required. The study confirmed the low number of query terms reported by Lau and Horvitz. The average number of queries per user session is low, with a peak at one and a long tail of few users submitting more than one query. This confirms that query modifications are rare, but still frequent enough to consider: 52% of all users entered more than one unique query. Of the modified queries, the majority involved query reformulations and specializations. Users typically do not add or delete many terms in their subsequent queries. Requests for additional results followed the same pattern as query modifications: the majority of 28.6% of users examined only one result page, another 19% looked at only two result pages. The median was 8 result pages viewed per user. Logical operators were hardly used; the AND operator was the most common, covering 3% of all queries. There was a small number of very frequently used query terms and a long tail of terms used less than three times. The set of most popular query terms mainly involve entertainment, download and adult content. Classification of a random sample of queries yielded the following top four of most popular categories: entertainment, adult content, e-commerce, and computers. Interestingly, the evolved interests do not completely match the distribution of information on the Web: about 83% of Web servers contained commercial content and only about 1% adult content.

3.4 Theoretical models

In the previous section we have seen a large number of studies aimed to provide knowledge on various aspects of user Web navigation. This kind of data is needed to guide future design and evaluation of user interfaces. However, the data is far from integrated and often it is not clear how results can be generalized to other situations. This means that usability engineers need to strongly rely on their intuition and experience, and possibly also on additional user tests to verify whether results still hold in this particular situation (Byrne 2003). A theoretical framework that relates the different findings and that models the underlying cognitive processes does provide usability engineers with quantitative background knowledge that can be used to derive quantitative predictions. In this section we discuss two theoretical models, the theory on *information foraging* (Pirolli & Card 1999) and the *CoLiDeS* model (Kitajima, Blackmon & Polson 2000). We continue with a discussion on a specific type of theoretical models, commonly called *cognitive frameworks*. Cognitive frameworks are working computer programs that simulate user interactions with an interface, making use of strategies that are assumed to reflect actual user strategies. From a behavioristic point of view, these frameworks can be seen as an 'explanation' of the underlying actions. A more practical benefit is that the simulations can be used for a variety of purposes; user tests can be replaced or complemented by a potentially large number of simulations, and user-adaptive systems can predict the most likely next user actions.

3.4.1 Information foraging theory

In 1999, Pirolli & 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.

Information foraging theory is an example of an *adaptationist psychology* theory, which draws upon evolutionary-ecological theories from the field of biology. Users who are trying to gather information are compared with food foragers, such as predators hunting for prey. Whereas the prey will provide the predator with energy once it is caught, the process of catching the prey costs energy. The predator will try to maximize the amount and quality of prey and to minimize the costs to catch the prey. Strategies for maximizing the *rate of gain* include moving to another hunting ground, once the current hunting ground is sufficiently depleted, and leaving animals that are hard to catch or that do not have too much meat on their bones alone. *Information foragers* are believed to follow similar strategies. Information relevant to a person's information needs may reside in some books in the library, or in (sections of) certain websites. Information is typically organized in a 'patchy' structure, with several clusters of relevant information - these clusters are called *information patches*. Within a current patch, users can decide to remain within the patch or to leave for another patch. This decision is based on personal preferences and the current *information scent*: 'the (imperfect) perception of the value, cost or access path of information sources obtained from *proximal cues*, such as bibliographic citations and Web links'. Once the information scent drops below a certain threshold, leaving the patch is likely to have a positive impact on the rate of gain. Formally, the rate of gain R can be defined as 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,

$$R = \frac{G}{T_B + T_W} \tag{3.1}$$

As can be observed from the above formula, both time spent within patches and time spent between patches have an equal impact on the rate of gain, which means that costs for both activities should be minimized to increase the rate of gain. However, thus far no assumption has been made on the *action costs*. In real life, patch leaving activities are quite expensive due to the physical distance that needs to be crossed. To a certain extent, this also yields for information on the internet: it costs less time to click on a hyperlink than to find an alternative resource, to become familiar with its structure and layout, or even to type in an url. However, during the past few years moving between patches has become 'cheaper' with the increased availability and usage of search engines; it has been observed that 'Google makes users leave your Web site faster' (Nielsen 2003). Within-patch navigation strongly depends on the information scent; results from the Katz and Byrne laboratory study (Katz & Byrne 2003) suggest that the site organization and menu structure have a strong impact on the information scent.

From the above it becomes clear that the *information environment* has a high impact on the users' foraging strategies. However, the information scent provided by the environment is far from deterministic; foraging through a large volume of information involves uncertainties about the location, quality, relevance and veracity of the information sought and the effects of foraging actions (Pirolli 2004). Moreover, proximal cues, such as hyperlink labels, are often ambiguous and do not provide information on the location it points to. This implies that in order to estimate the expected benefits of a possible next action, users need to learn about and to take into account the specific unreliabilities related to the information patch. An additional source of uncertainty may be the *task environment*. As mentioned several times before, information gathering tasks - let alone browsing tasks - are often not fully specified; the information needs associated with the task are gradually recognized during the process of *orienteering*.

Scent-following activities serve to exploit the information domain. However, users also often optimize the information environment to improve future information gains. These activities are generally called *enrichment activities*. Enrichment activities aim to decrease information finding costs by organizing or filtering the available items. However, there is an obvious trade-off to be faced: at some point the efforts spent on enrichment might not be sufficiently compensated by the resulting increase in efficiency. Several information enrichment actions may be thought of, targeted at either within-patch or between-patch activities. As mentioned earlier, enrichment of between-patch activities will decrease costs of moving from one information patch to another, which will most likely lead to a decrease in the within-patch time - the threshold for moving to another patch has become lower.

Information filtering - or *diet selection* - is an important category of enrichment activities. By iterative query refinement activities the candidate information patches most likely will be limited in number and of high quality; between-patch navigation actions will be cheaper for smaller, more relevant selections. A higher threshold for the minimum profitability of an information patch or an individual information source, will most likely increase the information gain, but does also increase the risk of lost opportunities. In fact, information foraging theory suggests that the decision between continuing exploring a patch or finding an alternative patch is bound to be an educated guess: even if you think that it is unlikely that you will find what you are looking for in the current information patch, it still might be just one click away. One thing is for sure: if you leave the patch, you will never know.

One of the assumptions behind information foraging theory is that humans exhibit *bounded rationality* or make choices based on *satisficing*. In that respect, information foraging may be regarded as *heuristic search* using local optimization techniques. Given this perspective, backtracking activities are implicitly modeled as a failure; if a local search does not end up with a sufficiently high local maximum, one backtracks to a point from which hopefully a path leads to a better local maximum. However, information foraging theory does not only involve blind scent-following activities, but it also takes information enrichment activities into account - activities that do not directly add to the information gained, but that will lead to a higher rate of gain in the long run. We think that the process of orienteering, in particular those activities that lead to an increased sense of context and a better knowledge on where to find navigational hubs, should be explicitly modeled as an enrichment activity. Furthermore, the profitability of an information source should not only be measured in terms of content, but also in terms of structure.

Information foraging is a theory rather than a cognitive framework; it does provide a framework for relating and understanding various empirical observations, but it is far from an integrative model of Web navigation or a computational model. However, a number of computational models - cognitive frameworks - have been built, based on notions from information foraging. We think this is the main benefit of the information foraging theory: it provides a way of looking at the issues related to information gathering from several perspectives; the insights may be used for building actual computational models, for rethinking design implications or for interpreting observations. In the following two subsections we discuss the CoLiDeS model and *cognitive architectures* that provide (semi-)algorithmic implementations of notions from the information foraging theory.

3.4.2 The CoLiDeS model

The information foraging theory provides a framework for reasoning about user navigation sequences in information gathering activities. In particular, it explains when and why users might leave a current information patch and look for another patch. Furthermore, it explains why users involve in activities that are not directly related to the information goal but that might increase navigational efficiency. What it does not explain, is the within-page user activities of visually scanning links on a page and the process of assessing links with respect to the user's navigational goal.

An alternative model, the *Co*mprehension-based *Li*nked model of *De*liberate *S*earch (CoLiDeS) (Kitajima et al. 2000) does take the actions of link scanning and assessing into account. The CoLiDeS model is based on earlier theory and models on text comprehension. The most important claim of the model is that comprehension of texts and images is the core process underlying Web navigation. The four cognitive processes most central in the CoLiDeS model are *parsing* the screen layout, *focusing* on a region of the screen, *comprehending* the available options, and *selecting* the most promising option. Web users are assumed to expect that conventionally designed Web sites have a number of standard screen elements, such as menus and navigation bars. The top-down process of distinguishing the constituent regions is accompanied by the bottom-up process of relating visually distinctive text elements - such as headers - to these regions. Focus is given on the region containing the most relevant or most promising elements. Subsequently, the user examines the region more detail and tries to comprehend its elements in terms of relevance to the user goal. Three independent factors are believed to define the degree of relatedness:

- the degree of similarity between the (interpreted) screen element and the user goal; the similarity is determined using a technique called *latent semantic analysis* (LSA) in combination with a proper corpus ('semantic space');
- the frequency with which the user has encountered the element before on the navigation path; screen elements that a user is familiar with - for example a tab menu or a search form - are more likely to be followed;
- whether the screen element has a literal matching with the user goal.

According to the CoLiDeS model, in the ideal case, a user only employs forward navigation actions, in which the user moves smoothly step by step toward the user goal. However, it may be the case that at some point no action matches a desired level of similarity, frequency or literacy - this situation is called an *impasse*. In order to recover from an impasse, users may focus on another screen region, try and use an alternative strategy - for example searching instead of browsing -, or backtrack to a previously visited Web page. Impasses are assumed to be caused by link labels that are either ambiguous, too general, too technical or simply inappropriate.

CoLiDeS shares with information foraging the concept of local search - or hill-climbing. In contrast to information foraging the CoLiDeS model explicitly chooses a technique for measuring similarity. In addition, the model has successfully been used in several *cognitive walkthroughs for the Web* (Blackmon, Polson, Kitajima & Lewis 2002), in which goal-driven user navigation is simulated through a number of Web sites. Results from these walkthroughs can be used to identify and 'repair' problematic link structures within a Web site. Whereas CoLiDeS has proven its use in this respect, it is far from a complete or reliable simulation of real Web users. The most important issue that we identified with CoLiDeS is the inflexible concept of 'impasses' and the resulting implicit definition of backtracking as 'recovering from a failure'. Further, the choice for LSA as a similarity indicator might simulate human assessment of Web links, but it makes a rather bad assessor; (Juvina & Van Oostendorp Submitted) reported an unrealistically high number of impasses caused by common problems associated with LSA, such as unknown words and undeterministic effects due to factoring techniques.

3.4.3 Cognitive architectures

Whereas the CoLiDeS model was not explicitly introduced as a simulation of a real Web user, its procedural approach provides an usability engineer tools to evaluate a Web site architecture in a relatively objective manner. It is believed that a next step toward understanding user behavior on the Web is the development of *cognitive frameworks*, also called *cognitive architectures*. Cognitive architectures aim to simulate human intelligence in a humanlike way (Byrne 2003), which includes attention, memory, problem solving decision making, learning and so on.

As cognitive architectures are intended to be running systems, they are expected to provide an alternative for usability tests and explicit representations of notions from 'vague' notions from cognitive psychology such as 'working memory' or 'mental model'; this provides a window into how these theories actually work and how they relate to one another. Cognitive architectures are mainly based on theoretical insights from psychology or related fields such as evolutionary-ecological theories, and built on experimental data. These aspects make cognitive architectures quite similar to the concept of *user models* as introduced in the previous chapter. However, the ambition appears to be a little higher: according to (Byrne 2003) cognitive architectures could serve as the usability equivalent of computer simulations as used by, for example, aerospace engineers. Whereas computer simulations of aerodynamics are built on accepted knowledge, they are not infallible, which implies that evaluation in real wind tunnels might still be needed for unknown situations.

In contrast to 'hard' engineering domains, cognitive architectures of Web navigation are far from comprehensive (Miller 2005). Given the diversity of experimental data and theoretical insights presented in this chapter, this is hardly surprising. With more knowledge on how all these various aspects interrelate, more complete models might be constructed, but Web navigation is non-deterministic by its very nature: it is carried out by users engaging in tasks that are in many cases not clearly specified and that become more specific while carrying out the task. Whereas insights on airflow in controlled laboratory conditions are still valid in more noisy real-life conditions, users are likely to behave different in laboratory conditions. As a result, a computer simulated 'wind tunnel' of user behavior will be a rough approximation at best.

The SNIF-ACT model of information foraging on the Web (Pirolli & Fu 2003) is based on insights from information foraging theory and the ACT-R theory and simulation environment (Anderson & Lebiere 2000). The two major components of the SNIF-ACT model are *declarative knowledge* and *procedural knowledge*. Declarative knowledge corresponds to things that we are aware we know and that we can describe to others, such as the content of Web pages and the working of Web links. Procedural knowledge specificies how declarative knowledge is transformed into active behavior - like pointing a mouse to a menu item or making use of a search engine. Procedural knowledge is represented as *production rules*, similar to the cognitive processes in the CoLiDeS model. If there is more than one possible action, the item which most closely resembles the user goal description - the *information scent*, described in terms of cosine similarity - is chosen. The SNIF-ACT model is evaluated with four users, each carrying out two different tasks. Results confirm the information foraging assumptions that links with a higher information scent are more likely to be followed and that users leave an information patch - a Web site - when the information scent drops below a threshold.

The SNIF-ACT model, as presented in (Pirolli & Fu 2003) does not provide a taxonomy of procedural knowledge for navigation activities. This might be explained by the fact that all individual users have their own strategies for reaching their goals and their own triggers for a particular action. In order to fill the model with rules such as 'if the information scent on this particular page is too low, and if the last used query can be specified, return to the search engine and add a new keyword to the query', empirical data and taxonomies as discussed in the previous sections are needed.

Some other flavors of cognitive models have been implemented, such as (Miller 2004) (Juvina & Van Oostendorp Submitted). It is beyond the scope of this chapter to discuss these in detail. A general characterization is that the models aim to simulate users' goal-directed information finding activities. The approach is quite similar to the approach followed by the CoLiDeS and SNIF-ACT approaches, with a focus on abstraction and simplification of the process. The main drawback of these models is the lack of empirical evaluation (Juvina & Van Oostendorp Submitted) or evaluation in oversimplified, unrealistic settings (Miller 2004), which hinders in making assumptions on the validity of these cognitive architectures in real-life situations.

3.5 Putting the parts together

In the previous sections a wide spectrum on empirical observations and theoretical insights has been discussed. We deliberately chose to summarize and discuss the earlier research efforts separately and in a systematic manner, in order to give an overview on the various focuses of attention and methods of research used. With all the 'bits and pieces' from the earlier sections in mind, it is about time to connect the pieces. In the first part of this section we revisit the navigation activities of searching, browsing and backtracking. We continue with a discussion on recurrent behavior on the Web and browser support for various kinds of revisits. In the third part of this section we propose a framework for look-ahead navigation support, based on insights from information foraging theory and data from Web usage logs. We conclude with an outlook on the upcoming chapters.

3.5.1 Searching, browsing and backtracking

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 a useful navigation strategy 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.

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 the CoLiDeS model (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 2005). 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 productive navigation strategy, more effective than linear forward navigation. The observational study by Teevan et al. (2004) provides a qualitative description of *orienteering behavior*, a concept that recognizes the value of backtracked navigation trails that do not directly relate to an immediate goal, but that will provide context and sense of location in subsequent navigation. What effective backtracking strategies look like and how this can be integrated in the information foraging theory, is not yet clear.

To shed a new light on the role of backtracking activities in information gathering, we conducted a laboratory study, which is described in chapter 6 and in (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.

In terms of information foraging, browsing and backtracking can be regarded as *within-patch* navigation. Search activities aim to generate a list of pointers to various navigation patches; returning to the list of search results and selecting an alternative result from the list can be seen as *leaving* for another patch. The Web search study by Lau & Horvitz (1999) shows that users often change patches by returning to a search result. Whereas issuing a new query results in a first selection of information patches, query modifications *enrich* the list of candidate starting points by better accommodating it to the user information need; often, the information need is rather general or unspecified and will change as the user proceeds. Analysis of search engine logs (Spink et al. 2001) shows that query modifications are quite rare; however, as the analysis did not take the user navigation on other sites into account, it is not known whether this bias implies that users rarely modify their queries while gathering information; it might be more likely that the high amount of sessions with just one query relate to short sessions in which users visit routine sites or are browsing for leisure.

To summarize, searching, browsing and backtracking are three closely related and complementary ways of navigating the Web. While navigating, users have to actively choose between these activities, making use of incomplete proximal cues that are different for each direction. 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. Theoretical models of Web navigation show the same bias. The laboratory study by Katz and Byrne (Katz & Byrne 2003) shows that search is preferred in situations where the information scent is low. To what extent users search rather than follow links, is yet unknown.

3.5.2 Recurrent behavior and revisits

The main focus of the long-term click-through studies is on recurrent behavior on the Web. From the results it became clear that users often revisit Web pages; the reported percentages of page revisits vary between 58% and 81%. Most revisits are focused on a limited set of highly popular pages and the pages that are visited very recently. There appears to be a stable ratio between the number of newly visited pages and revisited pages.

Unfortunately, the studies only report on the *amount* and the *distributions* of the page revisits. No data is available on *what* sort of pages are popular and *why* and *how* they are revisited. Whereas Tauscher and Greenberg (Tauscher &

Greenberg 1997) explored the theoretical performance of the lists of most popular and most recently visited pages, it is not known whether these lists will actually solve actual user problems in relocating pages visited before. Based on the list of current most popular sites (comScore Media Metrix 2004), which is dominated by search engines, entertainment and news, and e-commerce sites, it is most likely that the lists of a users' most frequently visited pages are populated by these kinds of portal sites, which are revisited on a routine, daily basis to initiate a search, to access a service, or to check the news. Moreover, as users seem to visit these pages on a very frequent basis, it is most likely that the users know the address by heart; most browsers will automatically complete partially entered addresses while typing, which means that even long or complex addresses of popular pages can be retrieved after having entered only a limited number of characters.

By contrast, the long tail of less frequently visited pages is likely to be more problematic; the more infrequent the page is accessed, the less likely it is that a user will remember the exact address - and to make things worse, the automatic address completion will not work if the page is not available anymore in the browser's history. A list of bookmarks, managed by the user, would support these infrequent visits. Unfortunately, these lists should be considerably large to cover a reasonable percentage of these less frequently revisited pages, which would require the user to spend time and effort on organizing and categorizing their bookmark lists - an effort that users are reported not to make (Cockburn & McKenzie 2001). In order to design support for recurrent activities, more insight is needed on which pages are visited very frequently and infrequently, and the elapsed time between the less frequently visited pages; are these pages visited, say, about once a week - or could one identify chunks of increased revisits to a particular page, separated by long periods in which the page is not revisited at all. Further, if browser history mechanisms do not support these less frequent revisits, what alternative strategies are used to return to the page.

From the studies it also does not become clear to what extent the dominance of revisits to pages visited very recently before is caused by backtracking activities; the reported recurrence rate of 81% by Cockburn and McKenzie (Cockburn & McKenzie 2001) does suggest that users are likely to have pages that they visit more or less continuously.

3.5.3 Sniffing around for enhanced information scent

Given the theory of information foraging and insights on orienteering behavior, it would be an appealing idea to put this knowledge into use. Below we propose a framework that extends an existing prototypic system, ScentTrails (Olston & Chi 2003). The basic idea is that while a user engages in information gathering activities, a 'scout' looks a number of steps ahead. Instead of relying on the limited proximal cues of hyperlinks, it assesses the relevance of pages based on their contents; the relevance is propagated to the current page in order to provide the user with navigation suggestions based on an 'enhanced' 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. 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.

Information foraging theory predicts that the action with the highest information scent will be chosen. However, as the proximal cues from the different categories are different, the choice between browsing, searching and backtracking will at least partially be guided by personal preferences for each of these actions. Moreover, as proximal cues such as link anchors are typically a poor representation of the destination page's contents, the cues might be misleading. The ScentTrails system (Olston & Chi 2003) aims to improve the user's limited information scent by annotation of browsing cues - hyperlinks - with an enhanced navigation scent, determined by the following procedure.

- Each information gathering is assumed to start with a query; the keywords from the last issued query are assumed to represent the user's *partial information goal*;
- The relevance of each page in an information patch is calculated using a conventional information information retrieval technique: cosine similarity with tf.idf weighting (Grossman & Frieder 1998);
- *Spreading activation* is used to flow the page relevances backward along links; each page's information scent is recursively defined as the sum of its own relevance and the relevance values of its children, controlled by a decay parameter;
- All links within a Web page are highlighted according to their information scent.

The approach was evaluated in a laboratory study with twelve participants, working on eight different tasks, making use of a moderately-sized prototypical

Web site. The average task completion time was significantly lower for participants using the ScentTrails system than for participants who either searched or browsed. A computationally less intensive approach, in which relevance is only propagated to the pages that directly link to the page, performed slightly worse than ScentTrails, but was still an improvement over merely browsing or searching. Participants indicated that the ScentTrails system permitted them to narrow down some aspects of the task with keywords. However, the link highlighting mechanism used - increasing the font size of textual hyperlinks - was criticized, as it gave Web pages a 'messy' appearance.

The ScentTrails system is an example of a cognitive architecture that actually serves as a user model for Web personalization. In contrast to 'real' cognitive architectures it goes a step further: as the user navigates, the system looks ahead and 'wafts' the estimated relevance of surrounding pages toward the user, providing an enhanced information scent as would have been perceived if users would have read the pages themselves. The idea of having a 'scout' that looks ahead and estimates the information scent of links appears promising; instead of taking over the navigation process from the user, it supports the user in orienteering behavior. However, as information gathering does not only involve forward navigation, we think that the assistance can be improved by including suggestions for search and backtracking actions.

In (Herder 2004) we proposed an alternative framework that does not only provide the user with scent-based browsing assistance, but that provides the user with navigation suggestions in all categories: forward browsing, search and backtracking. In addition to the ScentTrails' looking forward for estimating the relevance of yet unvisited pages, it also looks back by taking the information scent of backtracking actions into account. Further, in order to support a user's individual navigation style, it would be useful to take the users' preferences for browsing, searching or backtracking into account in the form of *action costs*; these action costs can be estimated by the prevalence of these actions in the users' navigation history, and by setting the threshold for patch leaving activities such as search, based on predictions similar to the SNIF-ACT system (Pirolli & Fu 2003). By taking these preferences into account, a user's natural navigation strategy can be supported.

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 ScentTrails system 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, Obendorf & Lamersdorf 2001). Suggestions for backtracking or search actions 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.

3.5.4 Conclusion and outlook

In this chapter an overview is given on empirical and theoretical insights on user navigation on the Web. Information finding and gathering are considered the most important activities on the Web. At the same time, a large number of Web activities involve recurrent behavior.

Laboratory studies and observational studies were mainly focused on information gathering. From the results it became apparent that information gathering involves search, forward browsing and backtracking activities; the process of gradually moving toward the information goal has been termed *orienteering*. Theoretical models of data gathering activities have a strong bias on forward navigation. From section 1.3.1 it became clear that usability problems experienced while gathering information generally lead to a decrease in performance. Based on the empirical findings, we believe that this decrease in performance cannot be directly translated into 'a higher amount of backtracking'. In chapter 5 we describe two laboratory studies that aim to identify successful navigation strategies, and to evaluate an adaptive hypermedia technique to support these strategies.

Several characterizations of recurrent behavior on the Web have been made in this chapter; most Web activities involve revisits to pages, most revisited pages have been visited very recently or belong to a small set of popular pages, and current browser history support - in particular the temporally ordered history list and bookmarks - are often not used. Instead, users often rely on known *waypoints* that lead them to the desired information. However, from these characterizations it is not clear to what extent pages are revisited as a result of backtracking activities or as a result of recurrent behavior. In chapter 6 we present results from a long-term study in which we have analyzed Web activities of 25 users for an average period of three months. In addition to the general statistics on revisits, we have explored individual differences. Further, by looking at revisit behavior in individual sessions, we were able to separate backtracking from recurrent behavior. We also took a closer and more qualitative look at the nature of frequently visited pages and temporal aspects of less frequently visited pages.

From the results of the long-term study it becomes clear that several aspects of Web usage have dramatically changed since the latest reported study by Cockburn and McKenzie (Cockburn & McKenzie 2001). Whereas thus far Web browsing has been considered an activity that mainly takes place in a single window, currently users often employ multiple windows or browser tabs simultaneously. As an example, by opening a new tab for initiating a trail from a search result page, users do not need to backtrack anymore; moreover, several results can be compared next to one another. Conducting several related or unrelated activities in parallel has various implications on browser history support and support for backtracking. Further, the prevalence of transactional Web sites has added a new dimension to the hypermedia concept of the Web: when providing information to obtain a product or a service, backtracking to an earlier page is often an *undo* activity rather than returning to an earlier visited navigational hub.

Whereas search is reported to be an important aspect of information gathering activities, search activities have mainly been investigated on themselves. In chapter 6 we investigate the prevalence of search activities in user navigation. Results indicate that some users rely on frequent query refinement by rapidly adding new keywords, whereas other users rely on longer initial queries and more extensive orienteering behavior.

The next chapter is a preliminary to chapter 5 en 6. In order to obtain Web usage data to analyze, user activities need to be logged. In chapter 5 we explain several issues in data collection and preparation, followed by a discussion on several techniques for analyzing and mining data on user navigation on the Web.

Chapter 4

Web usage mining - finding patterns in Web navigation

4.1 Introduction

Web personalization is a specialization area of adaptive hypermedia, specifically aimed at improving the usability of Web sites (Eirinaki & Vazargiannis 2003). Adaptive hypermedia systems build a model of the user and use this model throughout the interaction with the user (Brusilovsky 2001). The term 'Web personalization' is also often used in e-commerce, with a slightly different meaning and serving slightly different goals. In the field of e-commerce, Web personalization usually means the segmentation of user groups in order to create and aim effective messages at customers (Rozanski, Bollman & Lipman 2001). Though the goals are different, the methods employed by both researchers and marketers are quite similar.

Similar to the layers of the adaptation process, the process of Web personalization can be split into three stages:

- data collection and preprocessing;
- knowledge inference;
- personalization

In this chapter we concentrate on the first two stages, which together are usually labeled *Web mining*. Web mining is an area of data mining that deals with the extraction of interesting knowledge from the Web and the way the Web is used (Facca & Lanzi 2005). This knowledge may be used as a starting point for usability analysis, or as a reference frame for the design of algorithms and user support; in addition, knowledge on the specific usage patterns of one user or a user group may directly be used for Web personalization purposes, such as the selection of news articles to recommend. In the literature, four general categories of data sources for Web usage mining are distinguished, of which the first three categories can be gathered implicitly (Srivastava, Cooley, Deshpande & Tan 2000):

- *content*: the real data in Web pages, as presented to the end user. The data can be simple text, images or structured data, such as information retrieved from databases. Applications of content analysis include finding Web pages that contain similar content and determining which Web pages match a user query or user preferences;

- *structure*: data that describes the organization of the content. The data can be either data entities used within a Web page, or data entities used to connect pages - most notably hyperlinks. Within-page entities - such as the number of words, images or hyperlinks on a page - reflect the function of a Web page (does it mainly provide content or is it used for navigation). The structure, as defined by connecting entities - such as hyperlinks - reflects the way the information is structured, which determines to a large extent user navigation patterns;

- *usage*: data that describes the pattern of usage of Web pages, such as IP addresses, access date and time, referring pages and other attributes that can be retrieved; from the raw Web usage data user groups may be distinguished, as well as groups of pages that are often visited together, popular pages and frequently followed paths;

- *user profile*: explicitly gathered data that provides demographic information about users of the Web site. This includes registration data and customer profile information. Such information can be obtained through registration forms or questionnaires or inferred from user actions. User profile data may help in explaining peculiarities in the Web usage data, but it can also be used as an additional source of information for personalization.

Depending on the main purpose and data sources used, Web mining can be categorized as Web content mining, Web structure mining and Web usage mining. Applications of *Web content mining* include categorization of pages based on their content, identification of topical clusters, and support for information finding and information filtering. Web content mining mainly makes use of information retrieval (IR) techniques. *Web structure mining* techniques aim to characterize Web pages, Web links, site structures and the structure of the Web in general. Applications include the design of efficient search and navigation algorithms. *Web usage mining* - which is the category of Web mining that we will focus on in this chapter - is the process of discovering interesting usage patterns from the interactions of the users while surfing the Web. Applications of Web usage mining include:

- understanding user tasks and user needs;
- recognizing usage patterns that indicate usability issues;
- target potential customers for ecommerce;
- enhance the quality and delivery of internet information services;
- improve Web server performance;
- identify potential advertisement locations;
- improve site design;
- fraud/intrusion detection;
- predict user's actions - for example for prefetching;
- personalization and adaptive sites.

As mentioned earlier, the first stage of Web usage mining is the collection and preprocessing of the data to be mined (Srivastava et al. 2000). In most cases, the usage data on itself is not sufficient for extracting interesting knowledge; background information on the information domain's content and structure, as well as on the user, is needed (Cooley 2003). In the next section we explain various approaches to collect the data and the preprocessing steps needed to make the data ready for analysis. Specific attention is given to a proxy-based approach of data collecting using the Scone framework. We have used this framework for several laboratory studies and a long-term field study. In the context of the latter study we found a high number of artifacts in the data that were not reported in earlier long-term studies or Web mining literature, most notably caused by advertisements and frame pages. Steps needed to remove these artifacts - which are peculiar to proxy-based and client-based data collection - will be dealt with.

The second stage is the inference of usable information from the observed data that can be used as an input for analysis or the Web personalization process. Many mechanisms can be employed for this purpose, including:

- *graph theory*: mathematical analysis of the node-and-link structure of Web sites and navigation paths. This technique will be dealt with in section three;
- *link analysis*: this method is closely related to graph-theoretic approaches, but link analysis includes more sophisticated methods for predicting the 'importance' of Web pages. The most well-known method is the PageRank system, originally used in the Google search engine;
- *probabilistic methods*: prediction of user navigation behavior, making use of stochastic models;

- *text analysis*: the indexing, scoring and categorization of textual documents, as well as discovering user interests making use of techniques developed in the fields of information retrieval and machine learning;

- *user segmentation*: the identification of groups of users, based on e.g. their demographics, background, interests, behavior and current user context.

In section three we present a graph-based framework and associated measures for knowledge inference from Web content, structure and usage. Emphasis will be put on measures that describe user navigation strategies and usability matters. You might find this section relatively hard-going on initial reading. If this is the case, feel free to limit the first reading to just grasping the overall ideas. When needed, we will refer to the corresponding sections when we use the measures in the studies described in the next two chapters.

After inferences of usable information, the outcomes need to be analyzed and interpreted. For usability analysis and user modeling purposes, this means understanding the meaning and implications of user actions. A next step is to make predictions on users' future actions, either in order to validate the results or to design personalized support mechanisms. This process is supported by statistical and machine learning methods and background knowledge from earlier studies and the literature. In addition to numerical analysis, visualization is a good mechanism for researchers to facilitate interpretation. These aspects will be dealt with in the last two sections.

4.2 Data collection and data preparation

The first stage of Web usage mining is data collection. Data can be collected from several sources. Each type of data collection differs not only in terms of the location of the data source, but also in the kinds of data available, in the preprocessing steps needed, and in the impact on user navigation behavior. In the first part of this section we will describe the pros and cons of data collection on server side, client side and using a Web proxy. We continue with a description of the Scone framework, which provides a mixture between proxy-based and client-based data collection. In the second part of this section we will concentrate on the preprocessing steps needed to prepare the data for analysis: session reconstruction, artifact removal and data enrichment.

4.2.1 Data collection

The most widely used source of navigation data is the Web server. Most servers store all transactions in a *Web server log.* Whereas the Web Consortium has defined a standardized log format, several Web servers have their own proprietary format. Typical fields included in the log file are:

- The *requested URL*;
- The *remote IP* of the machine from which the page request originated. This may be a user's computer, a provider's proxy server or any other machine from which a page request can be made;
- A *timestamp* of the date and time of the page request;
- The *method* used for requesting the page;
- The *status code*, which indicates whether the page was retrieved successfully;
- The *number of bytes sent* to answer the request;
- The *referrer*: if the user followed a link, the originating url is listed here;
- The *user agent*, which indicates the browser used, or which robot requested the page.

Obviously, the Web server log cannot provide details on all user navigation activities, as it contains only requests to one particular Web server. Requests to pages from other Web servers before, after, or in between a session with the Web server cannot be inferred. Several other issues need to be dealt with before the data from the Web server log can be analyzed (Cooley, Mobasher & Srivastava 1999) (Baldi et al. 2003) (Pierrakos, Paliouras, Papatheodorou & Spyropoulos 2003) (Facca & Lanzi 2005). The first issue is *data cleaning*. This step consists of removing all the requests that are not explicitly initiated by the user, such as additional files that are embedded in the originally requested page. Typically, in this step graphical content such as jpg and gif images are removed.

The second issue is *user identification*. Whereas each computer has its own unique IP address, several providers use a proxy server to balance Web traffic. Heuristics or technical tricks are needed to separate users behind the proxy. A possible solution is to send a cookie to the user computer, or to use a unique session identifier. Should this not be possible, it might still be possible to distinguish users based on their browser identifiers. Should none of these methods be of any avail, one can analyze whether a requested page is reachable from the last requested page. Further, not all users may be human. Sites such as search engines use *robots* for building and updating their database. The common politeness of having robots identifying themselves is not followed by all developers. Robots may be recognized by the fact that they explore the site more systematically and faster than users do, and return to the site at regular time intervals.

The third issue is *session identification*. It may be possible that users visit the site more than once. The goal of session identification is to divide the page accesses of each user into individual sessions. A session identifier might help to group these page requests in sessions. However, this breaks down if a user closes and reopens the browser window. A common heuristic is to assume that if a user has not been active for at least 30 minutes, the next page request will belong to a new session. The issue of session identification is dealt with in more detail later on in this section.

The fourth and last issue we mention is *path completion*. In practice, not all page requests reach the Web server. Browsers may store a page in their local cache for improving response time. In addition, several providers use a proxy server that stores popular pages for improving response time as well as for reducing Web traffic. A common approach in situations when two subsequent page visits to a and b are recorded, while no link between a and b exists, is to assume that a user backtracked from a to some page c, from which a link (b, c) exists.

With access to the server log, it is most likely that one also has access to the site content and structure. The url is a poor source for usage mining, as it does not provide any information on the page content. Additional information on the page title, contents and embedded objects can be retrieved from the Web server, as well as information on the site's link topology. As argued by (Cooley 2003), and dealt with in section 5.3, information on the site structure is essential for interpreting user navigation. In practice, in particular for dynamically generated sites with continuously changing content, it is often impossible to reconstruct individual page views.

In contrast to Web server logs, *client side logging* provides data on user navigation on multiple servers. As the data is collected on the user computer, problems associated with user recognition and path completion are eliminated or reduced. A popular early technique for client-side logging is the use of instrumented browsers (Claypool et al. 2001) (Tauscher & Greenberg 1997). However, browser modification is not a trivial task for modern browsers, even when their source code is available (Pierrakos et al. 2003). However, in particular for long-term studies it is undesirable to replace the user's preferred browser with a custom browser. An alternative approach is the use of agents that are embedded in Web pages, for example as Java applets, which are used to collect information directly from the client. IBM's SurfAid system (Facca & Lanzi 2005) features a JavaScript called Web Bug, which requests a 1x1 pixel image upon the loading of a page; the requests is generated with parameters identifying the Web page containing the script. A more intrusive method for client-side logging is the use of *spyware*; the software used by Kelly & Belkin (2004) is identified as a security thread by several security companies. Potentially, a large number of events can be recorded on client-side, from low-level events such as keystrokes and mouse clicks to higher-level events such as page requests. There are several limitations to client-side logging. First, client-side modifications might alter the user's normal behavior (Cockburn & McKenzie 2001). In addition, end-users may be unable or unwilling to install special software that might cause problems or that might form a security thread. Further, some mechanism is needed to retrieve the data

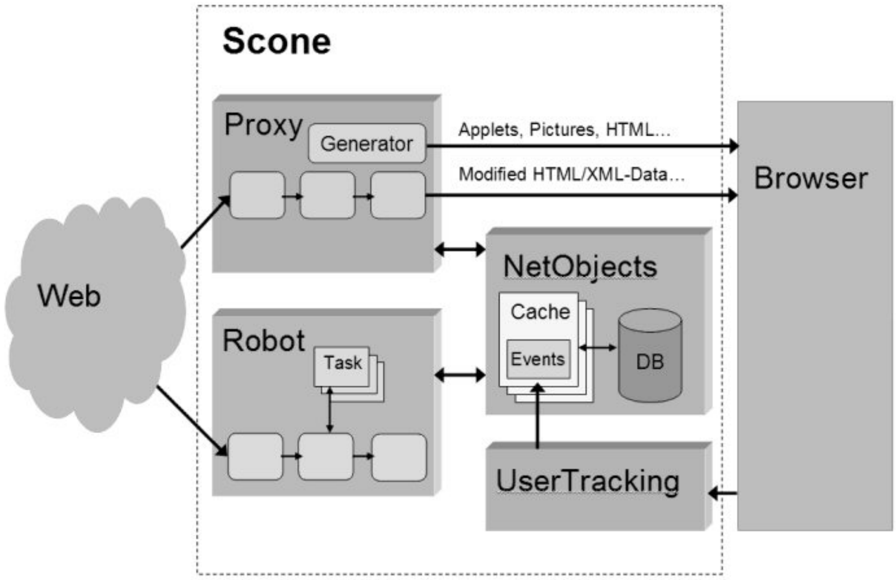

Figure 4.1: *High-level architecture of the Scone framework*

from the client computer (Waterson & Hong 2002).

As an alternative solution, one could make use of *proxy-based* logging techniques. A proxy server acts as an intermediary between client systems and the Internet. Many Internet service providers run a proxy server to improve their customers' navigation speed through caching. Proxy servers also use access logs, with similar format to the logs of Web servers. The main difference is that proxy servers collect data of *groups of users* accessing *many different Web servers*. In addition, proxy servers may be configured to modify pages being served through the proxy. The Web Browser Intelligence system (WBI) from IBM Almaden (Barrett, Maglio & Kellem 1997) is an extendable framework for Web personalization. It is composed of agents that monitor user actions, intercept and modify the communication stream, and provide additional services. Whereas the system was designed for Web personalization - such as personal history and shortcuts - it can also be used for inserting Javascript agents that communicate browser actions back to the server. A proxy-based solution has as an advantage that it requires less effort from users - they only have to instruct the browser to make use of the proxy server. Further, it can be used for multiple browsers and platforms, and any updates need to be applied only to one central instance. Like server-based logging, user identification may be a problem, as well as caching mechanisms. This may be solved by requiring users to register themselves and to ask them to turn off their browsers' cache during the registration procedure.

The Scone framework

In our studies we have made use of Scone (Weinreich et al. 2003). Scone is a Java-based framework for rapid development of Web enhancement and evaluation of Web tools, based on the WBI system (Barrett et al. 1997). The Scone framework is developed by a research team at the University of Hamburg. The core components of Scone are depicted in figure 4.1.

The *proxy* component extends the WBI proxy. All Web pages are delivered through the proxy. The information stream can be directed through a series of *plugins* for content extraction or modification. These processes are facilitated by html tokenizers, which split the content into tags and their associated parameters, text and comments. Tokens may be added, removed or changed in real-time.

The *user tracking* component analyzes and records user actions and page content. The component offers a user registration interface. In each page, a piece of Javascript is included that sends messages to the Scone server when a page is completely loaded and when the page is closed. Information in these messages includes a timestamp, a window identifier, an indication whether the html document is part of a frame page, and the navigation action used to request the page - following a link, using the back, next or reload button, manually entering of a url, opening a new window or selecting a bookmark. In addition, each page and each link receives a unique identifier. While parsing the page, the number of words, links and embedded images is counted. All information is stored in an object model, the *NetObjects*, which is permanently stored in a MySql database or written in a log file.

A *robot* component is available for crawling the Web for pages that users have not visited yet. The robot can be instructed to use a variety of crawling strategies, from breadth-first to depth-first search. Depending on its task, it explores only links within a site, only external links or both, until a specified depth is reached. Typically, the robot starts its crawling task at the moment a user loads a page. Alternatively, the robot can be used post hoc, for analysis purposes.

Various other help components are available. The most important of these is the *browser control* component, which can be used for automatically opening, closing or resizing browser windows, or to instruct the browser to load a certain page. This functionality can be used for offering the user tutorials or guided tours, or for automating usability studies. An optional Test Environment Automation (TEA) module, a Java-based user test environment, shows up on the left part of the screen and provides questions or additional information relevant to the current user context (Obendorf, Weinreich & Hass 2004).

Scone is developed as a tool for rapid prototyping of Web enhancements. An example enhancement is HyperScout (Weinreich & Lamersdorf 2000), a plugin that creates pop-up windows with additional information on each link, including the destination page's title, size, language, load time and when the user last visited it. The BrowsingIcons project (Mayer 2000) creates a trail visualization of the user past navigation. We developed task-related link annotations in the context of a laboratory study, reported in (Herder & Juvina 2004) and in chapter 6. The main benefit of the Scone system for the fields of user modeling and adaptive hypermedia is that it provides an integrated environment for prototyping personalization tools, creating controlled studies and monitoring user actions. We used the latter functionality in the context of a long-term study, of which the results are reported in chapter 7.

We started using the Scone framework in an early stage of development. Our main contribution to the framework is extensive testing, error reporting and debugging as a preparation for our laboratory studies. To support analysis of the recorded data, we developed a number of tools for calculating navigation measures - which will be dealt with in section 4.3 -, creating files that can be loaded into statistical packages, and visualizing user navigation paths and site structures. These tools are integrated into the *Navigation Visualizer*, which is described in section five.

Limitations are inherent to prototyping tools such as Scone. The main drawback of Scone is that it is only able to analyze and modify pages written in html. Sites that make use of binary formats such as Flash or Java cannot be analyzed properly, let alone modified. Secure connections cannot be analyzed either; however, for privacy reasons, one would discard data transmitted through secure connections anyway. A further inherent drawback is that Web enhancements can only be embedded into the Web page and not directly integrated into the Web browser. Hence, personalization of history mechanisms - which are currently part of the browser interface - can only be presented within a Web page or in a separate window. The response time of the Web becomes a bit slower as well, as all communication is directed through the proxy server; moreover, each page needs to be parsed, which costs additional time. A more disturbing drawback is the rather primitive html tokenizer. Current Web browsers have various mechanisms for dealing with erroneous coded html pages, which are manifold on the Web. With Scone, these errors could lead the parser to break down. In the past few years many mechanisms to deal with errors encountered during our studies have been incorporated into the tokenizer.

To summarize, Scone is a relatively stable proxy-based framework for Web usage mining and Web personalization. As several logging actions take place on the user's computer - the JavaScript included in the code - the data collection can be categorized as a mixture between proxy-based and client-based. In addition to laboratory studies, the framework has been used by twenty-five users for an average period of three months. Although the server did crash occasionally, and still a number of pages were not parsed correctly, the participants of the long-term study indicated that they did not experience major problems. Constant monitoring of the server status was necessary, though.

4.2.2 Data cleaning and data enrichment

The Scone *NetObjects* data structure, and the underlying database structure treat users, pages, links and page visits as related objects. The object-oriented approach is intuitive for constructing several queries, such as:

- how many users have visited this particular site;
- how many different pages are visited by this particular user;
- which pages are visited by the highest number of users;
- do users who visit more pages, also visit more sites.

Before mining the data, it is desirable to have a clean and rich data structure. In the previous section we identified several data cleaning steps: removal of embedded graphics from the log file, user identification, path completion and session identification. In this section we deal with session identification in a little more detail. Further, we present heuristics for a data preprocessing step that is not reported earlier, but that has become highly important for analyzing logs collected on client-side or at a proxy server: removal of artifacts caused by frames, advertisements and automatic reloads. We conclude with a discussion on data representation and enrichment.

Session identification

The goal of session identification is to split the list of user page accesses into individual *sessions*. It is a likely assumption that users do not access the Web continuously, but rather have several periods of increased navigation activity and periods in which they do not access the Web at all. These periods of increased navigation activity are assumed to represent a user task, or a set of user tasks. In order to analyze the navigation activity, session identification - or rather session reconstruction - attempts to split the Web log into meaningful chunks, which are called sessions. Whereas these sessions are related to the user activity, they are merely a technical concept, defined by the heuristics used during the reconstruction process.

The most common heuristic used for session reconstruction is a time-out mechanism; if the time between two subsequent page requests exceeds a certain time limit, it is assumed that the user is starting a new session (Cooley et al. 1999). Catledge and Pitkow (Catledge & Pitkow 1995) empirically derived a time-out of 25.5 minutes. They measured an average time between two page requests, which was 9.3 minutes, and added 1.5 standard deviations to this number. This is a standard statistical procedure to remove outliers from a dataset. However, this approach only makes sense for normal distributions. As reported by (Cockburn & McKenzie 2001), Web browsing is a rapid activity in which users spend only

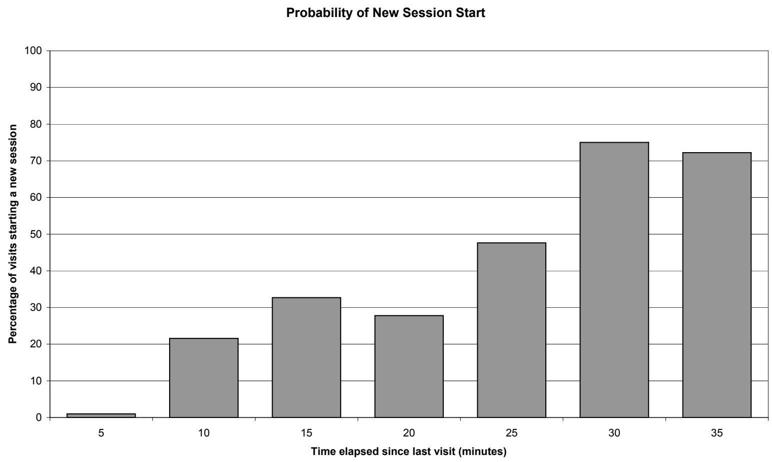

Figure 4.2: *At 25 minutes there is a 50% chance that a next visit will initiate a new session. The large histogram bins of five minutes is used to ensure that the number of samples is large enough*

little time on the far majority of pages. As may become clear from figure 4.4, 25.5 minutes is part of the long tail of the distribution and it would not make much difference to take twenty or forty minutes as a cutting point instead, based on the distribution itself. Nevertheless, based on the established 25.5 timeout, most researchers and commercial Web usage mining packages use this cutoff, in many cases rounded to 30 minutes.

Given the obviously faulty manner in which the 25.5 minute session timeout is derived from empirical data, it is surprising that it is frequently cited and used in the context of Web usage mining, without questioning its validity - for example (Baldi et al. 2003) (Cooley et al. 1999) (Spiliopoulou, Mobasher, Berendt & Nakagawa 2003)(Pierrakos et al. 2003). To find further justification for the 25.5 minute session timeout, one of the participants of our long-term study, reported in chapter 7, actively looked through about four months of his Web activities as recorded in the log file and manually marked the borders between subjectively meaningful sessions that coincide with separate user tasks. From figure 4.2 one can observe that the probability that a page visit marks the beginning of a new user task starts to exceed a threshold of 50% at about 25 minutes. Whereas this informal verification does not provide 'hard' evidence, it provides an explanation why the session timeout seems to work quite well in practice.

Several other heuristics for session reconstruction have been summarized and proposed by Spiliopoulou et al. (2003), mainly based on the site structure or the referring pages. She concludes that there is no best heuristic for all cases and that different heuristics lead to significantly different results. However, in order to split a large log file into individual chunks, some heuristic should be used.

Given the tradition of using the 25.5 minute session timeout, and the informal justification presented above, it makes sense to follow the tradition. However, when doing so, one has to realize that a session may consist of multiple tasks, and that one task may be split over multiple sessions.

Removal of artifacts

The first analyses conducted in the context of our long-term study showed that a preprocessing step not reported before was needed to get meaningful results, as a large number of page requests were not initiated by the users themselves. As will be explained later on in this section, it is of extreme importance to exclude these requests from the mining process, as - due to the large number - they interfere with many analyses, such as page popularity ranking and frequent path mining. Before explaining why this is the case, we first describe the artifacts found in our data.

A major artifact was formed by *framesets*, which break the page metaphor - what the user sees as a single page, may be constructed from multiple html files. As each html file causes an additional page request, multiple entries in the log file may relate to one user action. Due to the stateless character of the http protocol, it is impossible to identify which html files are loaded as part of one frameset and which files were loaded by user actions within the frameset. For this reason, we used a heuristic based on the window identifiers - which were embedded in the code by Scone - and a timeout mechanism. If an html file was loaded into an embedded window within a certain time interval after a frameset was loaded, it was assumed to be part of the frameset. Based on the page collections for various alternatives, we established a timeout of 3 seconds for slow ISDN connections and 1 second for LAN connections. For most analyses we excluded the frames that were most likely part of a frameset. In certain cases it seemed more appropriate to exclude all frame pages from analysis.

An even more severe problem was created by *advertisements*. JavaScript-initiated advertisements in pop-up windows are not deliberate user actions, and therefore excluded from the study. A statistically even more relevant advertisement technique is based on iFrames that allow embedding other HTML pages in a Web document. Our data showed that iFrames are currently a popular technique to dynamically include advertisements. We identified advertisements by different lists of known servers, typical patterns in Web addresses and equivocal frame names. For the group of participants who did not apply any kind of ad-blocker frame and advertisement artifacts represented over 28 percent of all page requests - see figure 4.3. This is remarkable, as it does not even consider online promotion realized as pure text, embedded images or flash animations.

A third source of non-user initiated page requests were *automatic page reloads*, mainly caused by news pages. Some participants had the habit of leaving a news Web site open in the background, while working on other tasks. Many news sites

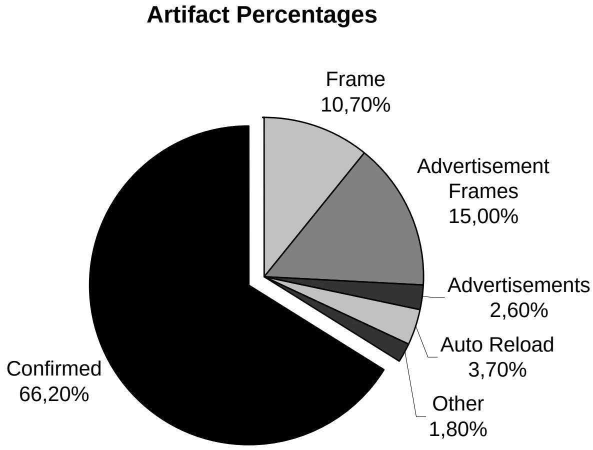

Figure 4.3: *Artifact percentages, generalized over all users.*

embed Javascript code that instruct the browser to refresh the page - and the headlines - after a certain time interval. These automatic page reloads became visible as peaks in the stay time distribution of certain users - see figure 4.4. In other cases external applications, like instant messaging agents, were responsible for the artifacts. These contributed nearly four percent of the page requests. However, the ratio differed severely between participants: some did not show any periodically reloaded pages, others over 20%.

As became clear during the analysis of the comprehensive datasets we had gathered, data cleaning and confirmation of user-initiated events were important to be able to relate recorded events to user actions. For example, many popular sites contain advertisements, which implies that each visit to the site also leads to one or more page requests to advertisement servers. As a consequence, some advertisements were present even in the list of ten most popular pages. Further, in several cases, session logs contained more requests to advertisements than to pages that deliver actual content. If one would build an user interest profile based on the content delivered, the profile would by highly skewed by these artifacts. Also, recommendations based on frequently found patterns would become useless, as many of these patterns would constitute of one user-initiated page request, followed by multiple automatically loaded advertisements.

Previous long-term studies did not use similar data consolidation methods, probably because the amount of such noise was lower in the past: in 1995, advertisements were still hardly known on the Web. Bruce McKenzie, who conducted the last reported long-term study (Cockburn & McKenzie 2001) reported that even in 2000 the effect of such requests could be neglected. The prominence

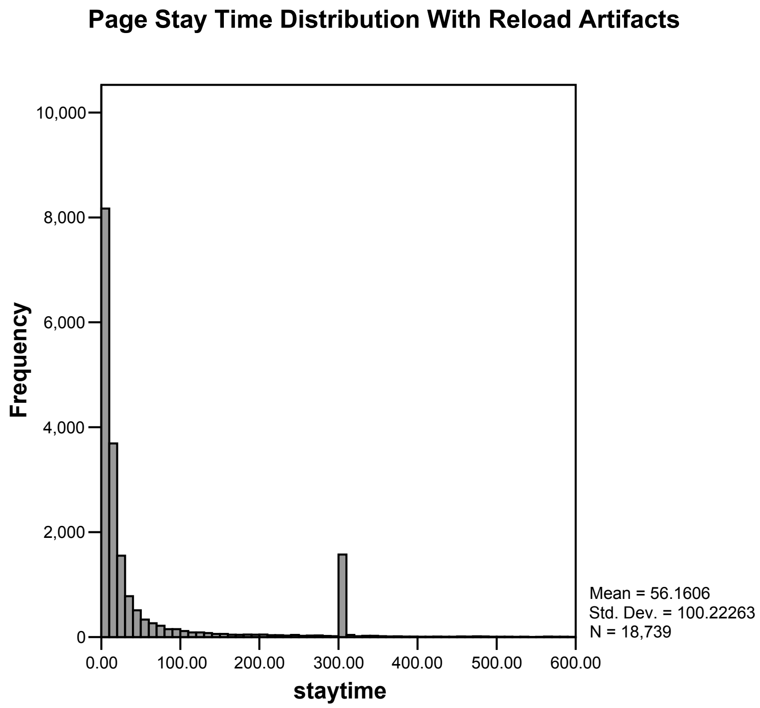

Figure 4.4: *Page stay time distribution. The peak at 300 seconds results from a news site automatically reloading its page contents every five minutes*

of advertisements can safely be interpreted as an indicator for the increasing commercialization of the Web (Network Working Group 2005). As said before, removal of artifacts due to advertisements has become a major issue in client-side and proxy-side logging. Most Web usage mining activity takes place at the server side, where these effects are not observed.

The heuristics presented in this section - timeout mechanisms for recognition of frame sets, lists of known advertisement servers and patterns in Web addresses, and peaks in the distribution of page stay times - have been developed mainly by extensive manual analysis. Whereas the resulting algorithmic artifact removal does remove the far majority of all artifacts, most likely it will not completely cleanse the data. Also, one has to be careful not to remove all suspicious Web activity. As an example, one of our participants had the habit of constantly pushing the reload button herself in order to receive the latest headlines. Therefore, the resulting data set may be considered ‘clean’, but not immaculate.

4.2.3 Data representation and enrichment

The exact representation needs and possibilities for the data depend on the richness of the data collected and the analyses envisaged. However, Web usage mining is typically an iterative process to find interesting patterns in user navigation. In order to stimulate curiosity and creativity, it is helpful to make the representation as rich as possible before starting the analysis or mining process.

As mentioned in the introduction, we need data on the page contents, site structure, usage and the users. Basic data that can be derived from most log files in a straightforward manner, possibly in combination with crawling the site for relevant content, include:

- *user*: demographic information, preferences, etcetera (see chapter 3);
- *page*: title, url, size, (statistics on the) content, name of site, number of visits;
- *link*: source page, destination page, link anchor, type of link;
- *visit*: page and user involved, time, duration, query or data submitted.

Given a larger log file, several additional objects and attributes can be created. These additional objects and attributes do not provide new data, but may be considered as *annotations*, which serve to bring the data to a higher level for interpretation. In an earlier framework for user navigation analysis, STRATDYN (Berendt & Bernstein 2001), a query language was proposed for analyzing user navigation. We believe a rich set of annotations provides a more flexible solution than a predefined query language.

As a first annotation, we split the log file into *sessions*, as explained in the previous section. This will facilitate comparison of page visits between sessions, in order to find association rules, such as 'if a user visits page a in a session, most likely page b will be visited as well'. It also serves to distinguish between backtracking activities and page revisits as a result of recurrent behavior. Each page visit that marked the beginning or the end of a session was marked as such. Further, all visits were numbered according to the order in which they were visited in the session.

A second annotation involves the type of page visit. Is it a first-time page visit, has the page been visited before in the same session, or has the page been visited before in earlier sessions, or both in the same and earlier sessions. As we will see in chapter 6, these different types of revisits tend to occur in chunks, which indicates that users employ the same strategy of alternatingly forward navigation and backtracking both for finding new information and revisiting earlier information. For each page revisit the time elapsed between the two visits, and the number of pages visited in between, were stored.

A third annotation reflects the use of search engines. As described in chapter 3, searching is a navigation activity that complements browsing and backtracking, which is mainly used to locate or relocate the *information patch* (Pirolli & Card 1999) for further exploration. Users do not only issue new queries; they often return to earlier result sets and iteratively refine their queries by adding, removing or modifying keywords. These refinement activities may be considered as *diet selection*, which is finding a new balance between the reducing the risk of not encountering relevant items and improving the chance of finding highly relevant items. Search refinement activities are dealt with in more detail in chapter 6. Further salient details on search engine use include the number of keywords used, and the presence of query modifiers.

Various further annotations were made. All site switching actions were marked as such, for determining the length of site visits. Further, an estimation was made of the number of windows opened at each action, based on the page access times and the stay time durations. A table with separate fields for date and time allows for recognizing recurrent behaviors at certain periods of the day. General statistics on each page's and site's number of visits, and the length and duration of each individual session facilitate distinguishing between different situations.

4.3 A graph-based framework

The Web and individual Web sites - and hypermedia documents in general - can be regarded as relational structures, consisting of a set of entities - the pages - and relationships between those entities - the hyperlinks. These structures are usually modeled as *graphs*. In such a graph, the pages are represented by *vertices*. Each hyperlink is represented by an *edge* that connects the corresponding vertices (Di Battista 1999). User navigation through the hyperdocument can be regarded as an overlay of the hypermedia document graph, consisting of only the pages visited and the transitions between these pages. These transitions are not necessarily the links in the document graph, as most hypermedia provide additional navigation options, such as backtracking, one-click access to the browser's 'home' page and switching between multiple windows.

There are many reasons to study the structure of hypermedia documents and user navigation paths through these structures (Baldi et al. 2003). The graph of the complete Web in particular is interesting from a theoretical or mathematical point of view, as it is a complex, growing and evolving network that reflects how humans organize information and that shares many properties with several other complex graphs found in social organizations and biological systems. From a more pragmatic point of view, the network structure is a rich source of information that can be used for algorithmic discovery of conceptual clusters, authoritative pages or sites, navigational hubs, structural complexity and various estimations of the navigation effort it would cost to navigate between any two pages - valuable knowledge for detecting potential usability problems and for developing efficient search and navigation algorithms. As will be explained in more detail in this section, the *user navigation graph* provides indicators of current, future and related user interests, navigation strategies followed and usability issues experienced.

The remainder of this section is structured as follows. In the first subsection we present a formal model of the *hypermedia structure graph* and the *hypermedia navigation graph* and various alternative aspects that can be represented by the model. We continue with an overview of algorithms, measures and characteristics of document graphs and user navigation through these graphs. Many of these measures have been used to relate navigation patterns to usability matters and success measures, with variable results. It has been argued Otter & Johnson (2000) that it is unlikely that one single measure can capture all aspects that are important to user navigation and he proposed to use 'a battery of measures' instead. In the last part of this section we argue that Otter's initial idea can be taken a little further, and we propose to use aggregate measures instead. This technique will be used in the analysis of our laboratory studies in the next chapter. We conclude with an overview on how the graph-based framework can be used as a basis for predictive Web usage modeling.

4.3.1 Graph models of hypermedia structure and user navigation

In order to describe the graph models of hypermedia structure and user navigation, we first need some basic definitions. The following is an informal introduction to graph theoretical concepts, mainly based on (Jungnickel & Schade 2000) and (Baldi et al. 2003).

Basic definitions

A (simple) *graph* $G = (V, E)$ is defined by a finite, non-empty set of *vertices* (or nodes) V and a set of edges E. An *edge* $e = \{v, w\}$ of E connects two vertices v and w ($v \neq w$). When two vertices are connected by an edge, the vertices are called *neighbors* of each other and *incident* with the edge. In regular (simple) graphs, edges are *undirected*, which means that no order is imposed on a pair $\{v, w\}$ of edge e. For a lot of applications, including hypermedia, it is useful to give direction to the edges and to call, in (v, w), v the *start vertex* and w the *end vertex*. Graphs with directed edges are called *directed graphs* or *digraphs*. Simple graphs can be generalized into *multigraphs*, which allow two vertices to be connected by more than one edge. Depending on the problem, it might be desirable to allow self-transitions - edges of the form $e = \{v, v\}$; these graphs are sometimes called *pseudographs*. In this thesis, when referring to a graph, we mean the *simple graph* as defined above.

Vertices v_0 and v_n are called *connected* if there exists a *path* between the two vertices, which is a sequence of edges $(e_1, e_2, \ldots, e_n)$ such that $e_i = \{v_{i-1}, v_i\}$ for $i = 1, \ldots, n$. If any two vertices of G are connected, G is called a *connected graph.* In directed graphs, if there exists a directed path between any two vertices of G, with all edges followed in forward direction, G is called *strongly connected*; a digraph is still called connected if any two vertices are connected by an undirected path following edges in either forward or backward direction.

For any vertex v of a graph, the *degree* is the number of edges incident with v. Analogously, in digraphs the number of edges having v as a start vertex is called the *outdegree* and the number of edges having v as an end vertex is called the *indegree.* If there is a path between two vertices v and w, the length d of the *shortest path* is called the *distance* $d(v, w)$. Vertices and edges can be given a *weight* that adds some meaning to the element. In network analysis the weights assigned to vertices are usually interpreted as the 'benefit' of visiting the vertex; the weights assigned to edges are usually interpreted as the 'cost' of traversing the edge.

A graph can be represented by an *adjacency matrix* M, where m_{ij} is 1 if there is an edge between vertices i and j, and 0 otherwise. The adjacency matrix of undirected graphs is symmetric, whereas the matrix is asymmetric for directed graphs. A different popular matrix representation is the *distance matrix* D, which may be derived from the adjacency matrix, after replacing the value of 0 of non-incident vertices with ∞. Each entry d_{ij} represents the distance between vertices i and j. A popular algorithm for calculating the distances between any two vertices is the Floyd-Warshall algorithm (Jungnickel & Schade 2000).

For several centrality measures it is not desirable to have infinite distances, which make it hard to determine the centrality of pages. For this reason, (Botafogo, Rivlin & Shneiderman 1992) introduced the *converted distance matrix* C, in which infinite distance values are replaced by a constant K, which may be a multitude of the number of vertices in the graph.

The hypermedia document graph

A model of a site structure expresses the way in which pages are related through hyperlinks. Conceptually, it is quite simple to obtain this structure: load the site's home page and recursively follow all internal links. Computer programs that autonomously navigate the Web and download documents are known as crawlers or spiders. However, for dynamically generated sites, especially sites that make use of personalization technologies, it is not trivial to determine which pages are equivalent. It is beyond the scope of this chapter to deal with the matter in detail. In (Cooley 2003) and (Baldi et al. 2003) the concepts are explained in more detail. A more conceptual problem is that an individual page view, the contents that are displayed in a browser window at some moment, may consist of several documents (Cooley 2003). For the sake of simplicity, we assume that each Web page that is modeled in the site structure reflects a possible page view, which may be either a single page or a page that consists of several frames or otherwise embedded pages.

As mentioned earlier, in the *hypermedia document graph* the vertices represent the set of pages P and the edges are associated with the set of hyperlinks L that connect the information contained in the documents. Depending on the problem and data at hand, several candidate sets for P come to mind. Example sets in the context of the Web are: the complete Web, one individual Web site, one cluster of a Web site, two related Web sites.

As explained in chapter 1, several definitions of the term 'Web site' can be thought of. From a user point of view, we think the most practical definition is by the uniform resource identifier's host part (Network Working Group 2005). In many cases, it is desirable to group several prefixes, as they refer to separate parts of one institution's Web site - for example www.foo.com, members.foo.com, www.cs.foo.com. Similarly, pages can be distinguished by the uri's path and query parts.

As we will see later on in this section, it desirable to take various properties of pages and hyperlinks into account, either or not in combination with analysis of the graph structure. For example, we might want to compare the link structure between pages within the same site with the link structure between pages of different sites. Or we might want to analyze the link properties of large pages. In order to facilitate these analyses, it is desirable to have an underlying *object model* (Rumbaugh, Blaha, Eddy & Lorensen 1991) from which various alternative graphs can be constructed. In the object model, each object (the page) and each relation between object (the link) can be given various qualitative and quantitative *attributes*. Similar to the approach followed by (De Bra, Aerts, Smits & Stash 2002), the selection of pages, links and attributes to take into account makes the resulting graph one of several possible *projections* of the domain model. As will be explained in section five, the attributes are essential for data filtering, marking and color coding for effective visual projections of the hyperdocument structure or user navigation.

For many analyses it suffices to treat all pages and links as equal by assigning them uniform weights. However, sometimes it is desirable to attach extra meaning - value, cost, relevance, popularity, authority - to them. Several weighting options, based on the page contents and its usage, are suggested by (Heer & Chi 2002) (Pirolli, Pitkow & Rao 1996) (Kleinberg 1999) (Claypool et al. 2001) and will be dealt with in more detail later on in this chapter. An important observation is that even weights based on objective observations represent a possibly subjective interpretation - for example, 'the more often a page is visited, the more important the page'. Another consideration to take into account is that the Web structure itself reflects various meanings: links from one site to another are often interpreted as an endorsement of the destination content's authority (Kleinberg 1999), whereas within-site linkage might better be interpreted as associative, causal or structural relations between source and destination. As will become clear from the next sections, interpretation is an important part of the analysis, as the graphs are not 'meaningless'.

The document graph may be an ordinary, undirected graph. Whereas some earlier hypermedia system did have bidirectional links or even *multilinks* that connected more than two pages, hyperlinks in the Web are unidirectional. Therefore, it would be natural to model the Web graph as a directed graph. For many measures this turns out not to be essential. Further, more than one link may exist between two pages, suggesting that a multigraph might be the best way to model the Web. As this will complicate analysis, multiple links between any two pages are commonly reduced to one edge.

Like (Baldi et al. 2003) we assume the document model in consideration is connected and ignore components that are disconnected from a given starting point - be it a Web crawler or a user session start. As we are interested in the document structure and user navigation from an adaptive hypermedia point of view, it does not make sense to take unreachable pages into account during analysis. If some choice of pages and links leads to unconnected graphs - for example all within-site links followed by some user - it would make more sense to treat the graph components as individual graphs in separate analyses and to compare the results.

The user navigation graph

Given the clickstream data of a user, we could model the navigation data as a time-ordered *sequence* of visits $V = \{v_1, v_2, \ldots, v_n\}$, where each $v = (u, p, t)$ represents a visit of user u to page $p \in P$ at time t. The sequence V is called a *navigation path* and contains as many elements as there were page visits.

The navigation path can be laid onto the hypermedia document graph following a visual metaphor: suppose we spin a chord by following the sequence of visits V in the document graph, starting at page p_i, which is associated with visit v_1. For each transition (v_i, v_{i+1}), the two corresponding documents get connected by the chord. For each transition, we might follow an existing hyperlink, or pass it in opposite direction - for example when using the back button - or we might have to pave our way through unexplored land when no hyperlink connection is present - for example when manually entering a url. The resulting 'Web' of connected pages $P' \subset P$ and created transitions is the *navigation graph*.

From this visual metaphor several observations can be made. First, once the navigation Web is spun, the sequential information is not directly visible; without retracing all steps, the best thing we can see is which transitions between pages have taken place and how often - the more often a link is followed, the more chords in the bundle that connects the pages. Second, pages that have been visited very often are now a gigantic knot. Third, suppose each user was given a different colored chord, transitions popular with many users will be cheerfully colored. Further, if we would obscure the hyperlinks and the unvisited pages, a graph quite similar to the document graph would remain: the *user navigation graph*.

Whereas the hypermedia document graph consists of all pages in the selected document set, the vertices in the hypermedia navigation graph represent the set P' of pages visited by a user or a user group, which is a subset of the corresponding page set P. Similarly, the edges represent the transitions between the pages T, as followed by the users. However, T may contain edges that are not member of hyperlink set L. These additional transitions may have been initiated by employing a browser's back button or by manually entered page locations.

The user navigation graph may be used to represent several alternative selections of user navigation, including:

- all pages from a Web site visited by a group of users;
- all pages in the Web that are visited by some user in a certain time interval;
- all pages in the Web that are visited by a group of users in a certain time interval;
- all pages in the Web that are visited by some user in one session;

Like the hypermedia document graph, several weights can be assigned to the vertices and edges in the user navigation graph. Traditional weighting approaches are to provide a page with a weight according to its estimated *benefit* for the user, and each link with a weight that represents the estimated *cost* for following the link, for example in terms of cognitive load, speed of the connection or the expected time for reading the next page. These weights are closely related to the *information foraging* framework presented in the previous chapter.

4.3.2 General Web structures

There is no direct evidence that browsing behavior and linkage statistics on the Web graph are related in any fundamental way. However, the success of systems that are based on this assumption, like Google's PageRank search algorithm, suggests that parallels can be drawn between the two (Dill, Kumar, McCurley, Rajagopalan, Sivakumar & Tomkins 2002). Moreover, the assumption matches the observation that the hypermedia document graph is the 'arena' in which user navigation takes place. In this subsection we describe several general properties of the Web arena and relate these properties to consequences for Web navigation.

The Web is the world's largest hypermedia document. It is a constantly evolving network in which pages are constantly added, edited or removed. These changes can have a dramatic impact on local scale. Pages that were previously not connected or distant from one another may become neighbors as a result of one added link. Existing paths may be broken (Teevan 2004). Whereas these changes occur in almost every region of the Web graph, due to the size of the Web several properties will continue to hold. Many of these properties reflect how humans organize information. Heavily linked regions indicate a close relationship between pages, which may be topological (a shared hobby), institutional (universities with similar research focus), geographical (companies in the same region) or any kind of relation that is recognized and possibly used in 'real life' as well. Similar to human societies, there is a relatively small number of locations - Web sites - that are highly popular. These sites are linked to from many locations and these sites attract a large number of visitors. These sites can be compared to celebrities in our society; everyone reads and talks about them. In contrast, there is a large 'working class'; John Doe's site on kite surfing will be visited and linked to by some friends and fellow kite surfing enthusiasts, a number that pales into insignificance compared to the attention received by the 'celebrities'. This notion is commonly known as the Pareto principle or the 80-20 rule: '20% of the Web sites receive 80% of the attention'. This principle can be observed in many real world contexts and is commonly approximated by a power-law distribution, also known as Zipf's Law:

$$P(X = k) = Ck^{-\gamma} \tag{4.1}$$

with $k = 1, 2, \ldots$, C a constant and $\gamma > 1$ a decay factor.

Zipf distributions can be found anywhere in reasonably large collections of Web data and Web usage data. Some examples (Broder, Kumar, Farzin, Raghaven, State, Tomkins & Wiener 2000) (Cockburn & McKenzie 2001) (Dill et al. 2002) (Baldi et al. 2003):

- the size of Web sites;
- the incoming and outgoing links from Web pages;
- the number of words in Web sites;
- the number of visitors to a site;
- the number of visitors to a page;
- the average length of a user navigation session;
- the average time users spend on pages;

In (Dill et al. 2002) it is shown that the structure of the Web is 'fractal': given a Web graph, the same (power-law) patterns can be found in almost any cohesive subregion - a cluster of Web pages that share a common trait, such as similar content. Due to this structure, the growth of the Web can be described in terms of *preferential attachment*: the more incoming links a page already has, the more likely it is that a newly added page will contain a link to this particular page. This preferential growth results in a robusts navigational backbone that ties together the collections of pages. Consequently, distances between any two connected pages are surprisingly small, even in large networks (Broder et al. 2000).

Given a large Web usage log - of one user or multiple user - the same structures can be found as in the Web graph. However, on the smaller scale of individual user sessions or a user's daily usage, the similarity to the 'random' graph of the Web is not that apparent. As users navigate by forward navigation actions (searching and browsing) and backward navigation, their trails have a more hierarchical structure in which linear sub paths can be recognized, cycles and vertices that have been visited more than once, and from which multiple trails are initiated.

4.3.3 Page measures

Local page measures are commonly estimates of a page's position, function, relevance and importance in a hypermedia document topology. The main reason for making use of the graph structure in addition to the information explicitly added by the author, is that the graph structure provides *implicit* information on these matters. In particular when a hypermedia document is gradually built by multiple authors - like the Web - no clear structure can be found in which users can orientate themselves. In the following we give an overview of existing approaches to implicitly inferring these page properties from their local structure.

Page position

In particular in sparsely linked hypermedia structures, the page position determines to a large extent the navigation effort needed to reach the page and the distance from this page to other pages. Whereas most larger Web sites provide keyword search as a shortcut, the link structure still determines to a large extent the sequence of pages followed as a result of *orienteering behavior*. The main reason for the interest in the page position is that in a document structure *central pages* are the most influential or most important ones. For this purpose, (Botafogo et al. 1992) defined two normalized centrality measures: the *relative out centrality ROC* (4.2) and the *relative in centrality RIC* (4.3) that make use of the converted distance matrix C. The higher the values of the *roc* and *ric*, the more central the page. In directed graphs, the ROC and RIC measures are in most cases different from one another; in undirected graphs, they are similar - due to the symmetry of the converted distance matrix. In this case, the ROC and RIC measures could be replaced by one *relative centrality* measure, RC.

$$ROC_i = \frac{\sum_i \sum_j C_{ij}}{\sum_j C_{ij}} \tag{4.2}$$

$$RIC_i = \frac{\sum_i \sum_j C_{ij}}{\sum_j C_{ji}} \tag{4.3}$$

Whereas the *ROC* and the *RIC* describe a page's distance to any other page, the *distance from the root* indicates the navigation effort needed to navigate from a site's starting point to a particular page. From the site graph a *spanning tree* (Jungnickel & Schade 2000) can be constructed, a subgraph that contains the minimum amount of edges needed to make it connected; the removal of any edge in the spanning tree would cause the graph to be disconnected. The edges from the spanning tree are called *hierarchical edges*; the remaining edges are called *cross-references* (Botafogo et al. 1992) or *associative links* (Park & Kim 2000). Most sites provide a spanning tree in the form of a site map or a menu structure built according to the site map. The spanning tree that provides the minimum distance from any page to the root is found by a breadth-first search. A physical metaphor for this procedure is 'picking up the graph at the root and letting it dangle will leave only a tree with the shortest paths taut'. Other links are cross-references.

Page function

In contrast to many closed hypermedia systems, the Web does not have an explicit structure or strong typing of pages and links (Pirolli et al. 1996) (Bieber et al. 1997). By lack of explicitly indicated page functions, various research efforts have been taken to infer the page function from its structural characteristics. These inferred functions do not necessarily match page functions as would be assigned by a human annotator. However, they do reflect the way the document is used and therefore may be useful if no explicitly defined categorization is available, or to verify whether a page is used for the purposes it was intended to be.

Botafogo et al. (1992) distinguished two categories of pages: *index pages* and *reference pages*. Index pages are defined as pages that serve as a guide to many pages; reference pages are linked to by many other pages and might contain background information relevant for all of the referring pages. Botafogo simply identified index pages and reference pages respectively by their outgoing and incoming links: the number needed to exceed the document's mean and a certain threshold. The method might work as a rough guide in traditional hypermedia documents. However, on the Web the distribution of outgoing or incoming links is a power law rather than a normal distribution, which implies that the mean is not a very relevant measure.

Pirolli et al. (1996) followed a more advanced approach and used additional measures to identify several categories of pages:

- *head page*: the first page one would visit in a set of related pages;
- *index page*: a page that serves to navigate users to relevant other pages;
- *reference page*: a page that is used to explain a concept or contains actual references;

- *content page*: a page of which the purpose is not to facilitate navigation, but to deliver information.

In addition to the in- and outdegree (Pirolli et al. 1996) formulated several hypotheses based on the page size, its textual similarity with the pages it links to, the length of paths than can be followed from the page and user page visit characteristics. *Head pages* are usually one of the first pages a user requests when visiting a site. Head pages have a reasonable number of outgoing links and most pages within the site can be reached from the head page. *Index pages* are relatively small, but they contain a large number of outgoing links. *Reference pages* are usually somewhere at the borders of the site, with few links pointing to and from them - note that this definition differs from Botafogo's definition of reference pages. *Content pages* are large in size, have a more central position in the structure, but they do not have many links, as their main purpose is to deliver information.

The head pages, as identified by (Pirolli et al. 1996), are usually a site's home page or the home page of a site's subsection, such as a university's departments. (Botafogo et al. 1992) used the term *root*, which is a common term in graph theory. According to Botafogo, 'the fundamental property of a root is that it has to reach every, or almost every, node in the hypertext' and 'the distance from it to any other node should not be too large'. In contrast to (Botafogo et al. 1992)'s condition that a root should not have too many children, many site's home pages can be categorized as index nodes with many children.

Kleinberg (1999) distinguished two other, closely related, categories of pages, which are defined in a recursive manner: *authorities* are those pages that are supported by many other pages linking to this page; of these linking pages, the most valuable ones are *hubs*, pages that link to many good authorities. The definitions of hubs and authorities are quite similar to Botafogo's definition of reference and index nodes. However, in addition to the page function, Kleinberg's definition includes the concept of *importance*.

Page importance

A different perspective on page typology is their relative importance.

The 'Hypertext Induced Topic Selection' (HITS) algorithm (Kleinberg 1999) aims to estimate the page relevance in a given set of pages $S = (V_S, E_S)$. It distinguishes relevance in two categories:

- page *authority* indicates the importance of a page, in terms of content;
- the *hub value* indicates to what extent a page links to important pages.

For specific queries there are very few pages that contain the required information, and it is often difficult to determine the identity of these pages. For broad-topic queries on expects to find thousands or even millions of relevant pages on the Web. The fundamental difficulty lies in the abundancy: the number of pages that could reasonably be returned as relevant is far too large for a human user to digest. To provide effective search methods under these conditions, one needs a way to filter, from among a huge collection of relevant pages, a small set of the most 'authoritative' ones. In the original formulation of HITS the set of pages consists of an unordered result set R_q of a *broad query*, expanded with all neighboring pages - any page linked to by a page in R_q and any page that points to a page in R_q. The basic intuition behind HITS is quite simple: a good authority is a page that is pointed to by many good hubs, and a good hub is a page that points to many good authorities. HITS is an iterative algorithm that starts with a uniform non-zero assignment of authority values a and hub values h. In each iteration i the authority and hub values are updated according to formula 4.4 and normalized. The algorithm is proven to terminate when the changes in authority and hubness do not exceed a certain threshold ϵ.

$$a_{p(i)} = \sum_{(q,p)\in E_S} h_{q(i-1)},\ h_{p(i)} = \sum_{(p,q)\in E_S} a_{q(i-1)} \tag{4.4}$$

HITS is originally designed for ranking pages in a query result set and ignores within-site links, as external links can be seen as *endorsement* whereas internal links most likely serve a different purpose - see the previous section. For other Web subgraphs or Web usage graphs the interpretation needs to be adapted, but the measures still carry some meaning. For a Web site, the authority and hubness values are better indicators for a *page function* as reference or index page respectively. In a navigation session, the values represent whether a page was *used* for reference or as a hub (index) to other pages.

The PageRank algorithm (Page, Brin, Motwani & Winograd 1999) is based on similar principles as HITS. However, unlike HITS, only one weight is assigned to pages. This weight may be interpreted as the probability that a *random surfer* will visit the page as a result of simply clicking on successive links at random. In other words, PageRank assumes a relationship between linkage patterns and usage statistics. If a page p is linked to by many pages, the probability that a random surfer will end up on that page will be high, in particular if some of the referring pages are linked to by many other pages. on the other hand, if all of the referring pages are hub pages with many outgoing links, the probability that a random surfer will follow one of the links to page p will be lower. A simplified version of PageRank iterates similarly to the HITS algorithm until the pagerank (formula 4.5) for each page stabilizes.

$$pagerank_p = \alpha \sum_{(q,p)\in E} \frac{pagerank_q}{outdegree_q} \tag{4.5}$$

As mentioned before, HITS takes as an input a subgraph matching some criterion - for example a query result set. By contrast, PageRank reflects the general importance of a page, independent of the query.

Several extensions and variations of PageRank and HITS have been developed, such as SALSA, PHITS (Probabilistic HITS) - see (Dhyani, Ng & Bhowmick 2002) and (Baldi et al. 2003) for an overview.

Page relevance

Page relevance is in principal a subjective criterion; it is the users who can judge best to what extent a page meets their general or current interests and the task at hand. Methods from the field of information retrieval are used for estimating the relevance based on the *similarity* between the page and some given task description or any other proximal cues - such as the queries issued by the user or recently visited pages.

The most straightforward method is *content-based ranking*, based on a similarity function on two pages p_1 and p_2, or on a page p and a query q. A classical approach is to represent a page p as a sequence of terms, $p = (t_1, t_2, \ldots, t_n)$, where n is the number of terms in consideration - which might be the set of all terms in the hyperdocument, or a relevant subset; t_i is a boolean that indicates the presence or absence of the term. The similarity function is the cosine of the angle formed by the two *page vectors*:

$$cos(p_1, p_2) = \frac{p_1^T p_2}{|p_1| \cdot |p_2|} \tag{4.6}$$

In practice, a weighting is used in order to distinguish rarely occurring terms from very frequently occurring terms, as rarely occurring terms are more likely to represent a page's distinguishing factors. The most common method is the tf.idf weighting. Given a collection of pages $P = \{p_1, p_2, \ldots, p_n\}$, each boolean t_{ij} is replaced by a weighted value x_{ij} according to the following formula, of which the first factor is called the *text frequency* (tf) and the second factor the *inverse document frequency* (idf).

$$x_{ij} = \frac{n_{ij}}{|p_j|} \cdot log(\frac{n}{n_j}) \tag{4.7}$$

n_{ij} denotes the number of occurrences of term t_j in page p_i, $|p_j|$ is the number of words in the page, and n_j the number of pages that contain term t_j at least once,

Page relevance can also be defined in terms of explicit user ratings or implicit feedback, based on user actions. Obvious interest indicators are retain actions: bookmarking or saving a page, writing down the address (Kelly & Teevan 2003). Several researchers have found that the time spent reading a page correlates positively with user interest (Claypool et al. 2001) (Kelly & Teevan 2003). However, reading times vary wildly between users and between tasks (Kellar et al. 2004). Given a large amount of data from multiple users, it is likely that there is small set of pages that many users find relevant. Pages that may be relevant for a particular user can be found using *collaborative filtering* techniques (McNee et al.

2003): a group of users is identified that visits similar pages; the pages most often visited by the group are most likely the most relevant ones.

4.3.4 Global measures

Global measures are used to describe the structure of a hypermedia document. From a usability point of view, important aspects to be considered are its navigational complexity (McEneaney 2001) - or *freedom* a document offers to the reader, and its size and distances between pages.

One of the most obvious measures of a hypermedia document is its *size*, defined as the number of pages in the document. The larger the document, the larger the distances one might need to navigate between any two pages, unless the number of links grows quadratical with the document size - in a fully connected document graph of size n each page would be linked to all $n-1$ other pages, resulting in a total of $n(n-1)$ links. Even for modestly sized documents it is undesirable to have a fully connected graph structure, as each page would need to contain an unreasonably high number of hyperlinks. This would clutter the screen and - moreover - disorient the user, due to the overwhelming number of navigation options and the lack of structure in the document; in a fully connected graph page locations are interchangeable, giving no point of reference for the question 'where am I'. However, in a sparsely linked structure navigation distances might become unreasonably large. In a densely connected hyperdocument, structure can be provided by suggesting or imposing an order of reading, for example by - manually or algorithmically generated - menus. Linear order of reading can be imposed by making the site structure more *hierarchical*. From these observations it becomes clear that the design of a document's navigation infrastructure is a balance between *system complexity* and potential navigation effort required in terms of *click distances*.

As explained earlier, the hypermedia structure determines to a large extent the navigation paths followed by the user; the navigation graph can be seen as a subgraph of the corresponding document graph, with additional edges that represent page transitions not initiated by links. Like hyperdocuments, navigation paths can be described in terms of size, distances and complexity. These terms will reflect the underlying document structure (Cooley 2003), the user's mental model of the structure (Dillon & Vaughan 1997) and the task at hand (Rauterberg 1992) (Rozanski et al. 2001). As described in more detail in chapter 1, a user's understanding of the document structure may vary from initial landmark knowledge to survey knowledge. A more complete mental model allows more complex - and possibly more efficient - navigation patterns, as users may confidently recognize and make use of shortcuts. However, complex navigation patterns may also indicate that users 'wander around', not being able to find what they are looking for. Rauterberg (1992) equated system (document) complexity (SC), the cognitive complexity of the mental model (CC), the complexity of user navigation behavior (BC) and the task complexity (TC) as follows:

$$BC - TC = SC - CC \tag{4.8}$$

Rauterberg (1992) assumes that the mental model CC of a document structure SC is always less complex than the actual complexity. The less complete the mental model, the more complex the navigation behavior BC will be, in comparison with the complexity of the task at hand TC. Equation 4.8 provides an intuitive reference point for comparing the various aspects involved in user navigation. Its practical use is quite limited, as in most cases the actual user task and the mental model are not known or cannot be inferred.

In this section we introduce several measures on document complexity and connectedness and associated navigational patterns are introduced. In the next section we explore in more detail how these measures can be put into for analyzing user navigation.

Size and connectivity of documents and navigation paths

Rauterberg (1992) defined four measures for measuring system complexity. The first measure 4.9 reflects the *document size*. As explained above, the larger the document size, the larger the inherent interaction complexity. Similarly, longer navigation paths indicate more complex interaction.

$$SC_{state} = |P| \text{ , } BC_{state} = |P'| \tag{4.9}$$

Rauterberg (1992) recognized the impact of the average number of links per page, which he reflected in the *fan degree* measure 4.10. The more navigation options per page, the more complex the document structure. Similarly, the larger the average number of transitions per page, the more complex the user navigation path.

$$SC_{fan} = \frac{|L|}{|P|} \text{ , } BC_{fan} = \frac{|T|}{|P'|} \tag{4.10}$$

.

As explained above, for larger documents the average number of links per page needs to be higher to avoid too large distances. Similarly, the more navigation options per page, the more likely the user is to initiate more transitions per page. In order to reflect this issue, (Rauterberg 1992) introduced a *density* measure, which is the ratio between the actual number of links and the theoretical maximum number of links 4.11.

$$SC_{density} = \frac{|L|}{|P| \cdot (|P| - 1)} \text{ , } BC_{density} = \frac{|T|}{|P'| \cdot (|P'| - 1)} \tag{4.11}$$

In a strictly hierarchical document, the only navigation option is moving forward. Cycles enable users to return to page visited earlier.

$$SC_{cycle} = |L| - |P| + 1 \ , \ BC_{cycle} = |T'| - |P'| + 1 \tag{4.12}$$

The Rauterberg measures on navigation behavior BC as introduced above have been used in several empirical evaluations (Rauterberg 1992) (Weibelzahl 2003). From the results it became clear that in particular the state and density measures on user navigation paths vary wildly with different tasks (Weibelzahl 2003). Therefore, in practice this means that in order to compare navigation complexity, the following two conditions should be met:

- the users must make use of the same hypermedia document;
- the users must work on the same task.

If one would want to compare navigation complexity in two related hypermedia documents - for example an adaptive system and its non-adaptive counterpart - the first condition can be toned down a little, ensuring that the documents under investigation are comparable in size.

Two more normalized measures on document complexity and connectedness have been proposed by Botafogo et al. (1992). The first measure, the *compactness* measure Cp, is related to Rauterberg's distance measure. A high compactness indicates that each page can easily reach any other page in the hyperdocument. The measure is defined independent of the document size, as a ratio of the sum of distances in the converted distance matrix C and the theoretical maximum and minimal total distances.

$$Compactness = \frac{Max - \sum_i \sum_j C_{ij}}{Max - Min} \tag{4.13}$$

$$\text{where } Max = K(|P|^2 - |P|) \text{ and } Min = |P|^2 - |P|$$

where K is the maximum value an entry in the converted distance matrix can assume.

The second measure, the *stratum* measure, suggests whether the hyperdocument structure implies an order for reading. A linear hyperdocument can only be read one way. If the hyperdocument is a cycle it structurally does not make any difference from what page to start. In fully connected hyperdocuments users can traverse the pages in any order they like. Of course, no structural difference does not automatically mean that there is no logical or semantic order. If a hypermedia author creates a page sequence $\{p_1, p_2, p_3, p_4\}$, it is most likely that the pages are intended to be read in this particular order. In strictly linear documents p_1 will not be reachable from p_4, and in hierarchical documents, p_1 may be reachable from p_4, but the distance will be quite high. Note that in undirected hyperdocuments - with bidirectional links - the distance from p_4 to p_1 is always the same as the distance from p_1 to p_4; no order is imposed. The stratum measure is based on an analogy with a company's hierarchical structure: in rigid organizations, the president only talks to the vice-presidents and for an employee at the bottom of the hierarchy it is very hard to reach the president without going through all the intermediaries; in more flexible organizations, contact between bosses and employees is accomplished much more easily. Companies of the first type are said to be *stratified*. All organization members carry some *prestige*; the highest prestige have those members who can easily reach other members - their *status*, but who are hard to be reached themselves - their *contrastatus*. Similarly, hypertexts can be structured in a very stratified manner - with few cross-references between the pages - or in a more liberal manner. The prestige of each page can be derived from the distance matrix D - *not* the converted distance matrix C using the following equations:

$$status_i = \sum_j d_{ij}, d_{ij} \neq \infty \tag{4.14}$$

$$contrastatus_i = \sum_j d_{ji}, d_{ji} \neq \infty \tag{4.15}$$

$$prestige_i = status_i - contrastatus_i \tag{4.16}$$

As the total status of the pages is equal to the total contrastatus, the total sum of a hyperdocument's page prestiges is always 0. The sum of the *absolute values* of the page prestiges, the *absolute prestige*, does represent a document's hierarchical organization. However, its value is not scalable; it will grow with the size of the hyperdocument. For this reason, in order to calculate the stratum, the absolute prestige is normalized with the *linear absolute prestige* (LAP), which is defined as the absolute prestige of a strictly linear document.

$$absolutePrestige = \sum_i |prestige_i| \tag{4.17}$$

$$LAP = \frac{n^3}{4}, \text{ if } n \text{ is even; } \frac{n^3 - n}{4}, \text{ if } n \text{ is odd.} \tag{4.18}$$

$$stratum = \frac{absolutePrestige}{LAP} \tag{4.19}$$

The compactness and stratum measures are related yet complementary. In general, more compact hyperdocuments are less hierarchical. Nevertheless, structures with large cycles will have a low stratum, but not a high compactness. (Botafogo et al. 1992) found that most hypertexts are far from hierarchical. As the normalization factor grows in the order of $O(n^3)$, it does not make sense to compare the stratums of hypertexts that have large differences in number of

	1	*2*	*3*	*4*	*5*	*6*	*status*	*prestige*
1	0	1	2	3	2	3	11	5
2	3	0	1	2	1	2	9	3
3	∞	∞	0	∞	∞	1	1	-9
4	1	2	3	0	1	4	11	5
5	2	3	4	1	0	5	15	11
6	∞	∞	∞	∞	∞	0	0	-15
contrastatus	6	6	10	6	4	15	47	

Table 4.1: *Example status and contrastatus table.*

pages. In addition, larger hyperdocuments might need to be more hierarchical after all, as more structuring is needed to prevent disorientation.

De Bra & Houben (1998) argue that the compactness and stratum measures do not take the browser's backtracking mechanisms into account. Hyperdocuments usually have a starting point - the root - to which users can backtrack. Consequently, even if there is no *directed path* between two pages, it usually still is possible to navigate between the two by backtracking to an earlier point from which the destination page can be reached via links. In order to take this property into account, De Bra & Houben (1998) proposed an alternative definition of the measures that allows links to be followed in backward direction. As a result, the hyperdocument's inferred compactness will be higher and its linearity will be lower - and, according to the authors, a better reflection of the document's properties. Even though this line of reasoning makes perfectly sense, a major problem associated with this approach is the fact that backtracking options are highly dependent on the path followed thus far. Moreover, for the compactness measure the converted distance matrix already reflects that any non-existing path may be followed at cost K.

Similar to the compactness and stratum measures as defined before, McEneaney (2001) defines the *path compactness* and the *path stratum* by only taking the set of visited pages P' and the page transitions T into account - the resulting path distance matrix D' is defined analogously. The normalization problem as experienced with the document compactness and stratum appears to be less severe, as no significant correlation was found between the path length and the associated measures. The path compactness and path stratum are suitable for comparing the complexity of user navigation paths if the underlying document structure is similar. From empirical results a significant yet weak correlation between the measures on compactness and stratum, and task outcomes was found (McEneaney 2001). In the next section on user navigation patterns we argue that second-order measures based on 'a battery of first-order measures' are more meaningful representations of user navigation behavior.

4.3.5 Navigation measures

The global measures as described in the previous subsection, were originally defined for hypermedia document graphs, but later applied to user navigation paths as indicators for the complexity of navigation. As mentioned in the introduction of this section, in the upcoming chapter we will describe two studies in which we infer navigation strategies from the (first-order) navigation measures by aggregating several navigation measures. In addition to the established measures described thus far, several other measures were taken into account as well and proven useful. These additional measures are described below.

The first measure is inspired by the concept of the *site diameter*, which reflects the maximum distance between any two pages in the hyperdocument. Mathematically, the diameter corresponds to the highest non-infinite value found in the distance matrix D. Sites with a large diameter impose potentially large navigation distances to the user. Correspondingly, a large diameter in a user navigation graph indicates either a strict linear navigation path - which occurs relatively rarely in practice in non-trivial navigation sessions - or that the user has made a 'huge leap' to a page visited many pages before. In a navigation graph with mainly backtracking to recent pages, most values in the distance matrix D will be either small or infinite. Backtracking to a distant page causes more pages to become connected, but with relatively long paths in between due to the size of the formed cycle. As a variation on the diameter, we define the *average connected distance* as the average distance between any two pages in the navigation path. This definition corresponds to the average over all non-infinite values d_{ij} in the distance matrix D. The larger the formed cycles, the longer users have waited before returning to an earlier point, which is an indication that they were confident that 'they will find their way back later'.

The second measure indicates the spread of page revisits. It is assumed that successful navigation strategies make use of the site's hub (index) pages to initiate several alternative trails for visiting content pages. Therefore, it is likely that this hub-oriented trail following will result in a few pages to be revisited relatively often. We define the *return rate* as the average number of visits to pages that have been visited at least twice. A low return rate indicates that users often visited content that they are already familiar with - which might be an accident caused by an incomplete mental model of the site structure, or an indication for wandering around. A high return rate indicates a more stratified navigation strategy.

Third, we mention several ratio measures. In most navigation models backtracking activities are considered as 'recovering from a failed attempt'. Whereas we argued in the last chapter that page revisits - in particular to hub pages - are part of successful navigation strategies, the back button itself does have a connotation of an *undo button*. For this reason, it makes sense to take the *percentage of back button actions* among all navigation into account as well by calculating the aggregated measures. In the last chapter we also presented the *recurrence rate* (Tauscher & Greenberg 1997), which indicates the percentage of revisits:

$$R = 100\% \times (1 - \frac{individual\, pages\, visited}{total\, page\, visits}) \tag{4.20}$$

One of the actions users are reported to take when experiencing problems, such as not being able to find what they need, is to return to the place from which they started (Otter & Johnson 2000). In experimental settings that involve only a number of Web sites, this can be measured as the percentage of visits to the site's home pages.

To conclude, we mention the *speed of navigation*. The time that users spend on pages is reported to be an important indicator for user interest and human factors (Shahabi, Zarkesh, Adibi & Shah 1997). People who are disoriented might speed up their browsing - comparable to anxiously running from one point to another - or they might just slow down. In our studies the *median view time* proved to correlate better with perceived disorientation than the average view time did. This is due to the fact that users spend only little time on the far majority of pages - they take a quick glance at a page and click a link that might bring them closer to their goals. Therefore, the average view time is overly influenced by the minority of 'high content' pages that they read carefully; the median view time indicates the average view time while browsing, not while reading.

4.4 Aggregated measures and predictive modeling

In the previous section a large number of measures on site structure and user navigation have been introduced. Whereas many of these measures have been used for analysis purposes, in most studies only a small number of measures has been used simultaneously. On the one hand, this is an advantage, as due to the sheer number of measures, it is almost inevitable that in a laboratory study at least some correlations between these measures and some experimental manipulation or questionnaires will be found. On the other hand, it is unlikely that all aspects of user navigation can be captured in a single measure. Otter & Johnson (2000) advocated the use of a 'battery of measures', independently used. Whereas we support this general idea, we think a better approach would be to combine the individual measures into aggregate measures that represent *patterns* in user navigation. These patterns may reflect user navigation strategies, the user task and usability issues encountered. This idea is further elaborated in the first part of this section.

In addition to the descriptive measures on empirical data, a substantial amount of research effort has been invested in the development of predictive models, based on cognitive theory or on statistical data. Descriptive data does not inherently provide explanations for the results, which makes it more difficult to generalize a given result to other circumstances. Predictive models describe the underlying processes that produce the recorded behavior and can provide insight in how users navigate through a hypermedia structure (Miller 2004). In addition, predictions on future actions can be used as a direct input for personalization applications (Baldi et al. 2003). A number of basic predictive modeling techniques used in the context of this thesis are presented at the end of this section.

4.4.1 Aggregate measures

User navigation can range from goal-directed task completion to more unstructured browsing and exploration of the availability of information or services (Shneiderman 1997). It is not unlikely that different navigation styles can be observed in a single session: free exploration may result in goal-directed search activities. Further, as described in chapter 2, domain knowledge and cognitive styles, such as field dependency, are likely to have an impact on the navigation strategies used while working on a task. Therefore, even with knowledge of the type of session in which a user is engaged, it is unlikely that one single measure on user navigation will be a reliable indicator of any usability issue.

For example, backtracking is often considered as 'recovering from a failure'. However, hyperdocuments are designed for non-linear navigation, making use of landmarks such as index pages. The Web in particular is shown to be a recurrent system, in which page revisits and backtracking are very common activities (Tauscher & Greenberg 1997). If we compare two studies that aim to predict a user's vulnerability to experiencing 'lostness in hyperspace', it becomes clear that one single measure can provide different results in different contexts. Smith (1996) takes as a starting point for her lostness measure the assertion that lostness should be viewed in terms of degradation of user performance. She proposes a lostness rating, based on the relative amount of revisits while searching, and on the number of navigation actions compared to the minimum number required. In other words, any form of backtracking is used as a piece of evidence for lostness. McEneaney (2001), on the other hand, concludes that users who employ shallow, hierarchical search strategies are more successful in their search than those who followed more linear paths. In other words, revisits to navigational landmarks can be seen as a sign that users have formed an accurate model of a Web site.

A logical approach to solve the inexpressiveness of single measures is to use a 'battery of measures', as proposed by (Otter & Johnson 2000). According to Otter, for measuring lostness the measures need to correlate well with one another and have been shown to measure lostness to some degree. However, given the complexity of user navigation, one cannot simply rely on a - possibly large - set of independent indicators for some navigation pattern or usability issue, as

- some indicators will be more important or reliable than others;
- indicators might be highly correlated, positively or negatively, or not correlated at all.

As the navigation measures aim to reflect underlying dimensions - such as user tasks, navigation strategies and usability matters - one can assume that these underlying dimensions act as *latent variables*. Hence, the 'battery of measures' should be grouped and reduced to a smaller number of uncorrelated *factors*. In addition, it should be made plausible that these factors are meaningful in terms of the underlying dimensions, and shown that these factors are useful for analysis or user modeling purposes.

In social sciences, the statistical technique of *factor analysis* (Field 2005) is frequently used for data reduction. The basic principle behind factor analysis is to reduce the matrix of correlations between variables to its component dimensions, by grouping variables that correlate highly with one another, but badly with other variables. Strictly speaking, this form of factor analysis is called *principal component analysis*. Each factor is represented by a linear equation of variables X_i, with a weight (factor loading) β_i:

$$Factor_j = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta n X_n \tag{4.21}$$

The most common method of calculating the factor loadings is *regression*. In this method, the factor loadings are adjusted to take account of the initial correlations between variables and to overcome differences in units of measurement and variable variances. Various statistical techniques exist for determining the number of factors to extract, based on the eigenvalue of the factors. In order to make it clearer which variables relate to which factors, the factor axes can be *rotated* - the absolute values of the variables are changed, whilst keeping their differential values constant.

The major assumption in factor analysis is that the factors represent the underlying, real-world dimensions of the data. Which dimensions they represent, can be found out by *interpreting* the factor loadings with the highest absolute values and correlations of these factors with other data - such as user test results. In a laboratory study, which is described in more detail in the next chapter, we applied factor analysis to navigation measures in order to find factors, which we called *navigation strategies*. One of these strategies, which we labeled *flimsy navigation* was characterized by slow navigation (short navigation path, long median view time), frequent backtracking (frequent use of back button, short average connected distance) and frequent revisits to a site's home page. Flimsy navigation was mainly found by users with little Internet expertise and a short working memory.

Several conditions should be met for successful factor analysis on navigation measures. First, the underlying dimensions should be known or hypothesized beforehand, and small in number. In our laboratory study, we limited the dimension to user navigation styles by giving the participants well-described tasks and by ensuring they would make use of the same Web site. If all aspects that influence user navigation are variable, it is most likely that no meaningful factors will be found; one will always find factors, but these are meaningless, or hard to interpret at least. A second condition is that the interpretation is as objective as possible. A researcher looking for a factor that supports his hypotheses, will most likely be able to find an explanation that does do. In our study, we opted for the solution of independently interpreting and labeling the factors found, ensuring that the interpretations were more or less similar, and to improve the interpretation by further discussion and (visual) analysis of the navigation behavior of users that scored either low or high on some factor. A third condition - or rather a consideration - is that factor analysis requires a large number of decisions from the researcher. This implies that the technique is not suitable for *proofs*, but rather for *exploratory* research.

A more goal-driven alternative for factor analysis is *regression*. Similar to factor analysis, a regression model is a linear combination of data (see equation 4.21. In addition to the data and their weights, a constant β_0 is added to the formula. Instead of a factor, the linear combination aims to explain the outcome variable - for example a user's experienced lostness. Several statistical tests should be used for testing how well the regression *fits* the outcome, tests which are covered in statistics books such as (Field 2005). Similar to factors, regression models may be interpreted by inspecting and interpreting the variable's weights. However, due to the goal-oriented manner in which it is constructed, the regression model cannot be regarded as a latent variable. Therefore, the regression model may tell that users who feel disoriented tend to visit less pages and to use the back button to a higher extend, but it cannot be interpreted as a 'disoriented navigation style'.

The basic statistical methods as dealt with in this section are rarely applied in user navigation analysis. A plausible explanation is given by Weibelzahl (2005), who observed a general lack of statistical approaches in adaptive hypermedia research. By contrast, in the field of Web usage mining statistical techniques are frequently used, in the disguise of machine learning algorithms for predictive modeling. Whereas predictive modeling has merits on itself - as will be explained in the next section -, it puts little emphasis on interpretation. Therefore, we think that techniques as dealt with in this section should become more commonplace in the field of user modeling and adaptive hypermedia.

4.4.2 Machine learning

The measures for Web graphs and user navigation graphs as described thus far in this chapter are mainly descriptive. The statistics that they provide are useful for finding and understanding navigation patterns, which give a quantitative insight in user tasks, user navigation strategies and usability issues. As argued in the last chapter, descriptive models have their limitations for modeling user navigation.

The simple and aggregated measures give a post-hoc picture of the navigation process, but fail to describe user navigation decisions at each individual step. Probabilistic models of browsing behavior attempt to fill this gap by predicting future navigation actions based on past observations. Probabilistic modeling can both generate insight in how we use the Web and provide mechanisms for making predictions that can be used for personalization. In the previous chapter we have seen predictive models based on information foraging theory and the CoLiDeS model. In this section some basic predictive modeling techniques are introduced.

The basic intuition behind predictive modeling is that some actions occur more often than other actions. The probability $P(a), 0 \leq P(a) \leq 1$ indicates the a priori likelihood of an action a. $P(a)$ will be high for actions that occur frequently and low for infrequent actions. The likelihood of an action a will become higher if an action b is observed that tends to be followed by a. The conditional probability $P(a|b)$ represents the likelihood that a will happen given *evidence* that b occurred. In practice, this is the ratio between the number of times that b is followed by a and the number of times that b occurred. Or, more generalized, the conditional probability $P(a|E)$ is the likelihood that action a will occur, given some evidence E. Bayes' theorem (4.22) is used for updating the conditional probability:

$$P(a|E) = \frac{P(E|a)P(a)}{P(E)} \tag{4.22}$$

A straightforward probabilistic model for page prediction is the (first order) Markov model. Under a first-order Markov assumption, it is assumed that the probability of each state s_t given the preceding states $s_{t-1}, \ldots, s_1$ only depends on the last state s_{t-1}:

$$P(s_n) = P(s_1) \prod_{t=2}^{n} P(s_t|s_{t-1}) \tag{4.23}$$

It can be shown (Baldi et al. 2003) that the probability of a state transition t_{ij} from s_i to s_j equals to the ratio between the number of transitions from s_i to s_j, n_{ij} and the total number of transitions from s_i, n_i observed thus far.

$$P(t_{ij}) = \frac{n_{ij}}{n_i} \tag{4.24}$$

In terms of the hypermedia document graph, a Markov model can be represented by applying a weight to the links, equal to equation 4.24. This weight should be interpreted as the probability that it will be followed.

In practice it is infeasible to use a simple Markov model for predicting page visits within Web sites, as the number of pages in a typical Web site is quite large. For this reason, in most applications the pages are grouped in a number of categories; in this case, the Markov model would predict which page category will be visited next. These categories can be defined in terms of page functionality, contents or position, but also in terms of their usage.

Various variations of first order Markov models have been proposed in the literature. The most obvious generalization is to use k-th order Markov models, which assume that the probability of each next state depends on the last k states. An alternative approach is to use *mixtures* of Markov chains, in which each navigation behavior is described by different Markov chains, each representing different groupings of pages.

Obviously, there are more parameters involved to predict the next page to be visited than just the last page or the last page category. As an example, we have found in our empirical data that the average page view time before backtracking actions is significantly lower than the page view time before forward navigation actions, such as link following. If we assume that all these parameters are independent of one another, we can use a *naive Bayes classifier*, which employs Bayes' theorem (4.22) to learn the probability of a page visit from the product of the probabilities for the individual parameters $a_1, \ldots, a_n$.

$$P(s_i) = P(s_i) \prod_{j=1}^{n} P(a_j|s_i) \tag{4.25}$$

The naive Bayes learning method involves learning the unconditioned probabilities $P(s_i)$ and the associated conditional probabilities of observing the parameters a_j, $P(a_j|s_i$, based on their frequencies over the training data (Mitchell 1997). Despite its naive assumption that the parameters involved are independent of one another, naive Bayes classifiers have been used successfully in a variety of systems. A more general model that does take dependencies between parameters into account are Bayesian belief networks, which takes dependencies of *subsets* of these parameters into account (Mitchell 1997) (Russell & Norvig 1995).

If it is likely that a user population consists of different user groups, or that different groups of pages can be distinguished, a *clustering* algorithm can be used to derive these clusters. The assumption behind clustering algorithms is that the probability of a certain event is the result of a mixture of k different (normal) distributions. As an example, if we have a user group that contains both German and Dutch users, it is likely that the probability distributions of page visits will be more similar between any two German users than between any Dutch and German users. One of the simplest and most widely used algorithms to find these clusters is the k-means algorithm, which works as follows:

- first, for each cluster an initial center point is initially chosen;
- each point in the data is assigned to the cluster associated with the closest center point;
- after the assignment, new center points are calculated, for instance by taking average values of all data points assigned to a cluster.

The last two steps are repeated until the fluctuations in assignment remain small. For a more extensive treatment of clustering techniques we refer to (Mitchell 1997). Clustering techniques have been used for discovering stages in task-based Web navigation, with the assumption that each page in a Web site has a different function in the task solving process (Hollink, Van Someren & Ten Hagen 2005).

Whereas some probability distributions and predictive models based on user navigation might not be directly useful for personalization due to an insufficient high predictive power, they do characterize user navigation. Ideally, a good predictive model - for example, based on information scent - should be able to construct navigation graphs with similar properties as ‘real’ user navigation graphs.

4.5 The Navigation Visualizer

The need for an interactive Web usage visualization tool arose during our laboratory studies, in which we were looking for specific user navigation patterns that indicate users’ disorientation while carrying out tasks on various Web sites (Herder & Juvina 2004). After having calculated a large number of measures, we wanted to see what these measures really meant. Despite the availability of several Web usage visualization tools, no tool was found that allowed for visualization and exploration of the various patterns in Web navigation.

As seen in the previous sections, the data that can be collected from Web usage logs is highly multidimensional, varying from basic statistics - such as the relative amount of back button usage - to more aggregate data, such as the linearity of user navigation paths and semantic similarity of documents to a query (Herder & Van Dijk 2005). While it is possible to put these data into tables and to analyze the data with statistical software or data mining applications, there is often still the need for visual analysis to enable researchers to interpret the meaning of the available data (Chi 2002). As visualizations allows analysts to survey a large and high-dimensional data set, higher order patterns that often remain unnoticed in a statistical analysis based on numbers, can be *detected* (Berendt & Bernstein 2001). Conversely, visualizations allow analysts to *interpret* patterns detected in statistical data. Unfortunately, it is hard - if not impossible - to visualize a large number of dimensions of numerous items in a workable manner, as discernibility and understandability decrease with the amount of data that is displayed (Herman, Melançon & Marshall 2000). Several approaches to reduce the amount of data, such as clustering, binding, filtering and hierarchization of documents or usage patterns, have been applied in various systems (Chen 1999) (Chi 2002). However, the resulting abstract views are often too complex to interpret or to compare. By contrast, traditional graph visualization of the node-and-link structure - for example WebQuilt (Waterson & Hong 2002) - is easy to understand, but badly scalable and it hardly allows visualizing additional metadata, such as page view times, back button usage, or page relevancy. Especially when evaluating various possible effects of adaptive hypermedia techniques or site redesigns, the interpretation of the recorded data has to be supported.

In this section we introduce a graph-based Web usage visualization tool - the Navigation Visualizer (Herder & Weinreich 2005) - that allows researchers to dynamically select and match the data to be shown and to interactively explore the visualization. Selection criteria for users, Web sites and time slices are used to narrow down or to broaden the number of items; various color codings and markers are available to compare several dimensions side by side; zoom and pan tools (Chen 1999) are used to explore larger graphs. The tool also offers means to display and export data in a tabular format to be further processed by statistical software.

The remainder of this section is structured as follows. First, we discuss related work on graphical overviews, Web usage mining and Web usage visualization. Using the insights drawn from the related work, we describe our design rationale. In the third part we give an overview of the actual system and its interface. We conclude with a summary and a short discussion.

4.5.1 Background and related work

Graph drawing is a common technique for information visualization and it has been subject of research since decades (Chen 1999) (Herman et al. 2000). Graphs are a natural means to model the structure of the Web, with the pages represented by nodes and the links represented by edges. One of the most important issues in graph drawing is how to effectively visualize large information spaces. Common criteria for a usable, well-drawn graph are minimized edge crossings, uniform edge lengths, straight edges and symmetry. In the graph visualization community, graph layout algorithms that attempt to adhere to these criteria, is still a topic of research. But even if it is possible to layout and display all the elements, the issues of viewability and usability arise, because it will become impossible to discern between nodes and edges. Usability becomes an issue even before the problem of discernibility is reached, as cognitive load grows with the size of the graph (Herman et al. 2000).

A common approach for enhancing usability in information visualizations is popularized by Ben Shneiderman as the Visual Information Seeking Mantra: 'overview first, zoom and filter, then details-on-demand'. The *overview* supports the first step of visual analysis of getting the 'big picture', which is needed for effectively focusing on the interesting details (Chen 1999). An effective mechanism for visualizing a large number of items is the *fish eye view* (Furnas 1997), in which items close to the region of interest are shown in most detail, and items in further regions in successively less detail. If the number of items is too large to fit onto one screen, *panning tools*, such as sliders, allow users to move the viewport. *Zooming* and *filtering* are two complementary strategies for moving from a global view to a more specific view. In contrast to zooming - the act of focusing on items of interest, typically using widgets that directly control the visualization - filtering operations filter out uninteresting items using dynamic queries, sliders, buttons or other control widgets. Once a visualization contains a manageable number of items, users may want to view *details on demand.* In addition to details specifically requested by the user, *bindings* can be specified between the information attributes and the visual attributes of the vertices and edges (Mukherjea & Foley 1995).

Various visualizations of user navigation paths have been developed in the past, in most cases as a support tool for analysis of usability studies. As the number of nodes in most experimental tasks is small, compared to the size of even relatively small Web sites, scalability and algorithmic efficiency are of lesser concern for these tools (Berendt & Bernstein 2001). Whereas user navigation is linear in time, it is common to use one node for each visited page, even if it is visited multiple times. This approach is quite similar to the concept of the *user navigation graph*, as introduced in section 4.3.1.

The WebQuilt Visualization System (Waterson & Hong 2002) aggregates navigation data from multiple users, as gathered using a proxy logger, into an interactive graph. In the overview mode, visited Web pages are represented by color blocks; the color indicates whether a page was an *entry* page or an *exit* page. The thickness of the edges indicates how heavily the corresponding link was traversed; the color of the edges indicates the average amount of time spent before traversing the link. The navigation paths can be compared with a *storyboard*, an optimal navigation path as defined by the analyst. Whereas WebQuilt provides zooming and filtering techniques, the system does not provide alternative binding options, which would allow the analyst to explore several aspects of user navigation. The Footprints system (Wexelblat 1999) provides a similar graph visualization, albeit with the different goal of helping users in information finding by showing the trails (footprints) of other users. Experimental results showed that the graph was of no help for 'naive' users, whereas experienced users - who had a proper mental model of the information domain - were a little faster in solving their tasks. The WebView system (Cockburn, Greenberg, KcKenzie, Jasonsmith & Kaasten 1999) adopts a different visualization approach by organizing the navigation trail in a 'hub-and-spoke' view, a structure similar to the Windows Explorer's collapsing file tree. An interesting feature is the dogear metaphor, which indicates how often a page has been visited. The STRATDYN system (Berendt & Bernstein 2001) is a visualization tool that allows analysis of navigation patterns. The graph layout is defined in such a way, that the x dimension (left-right) encodes temporal order of page visits; the y dimension encodes temporal order as well; multiple links followed from one node are temporally ordered from top to bottom. Unfortunately, the edges are restricted to horizontal and vertical lines, which tends to result in very dense maze-like structure.

All of the systems mentioned above employ their own data format, which implies that they cannot be directly used in other contexts. Most existing systems did not provide extensive data filtering techniques, whereas in particular for analysis of larger usage logs it is desirable to limit a visualization to some time period, one or more Web sites related to a particular (sub) task or a certain user group. Moreover, whereas all systems did use bindings between information attributes and visual attributes, these bindings cannot be changed; given the many attributes that can be derived from Web usage logs, it is desirable to be able to inspect and compare these attributes. For these reasons, we decided to build a more interactive visualization tool that provides analysts more freedom to find the matters that they are interested in.

4.5.2 Design rationale

As mentioned in the introduction, visual analysis allows analysts to explore a high-dimensional data set and to search for possibly relevant statistics for further investigation. In addition, visualizations facilitate interpretation of patterns found in the statistics. This implies that visual and numerical analysis are complementary activities, which ideally should be combined; in addition to a visualization, the system should facilitate statistical analysis. Given the functionality of existing (commercial) packages for statistical analysis and data mining, we chose to concentrate on the *preprocessing* step, which involves calculation of the various measures introduced earlier in this chapter. These measures as well as raw data can be exported in a file format compatible with regular statistical and data mining packages.

The primary focus of the system is the visualization of navigation paths. As most analyses will concentrate of individual navigation sessions, some defined time interval or a certain part of the internet - depending on the research question - we expect the analysist to specify these restrictions beforehand; alternatively, one could gradually restrict or loosen the restrictions. The process of restrictions specification is divided into three consecutive steps:

- select the user or group of users;
- optionally one can specify a session or a time interval;
- optionally restrict the visualization or analysis to one Web site or a number of Web sites;

Of the various graph layout mechanisms available (Di Battista 1999) a basic hierarchical tree layout is the most suitable choice for a variety of reasons. Recall from section 4.3.3 that a breadth-first spanning tree can be obtained by 'picking up the graph at the root and letting it dangle'. A major advantage of the spanning tree layout is that it preserves temporal order to a larger extent than iterative layout algorithms such as force-directed layout. Similar to the STRAT-DYN system, the x and y dimensions encode the temporal order of page visits.

Further, it matches our intuition of forward and backward navigation; forward navigation is visualized as further going down the tree and backward navigation - revisiting pages - is visualized as edges that point to a different direction. Even more important, imposing an hierarchy on the graph makes it less cognitively demanding, even if the graph is not hierarchical itself (Mukherjea & Foley 1995).

Several visual attributes can be used and varied for representing the various information attributes. The most prominent ones are color, shape, size and shading. However, if you ever have played the game of Set - the goal of the game is to find sets of cards that are all alike or all different in color, shape, number and shading - you might have experienced that it is hard to examine multiple continuous scales at once. For this reason, we chose to use just one continuous attribute - color - and two additional markers, visualized as a differently colored border of the node. The markers are used for distinguishing the top n or bottom n items from some ranking scale. The colors and markers are interactive: the user is able to dynamically change the binding with information attributes.

It is known that neither page title, Web address or thumbnail is sufficient for easily distinguishing between pages (Cockburn & McKenzie 2001). As it would consume too much screen real estate to display these three elements simultaneously, the user can switch between page title and Web address. We left the thumbnail option out for two reasons: first, it breaks down the color binding and second, thumbnails should be reasonably large to be useful. Instead, we chose to provide the page content as details-on-demand.

4.5.3 System overview

Upon starting the Navigation Visualizer, the user is presented a dialog as displayed in figure 4.6. The first step to generate a navigation graph is to select the data source, which is a MySql database containing data in Scone format (see section 4.2.2. If needed, additional tables can be created that enrich the data with annotations. In addition to creating a graph visualization, the user can also choose to analyze the data; this option results in a set of comma separated value (csv) files with navigation measures of each user and each individual navigation session. The csv file format is supported by most statistical packages. An alternative, linear *enhanced log file visualization* in the form of html files can also be created. In this section we assume that the user wants to inspect a navigation graph. We will come back to the linear log file visualization later.

The data selection is done in three steps. The first step, in the bottom left part of the dialog, is selecting the user or group of users whose navigation we want to be visualized. After selecting the user or users, by default it is assumed that the entire navigation history should be visualized. The second step involves restricting the time window to be visualized, either by adjusting the start and end time, or by choosing a navigation session. The third, optional step, is to only take one or more Web sites into account. This option is particularly useful if we

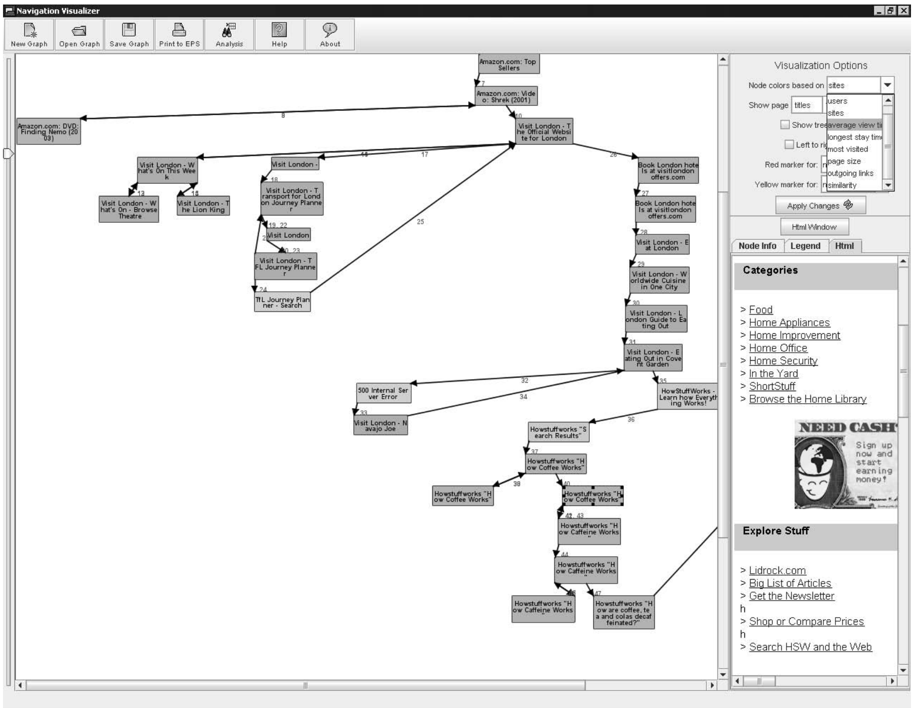

Figure 4.5: *The main window of the Navigation Visualizer*

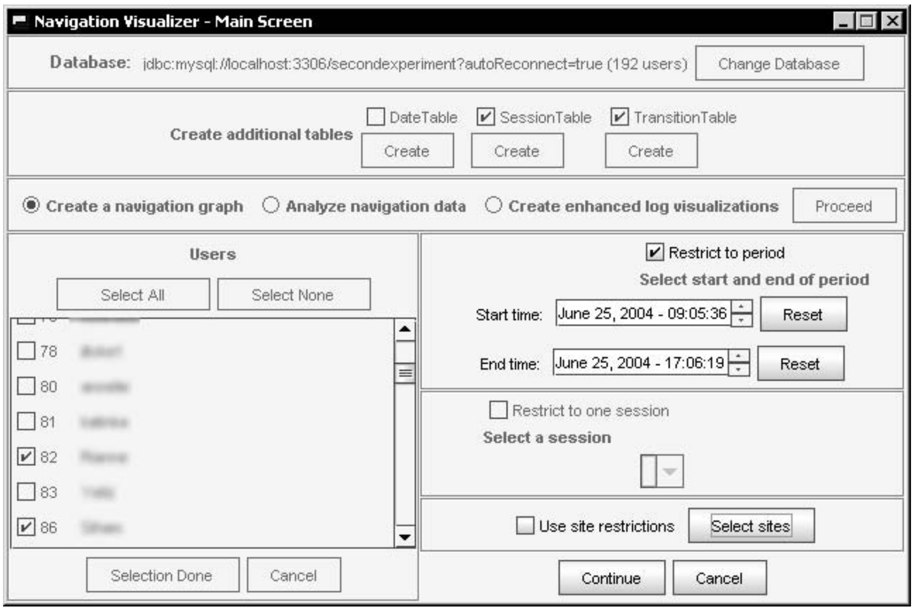

Figure 4.6: *Data selection. Names in the user dialog box are blurred for privacy reasons*

want to see how people navigate through, for example, a course Web site on a given day, without being bothered by intermixed spare time activities, such as checking news sites.

Once the data selection is completed, the navigation graph is generated. As can be seen in figure 4.5, the largest part of the window is occupied by the graph itself. Each page is represented as a filled rectangle, with directed edges representing the page transitions. From the edge labels the temporal order of the navigation path can be retraced, as the numbers indicate the order of navigation. Upon clicking a page, details are presented in the bottom right part of the screen. This is either a list of descriptive statistics or a preview of the page, which can be displayed in a separate window on request. To support visual inspection, global measures of the navigation graph can be displayed in a separate window as well.

The graph can be manipulated using the tools on the top right part of the screen. The most important manipulation is the selection of color binding and additional markers. Switching between colors is a powerful way to distinguish between dimensions such as user characteristics, page access statistics and textual similarity. Currently, the following color codings are available:

- each user is represented by a different, randomly generated color; pages that are visited by more than one user, are given a gray shade - the lighter, the more users have visited the page;
- each site is represented by a different, randomly generated color;
- the color indicates the relative average view time, which varies from blue (short view time) to red (long view time).
- the color indicates the view time of the longest visit, from blue to red;
- the color represents the number of visits paid to the page, from blue to red; this is somehow redundant with the incoming or outgoing edges, but useful in larger graphs;
- the color represents the page size, in terms of number of words, from blue to red;
- the color represents the number of outgoing links, from blue to red - the number of outgoing links is *not* the number of links actually followed from this page;
- the color represents the textual similarity to a given task description, as determined by the cosine similarity (formula 4.6).

In addition to the color codings, markers can be used to facilitate comparison between multiple dimensions. Markers are visualized as differently colored borders around the nodes that form the top n or bottom n according to the selection criterium. Red and yellow markers are available; pages that satisfy both marker conditions are marked with orange.

As neither the page title nor the Web address is sufficient for easily distinguishing between pages (Cockburn & McKenzie 2001), one can switch between the two. Alternatively, one can decide to only show the title or address on marked pages, or not at all. This may be needed to create more screen space, if one is only interested in the structure of longer navigation paths.

The layout of the graph is based on a spanning tree, which is built using the procedure described in the previous section. By default, if two links are followed from a page, both *children* are centralized below the parent page. A more space-consuming alternative is to use a strict left-to-right ordering, which better reflects the temporal order of the navigation path. If the default layout is not entirely satisfying, the pages can be manually relocated by dragging. As larger graphs with a high amount of backtracking have a high number of edges that are not part of the spanning tree, these edges can be hidden to avoid cluttering.

If needed, one can return to the data selection dialog and adjust restrictions on users, time window and Web sites. We have used this visualization tool extensively during the analysis phase of our laboratory studies, as described in the next chapter. In addition, the visualization was used for analysis of navigation paths created by students following a two-week summer course. The interactive selection of users, time windows and Web sites turned out to be essential for answering questions, such as on what course pages do students return to the help section or search for additional information.

Linear log visualization

Whereas the navigation graph is a suitable visualization for usability analysis, it is obviously too complicated to be used as a browser history tool. Also, the loss of temporal order makes it hard to retrace a navigation session step by step. As we wanted to discuss several sessions with the participants of our long-term study - described in chapter 7 - we created an alternative visualization that preserves the temporal order, but that still allows for distinguishing between sites, navigation actions, page view times and session borders.

An example visualization is depicted in figure 4.7. Each box represents a page visit. The page title, a clickable link and the date, time and session identifier of each visit are displayed next to the box. After a forward navigation action, such as link following, the next box is shifted to the right; after a backward navigation action, such as using the back button, the next box is shifted to the left. The color of the box is associated with the Web site the page belongs to, and the progress bar at the bottom of the box indicates the duration of the page visit. Additional information was mainly displayed for analysis purposes.

The linear log visualization has several advantages over a hierarchical list of page visits, such as the browser's history list. It provides a visual metaphor

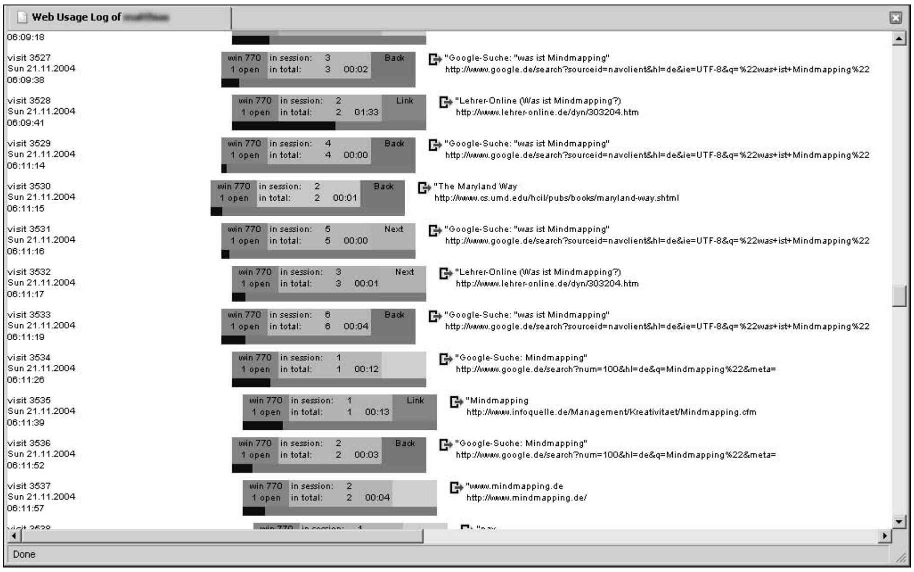

Figure 4.7: *Linear log file visualization. From the alternating left-right pattern it can be seen that an alternating strategy of forward and backward navigation was used.*

of forward and backtracking activities and it allows for visually distinguishing between sites. The strict temporal order makes it possible to retrace a session quickly step by step. During the interviews we conducted in our long-term study, several participants commented spontaneously that they would love to have such visualizations available to retrace their earlier activities, 'if only it wouldn't be so large'. For analysts it is beneficial to have both visualizations, which allows them to switch between pattern analysis using the graph-based visualization, and session retracing using the linear log visualization.

4.5.4 Summary

The Navigation Visualizer is a dedicated visualization tool for user navigation paths. The tool combines the mathematical, high level approach of Web usage mining with interactive graph-based visualizations and enhanced linear log visualizations. It facilitates tracing and understanding user actions, which would be harder to do with either one of the approaches. Furthermore, the Navigation Visualizer provides means for preprocessing complex user data for further analysis in statistical packages.

As the tool is constructed as a research instrument, it does not provide the flexibility of multi-purpose visualization toolkits, such as the prefuse system (Heer, Card & Landay 2005), which was introduced simultaneously with the Navigation Visualizer (Herder & Weinreich 2005). However, it does provide functionality for data filtering and bindings between attributes and visual elements not directly available in multi-purpose visualization tools. In addition, the dedicated interface provides focus on aspects important to user navigation. A next step in the development of the Navigation Visualizer would be the integration with an open source framework like prefuse, which provides more time-efficient rendering mechanisms and more flexible layout options. A further improvement would be to promote several data filtering options to data zooming options, for example in the form of a time window that can be adjusted and that directly affects the graph. Given its current limitations, the Navigation Visualizer already served its goal in our studies.

4.6 Concluding remarks

In this chapter we presented a graph-based approach to Web usage mining. The first part of this chapter dealt with the issue of acquiring, preprocessing and representing the data. In the second part we presented a graph-based framework for the analysis of hypermedia document graphs, and user navigation graphs in particular. Several measures for inferring the function, relevance and importance of pages for users were introduced. We continued with global measures on hypermedia structures and the structure of navigation paths. All these efforts were needed to come to the concept of *aggregated measures* on user navigation, which reflect user navigation strategies and patterns. In the last part of this chapter we showed a tool for visualizing these navigation paths, the Navigation Visualizer.

Due to the number of issues discussed in this chapter, it has become rather lengthy, even though many concepts are explained only briefly or just touched upon. Due to the inherent freedom of user navigation in non-linear hyperdocuments, and the differences in users and their tasks, a large variety of perspectives on the matter is needed to draw inferences. The reader's patience will be rewarded in the next two chapters, where we will come to the nuts and bolts, and apply the measures for analyzing actual navigation logs.

Chapter 5

Laboratory studies

5.1 Introduction

In this chapter we present results from two laboratory studies. Both studies aim to relate navigation styles to usability matters. In the first study we tried to discover individual navigation styles, and to find usability matters that are associated with these differences in navigation styles. For this purpose, we recruited a number of participants and collected user characteristics through questionnaires and tests, followed by a navigation session in which the participants were to solve several tasks in the field of personal finance. As can be read in the next section, we discovered two navigation styles, which we labeled *flimsy navigation* and *laborious navigation* that were highly correlated with the participants' perceived disorientation and Web site usability.

Given this background knowledge, we conducted a second study in which we wanted to find out to what extent adaptive hypermedia techniques would help in solving or reducing these usability matters. This study is described in section three. We decided to focus on one specific hypermedia technique - task-related link suggestions - to ensure that we did not merely test one system against one another, but evaluated the benefits of one particular technique. Whereas we tried to make the tasks as realistic and exploratory as possible, we fixed the user context by providing the participants a background scenario with some open issues to be solved. Results showed that users appreciated the link suggestions, that task execution times were reduced and that their navigation styles were more structured.

Both studies have been reported earlier in conference proceedings (Herder 2004) (Juvina & Herder 2005) [1]. In the last section of this chapter we further discuss the studies and position them in the 'tradition' of empirical studies in

[1]The second study won the James Chen Family Best Student Paper Award at the User Modeling 2005 conference.

the field of user modeling. We conclude with a number of recommendations for future user studies.

5.2 Discovery of individual navigation styles

Individual differences, ranging from gender differences through system experience to cognitive styles, significantly influence the way that people navigate through hypermedia systems (Chen & Macredie 2002). Many of these individual user characteristics can be gathered using questionnaires or standardized tests. However, for adaptive hypermedia systems this approach is often undesirable, as it requires time and effort from the users, which might eventually put them off. Moreover, not all user characteristics are stable or easily measurable: as an example, a user's motivation and concentration is most likely to change over time.

For this reason, it makes sense to provide users with adaptive navigation support based on users' navigation styles (Herder & Van Dijk 2003). With knowledge of the strategies that users follow, it is easier to recognize patterns in their navigation paths that indicate usability problems that need to be solved. A typical usability problem is that users become disoriented, or lost in a Web site (Thüring et al. 1995), which means that they are unable to keep track of their positions: at some point users might not know where they are, how they came there or where they can go to. Several characteristics of user navigation, most importantly those related to page revisits, have been related to success measures, such as task outcomes and user's perceived disorientation (Chen & Macredie 2002) (Herder & Van Dijk 2003) (McEneaney 2001).

In this section we present the results of a study that was aimed at finding patterns in user navigation that indicate a user's vulnerability to perceive disorientation while working on goal-directed tasks that require a fair amount of navigation to complete them. We were able to extract two navigation styles - which we called flimsy navigation and laborious navigation - that performed well in predicting the user's perceived disorientation. We start with describing shortly how individual differences influence user navigation. We continue with a discussion on navigation styles and measures for user navigation, which is followed by a presentation of the study and its results. We conclude with a discussion on the generalizability of the study and the implications for adaptive navigation support.

5.2.1 Individual differences in Web navigation

There is a vast amount of literature showing and analyzing individual differences involved in Web navigation. In (Eveland & Dunwoody 1998) it is noticed that novices tend to make use of a linear structure in hypermedia systems, when it is made available, while experts tend to navigate non-linearly. MacGregor (1999) demonstrated that students who had more *domain knowledge* displayed more purposeful navigation and allocated time more variably to different pages. *Spatial ability* is an important determinant of hypermedia navigation performance, as reported in several studies - e.g. (Chen 2000); users with low spatial abilities have difficulty in constructing and using a visual mental model of the information space. *Field-dependent* users tend to navigate through link-following and backtracking activities, whereas field-independent users employ more analytical navigation strategies, making use of tools for freely jumping through the hypermedia structure (Chen & Macredie 2002). Students with an internal *locus of control* are reported to be better able to structure their navigation and to take advantage of hypertext learning environments (MacGregor 1999).

Research on cognitive mechanisms involved in Web navigation gains increasing influence in the HCI community. A cognitive model of Web navigation should be able to simulate the navigation behavior of real users, producing the same navigation patterns as actual users would do. Many approaches to user navigation modeling are mostly inspired by the theory on information foraging (Pirolli & Card 1999). Information foraging theory assumes that people, when possible, will modify their strategies in order to maximize their information gain. More specifically, users continuously compare the benefits of alternative actions, for example digging further into one information resource versus looking for a different resource. Process models that are based on these theories can analyze or simulate users' actions in terms of their individual evaluations of their expected utility.

5.2.2 User navigation styles

A related line of research aims at directly modeling the user's navigation behavior in order to provide adaptive navigation support in Web applications (Herder & Van Dijk 2003). A dynamic user navigation model could include:

- *syntactic information* - e.g. which links are followed, what does the navigation graph look like, what is the time that users spent on each page;
- *semantic information* - i.e. what is the meaning of the information that the user encountered during navigation;
- *pragmatic information* - i.e. what is the user using the information for, what are the user's goals and tasks

In the following parts we focus on the syntactic information. Our aim is to identify patterns in user navigation that indicate problems associated with disorientation, as experienced by the user. In the following part we characterize several user navigation styles. In the second part we summarize several measures that can be used to capture these navigation styles, as dealt with in more detail in the previous chapter.

User navigation strategies

User navigation can range from goal-directed task completion to more unstructured browsing and exploration of the availability of information or services (Eveland & Dunwoody 1998). Routine browsing is an integral part of Web navigation, nowadays; typically, users have a small collection of favorite sites that they visit very frequently (Chen & Macredie 2002). Several taxonomies of Web browsing behavior are presented in the literature. One of the finer grained taxonomies is presented in (Rozanski et al. 2001), a white paper that is clearly targeted at the e-commerce community in which seven patterns are categorized, based on session length, average page view times and the amount of revisits during this session.

Within a navigation session, users often return to pages that serve as navigational hubs. Extensive use of these hubs is reported to be an effective navigation strategy (McEneaney 2001). When looking for information, users often employ search strategies that are quite similar to graph searching algorithms, such as depth first, breadth first and heuristic search (Berendt & Bernstein 2001). With knowledge of the type of session that users are involved in, and the navigation styles that they employ during these sessions, it is possible to compare navigation patterns and to recognize patterns that might indicate usability problems. In this study we let users work on a set of predefined tasks in a restricted part of the Web in order to find out what these patterns look like.

Measures of user navigation

As discussed in detail in the last chapter, user navigation paths can be modeled as graphs, with the vertices representing the pages visited and the edges representing the links followed (Herder & Van Dijk 2003). Several - mostly graph-theoretic and statistical - methods can be used for analyzing this structure. Typical measures include the total number of pages visited to solve a task, the total time needed to solve a task and the average times spent on single pages (Berendt & Bernstein 2001). Within the navigation paths, patterns may exist that indicate a user navigation style or problems encountered. In our study we made use of a collection of navigation measures that together describe these patterns. The measures relevant to the results of the study are summarized below.

Although disorientation and the problem of 'lostness in hyperspace' are considered as an important issue, very few attempts have been made to quantify these issues (Otter & Johnson 2000). Moreover, these few attempts have led to contradictory results. Pauline Smith (Smith 1996) takes as a starting point for her lostness measure the assertion that lostness should be viewed in terms of degradation of user performance. She proposes a lostness rating, based on the relative amount of revisits while searching, and on the number of navigation actions compared to the minimum number required. In other words, any form of revisitation is used as a piece of evidence for lostness. John McEneaney (McEneaney 2001), on the other hand, concludes that users that employ shallow, hierarchical search strategies are more successful in their search than those that followed more linear paths. In other words, revisits to navigational landmarks can be seen as a sign that users have formed an accurate model of a Web site.

Both studies report that their lostness measures are validated by user studies, in both cases making use of a system that provides teenagers with study advices. Although the results provide evidence that measures of user navigation can be used to predict disorientation, it appears that simple revisit rates are insufficient for capturing a complex issue as disorientation. By capturing various aspects of page revisitation, we aim to find *revisitation patterns* rather than the amount of revisitation. The following measures were taken into account:

- the *path length* is the number of pages that the user has requested during a navigation session, including page requests that involved revisits;
- the *relative amount of revisits* is calculated as the probability that any URL visited is a repeat of a previous visit. We adopted the formula that is suggested by (Tauscher & Greenberg 1997);
- the *page return rate* indicates the average number of times that a page will be revisited. The return rate is calculated by averaging the number of visits to all pages that have been visited at least twice. A more extensive use of navigation landmarks will most likely lead to a limited set of pages that is visited very frequently;
- *back button usage* indicates the percentage of back button clicks among the navigation actions, including backtracking multiple pages at once using the back button;
- *relative amount of home page visits* is a self-descriptive label. 'Relative' refers to a correction of home page visits based on path length.

The average time that users spend at Web pages is reported to be an important indicator for user interest and human factors (Shahabi et al. 1997). Besides the average view time, the *median view time* was also taken into account, as users generally spend only little time on the large majority of pages before selecting a link (Broder et al. 2000). The median view time is not affected by the few 'high content' pages that were inspected more carefully, and thus provides a better indicator for the average view time while browsing.

Navigation complexity can be defined as 'any form of navigation that is not strictly linear'. Complexity measures are mostly derived from graph theory and used frequently for assessing hypertext and its usage (Herder & Van Dijk 2003). Typical measures reflect the cyclical structure of the navigation graph and the length of navigation sequences within the graph. Several commonly complexity measures, as introduced in the previous chapter, were taken into account:

- the *number of links followed* per page ('fan degree') (Rauterberg 1992) represents the ratio between the number of links followed and the number of distinct pages visited;
- the *number of cycles* (Rauterberg 1992) is calculated as the difference between the number of links followed and the number of pages visited. As the number of cycles grows with the length of the navigation path, it can only be used for a fixed time window;
- the *path density* (Rauterberg 1992) compares the navigation graph to the corresponding fully connected graph. A higher path density indicates that a user makes use of short navigation sequences and regularly returns to pages visited before;
- *compactness* (McEneaney 2001) is a measure similar to path density. It indicates that users follow a 'shallow' search strategy. In contrast to the path density, it compares the average distance between any two pages in the navigation graphs to a theoretical minimum and maximum;
- the *navigation stratum* (McEneaney 2001) is a measure designed to capture the linearity of user navigation; a lower stratum indicates less linear navigation, as the 'hierarchical' distance between any two nodes is quite low;
- the *average connected distance* (Broder et al. 2000) indicates the average length of a path between any two connected pages in the navigation graph. A higher average connected distance indicates that users do not return to a page very soon, but only after having browsed for a while. They also return using a link rather than using the back button. In short, the average connected distance measures the users' confidence in that they 'will find their way back later'.

In figure 5.1 some of the complexity measures are exemplified.

The navigation measures that are described above are labeled first-order measures in this study, because they are derived directly from the raw data, without taking into account that the measures might be correlated, which most likely would be the case. As an example, the average connected distance is calculated independently of back button usage, without taking into consideration the fact that usually low values on the former measure are associated with high values on the latter and vice-versa. This aspect was dealt with by calculating second-order measures - or navigation styles, as will be explained in the next section.

5.2.3 Experimental setup

In our study we were interested in what navigation styles occur when users perceive disorientation when performing several goal-oriented tasks. In order to

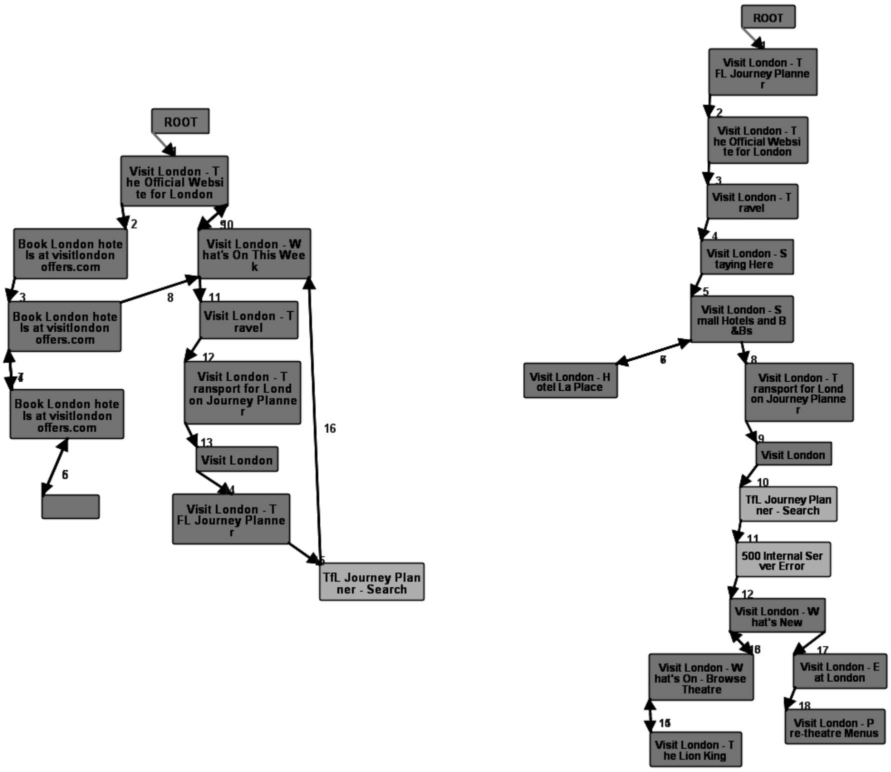

Figure 5.1: *Two visualizations of user navigation paths with highly different values for compactness and stratum. On the left a path with high compactness (0.80) and low stratum (0.38). On the right, a path with low compatness (0.45) and high stratum (0.86). As expected, the back button usage in the left picture is lower (12%) than in the right picture (16%). Due to the small size of this example, no difference in average connected distance can be observed.*

better interpret the outcomes, we also collected several user characteristics that are likely to differ between individuals, as well as users' evaluation of their navigation activities. The experimental setup and the results will be discussed in this section.

Participants and procedure

The study consisted of individual sessions with thirty subjects, all undergraduate and graduate students from two Dutch universities in the age range 19-28, with an average age of 21.5. Participants were selected randomly out of the student lists of both universities, while making sure that males and females were equally presented.

Each individual session lasted 2.5 hours. In the first 1.5 hour we collected data on the participants' user characteristics. The actual navigation tasks took forty minutes. The subjects were asked to answer questions by browsing through different Web sites on a laptop computer. Two of these Web sites provided information and tools in the field of personal finance; the third Web site was an online store. Some questions were meant to invite open-ended browsing, other questions were more goal-directed. A short evaluation survey was given after the navigation session.

User characteristics

Several user characteristics were collected in the first stage. The characteristics that are relevant in the context of this paper are briefly described below.

Spatial ability, *episodic memory* and *working memory* were measured with computerized cognitive tests provided by the Dutch institute TNO Human Factors. The *spatial ability test* uses the classical mental rotation task, and the spatial ability score is the number of correct solutions obtained by rotating three-dimensional objects. The *episodic memory test* presents three lists of sixty images each and participants must loudly name the images in the first two lists; after the second list a distraction task is given - we used the spatial ability test as a distraction task, to efficiently use the testing time - after which a third list is presented which contains a mix of images presented before in list one and two, and new images. The participants have to recognize the images that were presented in the first list. The *working memory test* uses a reading span task (Linderholm 2002), which presents participants with series of phrases that are to be read aloud. After each series, the participants are asked to recall the last word of each phrase in that particular series; to ensure that the participants treat the whole content not only the last missing words, for one random phrase from the series two missing words are asked to fill in. This test is more complete and more adequate than digit span tests for working memory capacity, since it takes into consideration not only information storage, but also information processing that is normally associated with working memory capacity.

Locus of control refers to the users' belief in how they contributed to their own success or failure. Research shows that users with an internal locus of control are better able to structure their navigation and take advantage of hyperspace features (MacGregor 1999). We measured locus of control with a standardized questionnaire (Pettijohn 2002). *Expertise* was divided in *internet expertise* and *financial expertise* (domain expertise). Both types of expertise were measured using self reported frequency of use and self-assessed level of knowledge and skills. A measure of users' affective disposition at the beginning of the navigation session was built based on users ratings of different affective states that they considered appropriate to describe their disposition. Subsequently, user ratings were factor analyzed and grouped into three basic moods: *active*, *enthusiastic* and *irritable.*

Navigation tasks

In the navigation session, subjects were asked to perform various tasks in the field of Web-assisted personal finance. This field includes using the Web to keep a personal budget, to perform financial transactions and decide to save or invest money. The tasks were designed in such a way that it would require a fair amount of navigation to answer or to solve them (Juvina & Van Oostendorp 2004). Three Web sites were used in this study, two of which are dedicated to personal finance. These sites provide users with advice and tools - such as planners, calculators and educators - to deal with their financial issues. The third site, an online store, was used as a reference. Some questions were meant to invite open-ended browsing, other questions were more goal-directed. A number of example questions are listed below.

- What are the definitions of personal budget and financial goal as they are presented within this Web site.
- Calculate how much the real value of 7,826 pounds will decrease in five years given an annual inflation of 3%.
- Suppose you are given 75 dollars to buy something for yourself. Find some products that you like in the section 'books, music and dvd'.

The participants had thirty minutes in total to solve the tasks. It should be mentioned that the two financial websites used were quite different both in size and structure. The first site was fairly small, with the look and feel of an office application, with pull-down menus, interactive calculators and short areas. The second site was much larger and more densely linked. Its lengthy pages were more similar to a news site, and its overall appearance was less structured than the first site.

After having solved the financial tasks, or when thirty minutes had passed, the participants were asked to fill a shopping basket with a collection of books, music or dvd's that they liked. This less goal-directed task was used as a reference frame to compare the more restricted task-based navigation behavior with their regular Web browsing patterns.

Evaluation survey

Task performance was recorded by the participants on a dedicated form and coded afterward by the experimenter. After the navigation session, the subjects were asked to evaluate their satisfaction with task completion and the usability of the different Web sites used. A standardized survey on perceived lostness (Ahuja & Webster 2001) was also included in the evaluation session. The participants were asked to indicate whether they agreed or disagreed with the statements on a Likert scale of 1 to 7. Some example statements are:

- it was difficult to find my position after navigating for a while
- quite often I unexpectedly returned to a page I have visited before
- the overall structure of the site is clear and easy to work with (reverse-coded) Our measure of disorientation was the sum of user ratings of these statements.

5.2.4 Results

As expected, a large amount of user navigation actions were page revisits. The revisit percentage of 38% that we recorded is lower than the 58% as reported by Tauscher and Greenberg (Tauscher & Greenberg 1997). However, this can be explained by the fact that users continue to add pages into their 'revisitation repertoire'. One forty-minute session is simply too short to generate such a broad repertoire. We found a significant difference between revisitation percentages while performing tasks on the personal finance Web sites (40.7%) and while browsing for interesting items in the online store (25.2%) - t=5.670, $p < 0.01$. This indicates that backtracking activities are more common for goal-directed tasks than for free exploration. Although the amount of revisitation was similar between the personal finance sites, the usage of the back button was not - 3.3% and 15.9% respectively, and 7.75% for the online store. Despite the large difference in back button usage between the personal finance sites, the general tendencies were similar (r=0.386; $p < 0.05$). These differences are most likely due to differences in site structure or site design, as the personal finance site that invited the lowest back button usage is small compared to the other site and its menus are clearer defined.

The return rate (the number of times that a page will be revisited) was about the same for all sites (3.1), as well as the average connected distance (3.6). Apparently these aspects of revisitation behavior are not influenced by the site designs; users recognize and find their way to navigational landmarks anyway - with more or with less ease.

Most subjects did not complete all tasks successfully. In general, they found correct answers for only two to four out of six questions. Therefore, it is not surprising that the subjects felt slightly disoriented. However, this did only influence their satisfaction about their performance to a certain extent (r=-0.310; $p < 0.01$). Neither did disorientation automatically lead to frustration.

The *return rate* proved to be a good indicator of disorientation (r=-0.417; $p < 0.022$). This provides evidence for McEneaney's (McEneaney 2001) observation that users that are better able to make use of the site structure - and of navigational landmarks - are more successful in navigation. The percentage of revisits did not correlate with perceived disorientation. Apparently, revisitation is a natural navigation behavior that does not necessarily indicate that a user is disoriented. The *median view time* is correlated with perceived disorientation, although not very strongly (r=0.365; $p < 0.05$). Apparently, users who feel disoriented spend more time on pages while browsing. The average connected distance of those people with a higher median view time becomes lower (r=-0.412; $p < 0.05$) - which indicates shallower backtracking. The *average connected distance* is correlated with some cognitive factors as well. Subjects with a higher working memory score tended to extend their revisitations (r=0.375; $p < 0.05$), as well as subjects that indicated to be in an active mood (r=0.413; $p < 0.05$).

Navigation styles

As it is most likely that patterns in the first order navigation measures occur simultaneously, second order navigation measures - linear combinations of the first order measures - were calculated. Principal component analysis with equamax rotation on twenty-two navigation measures resulted in four factors that together explained 86% of the variance. Factor loadings in first-order measures and the correlations between the factors and user characteristics and task outcomes were used to interpret the content of each factor in terms of navigation styles as follows.

The first factor, which we called *flimsy navigation*, is characterized by slow navigation, mainly focused around the site's home page. Time is spent on reading content rather than exploring the hypermedia structure. The second factor, *content focus* groups together all the view-time measures. Users that score high on this factor stay relatively long on pages, supposedly to read the contents. This style is not associated with any user characteristics or task outcomes. The third factor, *laborious navigation* is characterized by rapid and frequent backtracking. The fourth factor, *divergent navigation* groups together the participants that were reluctant to revisit pages, as expressed by the navigation paths' low compactness, high stratum, low homepage use and high average connected distance. This navigation style is only associated with a high propensity to trust.

We will focus on the first and third factor, which account for 27% and 23%, respectively, after rotation. We labeled them *flimsy navigation* and *laborious navigation*, based on their correlations with first-order measures and user characteristics. It should be noted that these styles do not exclude one another. All correlations mentioned are significant with $p < 0.05$.

High scores on the flimsy navigation style are associated with:

- small number of pages visited (r=-0.80)
- high path density (r=0.80)
- high median view time (r=0.77)
- short average connected distance (r=-0.70)
- low number of cycles (r=-0.53)

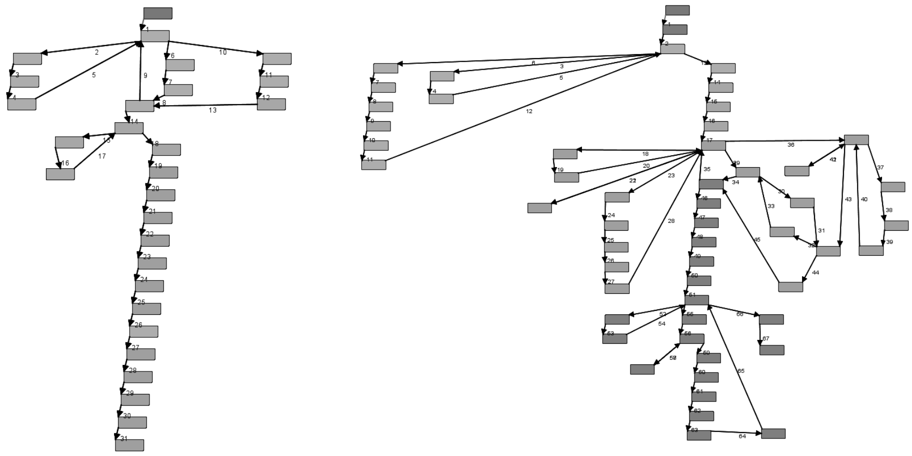

Figure 5.2: *Flimsy (left) versus sturdy (right) navigation. From the figure it can be observed that flimsy navigation is characterized by short navigation paths and a low number of cycles in the navigation graph. The page revisits that did take place in the flimsy navigation path were made using the back button.*

- high rate of home page visiting (r=0.48)
- high frequency of back button use (r=0.39).

Flimsy navigation appeared to be a weak navigation style. Most of the navigation takes place around the site's home page and users regularly return to their starting points. Time is mostly spent on processing content instead of actively locating information. The short average connected distance indicates that users return to a page very soon. Users also prefer to return by using the back button instead of by following links. The low number of cycles indicates that users employing this navigation style do not make extensive use of the means for revisitation available within the sites. High scores on the flimsy navigation style are associated with low scores on Internet expertise, current active mood, working memory and locus of control. Based on these correlations, it is likely that flimsy navigation is mostly employed by inexperienced users who are not able or not inclined to reconstruct their past actions; rather, they continue along the same path or eventually start over again. For these reasons, we might expect that flimsy navigation is related to users' perceived disorientation.

High scores on the laborious navigation style are associated with:

- high number of links followed per page (r=0.95)
- high revisitation rate (r=0.94)
- high number of cycles (r=0.79)
- high return rate (r=0.73)

- high frequency of back button use (r=0.71)
- high path density (r=0.43)
- high number of pages visited (r=0.40)
- short average connected distance (r=-0.39).

This navigation style involves intensive exploration of navigational infrastructure provided by the site. Users seem to employ a trial and error strategy; they follow links merely to see if they are useful or not. They figure out quite fast when paths are not leading toward their goal and return. Revisits are numerous but not redundant: once a page is revisited a different link is followed than before, which constitutes another trial. This behavior is particularly observed on navigational hubs, such as menus and index pages.

High scores on laborious navigation are associated with high episodic memory, and low spatial ability. This style indicates a revisitation pattern that does not lead to disorientation; instead, laborious navigation appears to help users in constructing a conceptual overview of the site structure and then to make use of this model.

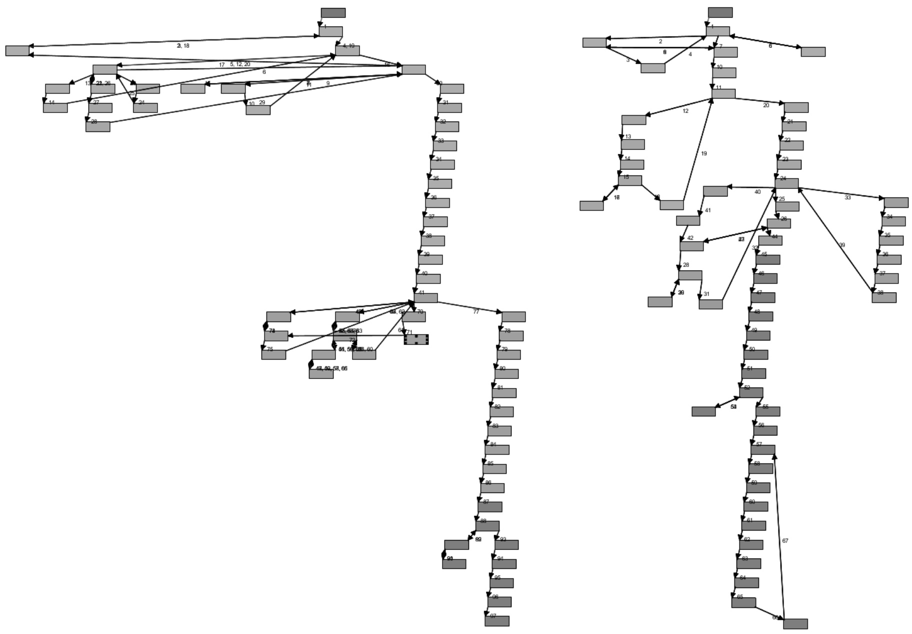

Figure 5.3: *Laborious (left) versus non-laborious (right navigation. From the figure it can be observed that laborious navigation is characterized by periods of extensive exploration - the two clusters in the left graph - followed by a more linear path. This pattern suggests that laborious navigators first build a mental model of the information domain before working on the task. Note that the revisit patterns in the right graph (non-laborious navigation) are less focused.*

Prediction of perceived disorientation

Multiple linear regression analysis was used to find out which navigation measures and navigation styles performed best in predicting the subjects' perceived disorientation. Including predictors in regression models was based on the stepwise method; the predictive power must be seen as the best one can get with the minimum number of predictors. It turned out that the flimsy and laborious navigation styles together best predicted the user's perceived disorientation ($R^2 = 0.29$) with a large effect size ($ES^2 = \frac{0.29}{0.71} = 0.41$) [2].

	B	β	t	σ
Constant	40.1		29.66	0.000
Flimsy navigation	3.92	0.46	2.85	0.008
Laborious navigation	-2.38	-0.28	-1.73	0.095

Table 5.1: *Prediction of perceived disorientation based on navigation styles. The regression model consists of perceived disorientation as dependent variable and flimsy navigation and laborious navigation as predictors. From the regression coefficients (B) one can observe the positive and negative correlations of flimsy and laborious navigation respectively with perceived disorientation. The standardized coefficients (Beta) show a larger relative importance of flimsy navigation as compared with laborious navigation.*

5.2.5 Discussion

The results of our study suggest that users' vulnerability to experience disorientation in large Web sites can be automatically diagnosed with an attractive level of accuracy. We identified two navigation styles, flimsy navigation and laborious navigation, which proved to be significant predictors with a large effect size.

The first-order measures that constitute the navigation styles confirm the assumption that patterns in revisitation are an important indicator for perceived lostness (Smith 1996) (McEneaney 2001). It has become apparent that in particular those measures that directly relate to known effects of users becoming disoriented (Otter & Johnson 2000) are valuable; flimsy navigation is characterized by slow navigation, unstructured backtracking and frequent returns to the site's home page. Laborious navigation indicates that users initially spend some time on building a mental model of the site structure; having a more complete mental model prevents users from getting lost.

The area in which these navigation styles have been identified, is rather limited: they apply to situations where goal-directed and performance-oriented tasks are performed on the Web. The domain of Web assisted personal finance might seem narrow and this is why we used three different Web sites and a relatively

[2]The effect size for regression is calculated with the following formula: $ES^2 = R^2/(1 - R^2)$. 0.02 is considered a small effect, 0.15 a medium effect and 0.35 a large effect size

complex and heterogeneous range of task. By choosing three different websites to be used in the pilot study, we attempted at randomizing factors pertaining to a specific site structure or interface design. Tasks were not only aiming at locating information but also at using this information to solve actual problems. These decisions were intended to constitute premises for ecological validity and generalizability of the results.

The number of subjects (thirty) was rather limited and relatively homogeneous, as they were students. New data is necessary to find out in what situations the identified navigation styles are relevant for predicting disorientation. Most likely, other styles will be identified as well that can explain other facets of disorientation.

Implications for adaptive navigation support

Prediction of users experiencing disorientation that is based on navigation measures has important practical consequences. From a usability point of view it is useful to identify those users who are at risk of experiencing disorientation and to assist them by adequate, and possibly personalized, navigation support.

Context information is important for effective navigation, as each navigation process is inextricably tied to the structure of the site. Two types of user context can be distinguished: the structural context and the temporal context (Park & Kim 2000). *Structural navigation aids* - such as site maps, menus and index pages - describe a user's current location and navigation options; *temporal navigation aids* - such as the browser's back button, bookmarks and visual navigation histories - describe the way that led to this position.

Users that navigate in a flimsy manner appear not to be able to reconstruct their navigation paths and therefore are prone to get stuck. The main cause of these problems is that they employ a strategy of link following and backtracking, which highly depends on the context information provided by links. In particular associative links, which are embedded in the running text, are known to be poor indicators of what is behind the link. Adaptations that are focused on link annotation should be considered for this category of users, to provide structure to their navigation strategy. Task-related link suggestions are expected to support forward navigation, as they improve the information scent and construct relevant trails to follow. Flimsy navigation also involves step-by-step backtracking, making use of the back button, and not making use of navigational hubs. Approaches such as the 'smart' back button (Milic-Frayling, Jones, Rodden, Smyth, Blackwell & Sommerer 2004) that recognizes and facilitates access to these hubs may help in structuring backward navigation. In addition, visual navigation histories help users in keeping track of where they already have been, which prevents occasions in which they accidentally return to pages visited before.

Laborious navigation is a strategy for building a mental model of the site structure. This strategy is particular useful if users are able to remember their recent navigation actions and compensates for a lower spatial ability. However, episodes of laborious *orienteering behavior* (Teevan et al. 2004) are time-consuming and users may be reluctant to spend effort on these seemingly inefficient activities. One approach to help users in recognizing the site structure is to improve and focus on contextual navigation support - such as site maps - and to show the user's current position and past positions in the structure. An alternative approach, which may be more effective - in particular when the user is expected to use the site for a longer period or to return to the site more often -, is to invite and support initial exploration by introductory tours along the most important landmarks.

Future perspectives

In this section we discussed how to address two navigation styles that might indicate, or that might lead to users getting disoriented in Web sites while working on goal-oriented tasks. The add-on navigation support, as discussed in the previous subsection, aims at improving the way users navigate rather than at forcing users to passively follow some ready-made paths. We believe that this should be the goal of adaptive hypermedia systems in general. Whereas the results of this study might be applicable only in the small domain of Web-assisted personal finance, the prospect of adaptive navigation support that fits the user's navigation style is attractive. As an example, users that prefer to extensively explore the sites that they visit, should be supported in doing so, instead of being urged to leave for a different site, unless the system is capable of making clear to the users that the benefit is higher than the cost of altering their strategy.

In order to investigate the potential benefits of add-on navigation support, we conducted a follow-up study in which we compared navigation paths of participants provided with link suggestions with the navigation paths of participants who were not provided with these suggestions. This follow-up study is described in the next section.

5.3 The impact of link suggestions

In contrast to most desktop applications, Web sites generally are designed for a general audience with varying goals (Shneiderman 1997). As it is hard to satisfy all categories of users with one design, adaptive hypermedia systems try to better support the users by personalizing content or link structure. Traditional techniques in the latter category involve link hiding, sorting, annotation, direct guidance and hypertext map adaptation (Brusilovsky 2001). When trying to find information related to a task, users have to rely on proximal cues such as the link anchor text to decide what their next action will be (Olston & Chi 2003). If the proximal cues are not clear enough, or if the users do not have sufficient insight in a site's structure, they might become disoriented: they don't know their current position in a Web site, how they came to that point or where to go next (Thüring et al. 1995). Various studies have been carried out to infer user goals from their actions, for example (Chi, Rosien, Supattanasiri, Williams, Royer, Chow, Robles, Dalal, Chen & Cousins 2003). Given these goals, the utility of the various navigation options on a Web page can be estimated (Kitajima et al. 2000) (Pirolli & Fu 2003) and communicated to the user by means of link relevancy indicators, or link suggestions.

While user-adaptive systems appear to be a good idea, it still is an open issue how the benefits from adaptations can be evaluated and compared to one another. Whereas the benefits of adaptive systems as a whole have been evaluated, little is known on the impact of individual adaptive hypermedia techniques, independent of domain and site structure (Weibelzahl 2003). It is common belief in the field of adaptive hypermedia that personalization techniques do have impact, yet no data is present on how user behavior changes as a result of the personalization. Given a suggested link, will it be followed 'blindly', will it be ignored, or will it be used as additional context. In order to find answers to these kinds of questions, we conducted a user study in which subjects were asked to carry out several predefined everyday tasks. In one condition, the subjects were provided with predefined link suggestions. Various indicators of user's behavior and perception were measured. We found evidence that link suggestions based on the user's goals have a positive impact: they cause the users to navigate in a more structured way, which makes them less vulnerable to perceive problems associated with disorientation (Herder & Juvina 2004).

The remainder of this section is structured as follows. In the next section we present our research questions. We continue with the setup of the experiment. After presenting the results, we conclude with a short discussion.

5.3.1 Individual differences, disorientation and navigation styles

There is a vast amount of literature on individual differences in Web navigation. In our previous study we found spatial ability and domain expertise to be the most important determinants of user performance in Web tasks (Juvina & Van Oostendorp 2004). It has been shown that there are differences in favor of men with regard to spatial ability (Sjolinder 1998), Web searching behavior, and learning performance (Roy & Chi 2003). Women are more likely to use a rote way-finding strategy - attending to instructions on how to get from place to place - whereas men are more likely to report to use an orientation strategy - maintaining a sense of their own position in relation to environmental reference points (Sjolinder 1998). For this reason, we expect women to benefit more than men from navigation support that supports rote-finding, thus compensating for the so-called 'gender gap' in Web use.

As mentioned in the introduction, disorientation is a major issue in Web navigation that is mainly caused by the non-linearity of Web sites; on each page, users have to decide between alternative options, which includes following links or backtracking to pages visited earlier. Although the problem has been given the label 'disorientation', it is hard to measure or to quantify. (Ahuja & Webster 2001) developed a questionnaire that is shown to indicate a user's perceived disorientation. Various attempts have been made to relate patterns in user navigation, most importantly patterns related to page revisits, to success measures and disorientation (Herder & Van Dijk 2003) (McEneaney 2001) (Smith 1996). In our previous study (Herder & Juvina 2004) we found a weak navigation style that was associated with perceived disorientation.

Based on these previous findings, we formulated the following hypotheses to guide the study:

- Link suggestions will generally be well received.
- Link suggestions improve perceived usability and reduce disorientation, as experienced by the users.
- Link suggestions will influence the way users navigate; the differences can be interpreted as an argument in favor of providing support.
- Women will benefit more from links suggestions than men.

5.3.2 Experimental setup

In order to check for differences in users perceptions and navigation behavior, as a result of link suggestions, an experimental approach was employed, which is described in this section.

Web navigation tasks

First matter of concern for the experimental setup was triggering realistic Web navigation behavior. Several participants of the previous study indicated that they could not completely identify themselves with the financially oriented tasks they were asked to solve. Based on this input, we sought for more everyday navigation tasks, with still underlying subtasks to be solved.

Five Web navigation tasks were created based on the collection of cases presented in (Morrison, Pirolli & Card 2000), following suggestions from (Kitajima et al. 2000). We encapsulated these tasks in the form of scenarios, of which an example is given below.

> This summer you will spend a long weekend in London with your girl/boy friend. Both of you would like to visit the top attractions and some museums. But, most importantly, you want to visit one of the great musicals in West End.
>
> Given facts:
>
> - You already have plane tickets to Heathrow Airport
> - You still have to book a hotel, preferably near the West End theatre district
> - You have sufficient money to spend during the weekend
>
> To do:
>
> - Go to http://www.visitlondon.com and find answers to the following questions:
> - Find a small hotel in the West End district
> - How do you get from Heathrow Airport to the city center?
> - In what theatre does the Lion King play?

Participants were instructed to start each task at a specified website's home page. They were allowed to use other websites than the indicated ones; the only restriction was to start at the specified websites. All of the scenarios have been assessed by test users and improved based on their comments, to ensure that the task descriptions were clear and unambiguous. The remaining four scenarios are summarized below:

- You want to buy or download a DVD movie and you are not sure yet which one of two particular movies you would prefer. You would like to decide based on how popular and how funny the movies are.
- You are searching in the market for a new digital photo camera. You want to check models and compare prices and also see some authoritative reviews of the available products.
- You have to write an essay for your science class about the use of coffee in daily life.
- Your grandfather appears to be suffering from memory loss. [...] You want to know whether these problems are part of normal aging, or signs of beginning dementia.

The order of the scenarios was randomized for each participant, to avoid measuring interactions between the tasks.

Experimental manipulation

Various strategies for generating link suggestions can be thought of. The simplest case is when the content providers indicate the most important sections of the content from their point of view. As users may have different goals for visiting a site, it might be a better idea to provide link suggestions that match the current context of use. In this experiment we provided users with suggestions that are relevant to their tasks.

Suggestions were generated based on simulations of a cognitive model similar in principle to CoLiDeS, a cognitive model of Web navigation presented in (Kitajima et al. 2000). CoLiDeS uses Latent Semantic Analysis (LSA) to calculate the semantic similarity between goal description and the available links on the current webpage. The link most similar in description to the task goal is selected to be clicked on. We used an augmented model, CoLiDeS+ (Juvina, Van Oostendorp, Karbor & Pauw 2005), which assumes that users do not only base their selections on relevance of incoming information, but also on whether a candidate selection is consistent with past selections or not. The main intuition behind the model is that users employ a hill-climbing strategy toward pages that are more similar to their current goal. On a given page, the semantic similarity (information scent) of each link text to the current goal is calculated. In addition, the semantic similarity of the links followed thus far plus the available links is calculated, the *path adequacy*. The link with the highest similarity or path adequacy is selected. If none of the links is better than the current path adequacy, the model backtracks to an earlier point. User behavior was simulated in advance based on the task descriptions. Semantic similarities between these task descriptions and the texts of the links leading to task solutions were calculated with LSA.

For each task, one or more successful paths to predefined goal pages - which contained any of the information asked in the task description - were generated. In the navigation support condition, links on these paths were highlighted to the participants - see figure 5.4. In the control condition, participants executed the same tasks without any support. The link suggestions were generated on the fly using the Scone framework for development and evaluation of Web enhancements (Weinreich et al. 2003).

Participants in the support condition were instructed that suggestions were automatically generated by a cognitive robot, which were meant to help participants in doing their tasks, and they could be followed or not. Participants got suggestions only when they arrived at specific pages.

Participants

Thirty-two participants, mainly students of various studies at Utrecht University, were recruited with advertisements. To qualify for participating, a minimum level of English language skills and Internet experience was required. The participants

Figure 5.4: *Example link suggestion: the two arrows point at the suggested link 'Good to the last drop'.*

were randomly assigned to one of the two conditions - sixteen participants in each condition.

Measures of user navigation and user perceptions

Several measures on navigation complexity and patterns of page revisits were calculated. For matters of brevity we limit ourselves to describing the most relevant measures in the context of this study and refer to chapter 4 for a more complete overview. The meaning of these measures is shortly summarized below.

- *back button usage* is the percentage of back button clicks among the navigation actions;
- the *relative amount of home page visits* is the number of visits to the Web pages that the participants used to start the different tasks, divided by the total number of page visits;
- *compactness* (McEneaney 2001) indicates that users follow a 'shallow' search strategy;
- the *navigation stratum* (McEneaney 2001) is a measure designed to capture the linearity of user navigation;
- the *average connected distance* indicates the average distance between any two pages in a navigation path. In short, it indicates how confident users are that they 'will find their way back later' (Herder & Juvina 2004).

A post-navigation questionnaire was used to measure user opinions on usability of the websites used and the users' perceived disorientation (Ahuja & Webster

2001). For each item of the questionnaire a 5-point Likert scale from 'strongly disagree' to 'strongly agree' was used. The 16 participants in the support condition were given four additional items on how they perceived the provided suggestions:

- The suggestions given by the robot were helpful
- I felt the suggestions were intrusive / annoying
- I believed I could trust the suggestions given by the robot
- I felt being manipulated by the given suggestions

The variable gender was added as an independent variable in the analysis phase to check whether it interacts with the fixed factor (support). The duration of each session was 55 minutes, of which 40 minutes were spent on carrying out the navigation tasks.

5.3.3 Results

In this section we present the results of the study described above. We start with the participants' opinions on link suggestions. Then we describe the impact of link suggestions on user perceptions and task execution time. We continue with the influence of link suggestions on user navigation behavior. We conclude with a brief look at gender differences.

Link suggestions are positively received

Table 5.2 shows the number of participants expressing their agreement or disagreement with each of the four questionnaire items concerning the way suggestions are perceived. It can be observed that most participants (13) do not perceive suggestions as intrusive, annoying or manipulative. A relatively high number of participants (11) trusted link suggestions; but there is no clear evidence that the suggestions are perceived as useful.

	Disagree	*Neutral*	*Agree*
Suggestions were helpful	5	4	7
Suggestions were intrusive or annoying	13	1	2
I could trust suggestions	4	1	11
I felt being manipulated by the suggestions	5	4	7

Table 5.2: *User perception of link suggestions.*

Suggestions improve user perceptions and decrease task execution time

Participants in the support condition disagreed to a larger extent than participants in the control condition with the following statements: 'It was difficult to find the information I needed on these sites' (t=-2.72, $p < 0.01$), and 'Labels of links and categories confused me' (t=-2.83, $p < 0.01$). Participants receiving suggestions agreed to a larger extent than participants in the control group that 'the websites can be used without previous experience' (t=2.33, $p < 0.05$). For all other items, differences were not significant.

When looking at aggregated measures of user perceptions - perceived disorientation and perceived usability - the differences between conditions appear to be non-significant. However, there is a marginally significant result: the level of disorientation is lower in the support condition, but this difference is significant only at an alpha level of 0.10 (two tailed). We also observed a significant interaction between the variable gender and the variable support in relation to perceived disorientation (F=5.12, $p < 0.05$); men and women benefit to different extents from link suggestions. This last result will be dealt with at the end of this section. When the interaction between gender and support is taken into consideration, the effect of support becomes significant (F=9.43, $p < 0.01$). Therefore, it is now clear that there is a significant effect of providing suggestions on perceived disorientation, but only for men.

When the two conditions are compared based on the average task execution time, a significant difference is revealed - see table 5.3. On average, participants in the control condition spent 558 seconds per task. In the support condition, the average time spent per task is 391 seconds (t = 5.99; $p < 0.01$). The spread of task execution times in the support condition is almost twice as low as in the control condition. This difference in variance between the two groups is a natural consequence of our manipulation; the aim of link suggestions is to prevent users from spending time on unsuccessful trials.

	N	Mean	Std. Dev.
Control	16	558.22	99.93
Support	16	390.98	50.05

Table 5.3: *Task execution times per condition.*

User navigation is better structured

The results presented in the previous subsections indicate that participants did believe that the link suggestions could be trusted and that they were considered slightly helpful. We now turn to the question whether the link suggestions actually changed the participants' approach to solving the tasks at hand.

As explained in section 3.4, we extracted a number of measures from the navigation paths that captured patterns of page revisits, page view times and navigation complexity (Herder & Juvina 2004) (Herder & Van Dijk 2003) . An independent-samples t-test was carried out to find significant differences in means between the two conditions. The result showed that participants in the support condition:

- used the back button less (t=-2.24, $p < 0.05$);
- the navigation paths had a lower compactness (t=-3.02, p¡0.01) and a higher stratum (t=3.42, $p < 0.01$), i.e. the paths were more linear;
- the average connected distance in the navigation path was higher (t=2.26, $p < 0.05$)

We also carried out principal component analysis on the twenty-two measures with equamax rotation to find linear combinations of the measures that indicate navigation patterns. Four factors were found that were quite similar to the four factors found in our previous study (Herder & Juvina 2004). We will concentrate here on the fourth factor, of which the means differed significantly between the two conditions (t=-4.01, $p < 0.01$). This factor correlates with the following first-order navigation measures ($p < 0.05$):

- high compactness (r=0.896) and low stratum (r=-0.861)
- many visits to the site's home page (r=0.496),
- short average connected distance (r=-0.388)
- frequent use of the back button (r=0.361)

Apparently, the link suggestions caused the subjects to navigate in a more linear manner and reduced the number of visits to the site's home pages. There are two possible explanations for this effect: either the subjects simply followed the suggested links, without bothering to explore the site structure (McEneaney 2001) - a negative effect, or the subjects got stuck less often - getting stuck usually results in returning to the site's home page to start another trial (Otter & Johnson 2000) - which is a positive effect.

Analysis of the navigation path visualizations led us to the firm belief that the latter is the case - see figure 5.5. In three of the five tasks the participants that were not provided with link suggestions typically appeared to randomly return to pages visited before and eventually return to the site's home page for another trial - effects that are frequently reported to be caused by disorientation (Otter & Johnson 2000) (Smith 1996). These effects were significantly less visible in the visualizations of the participants provided with link suggestions, yet highlighted links were often not followed - which is a good sign, as the paths generated

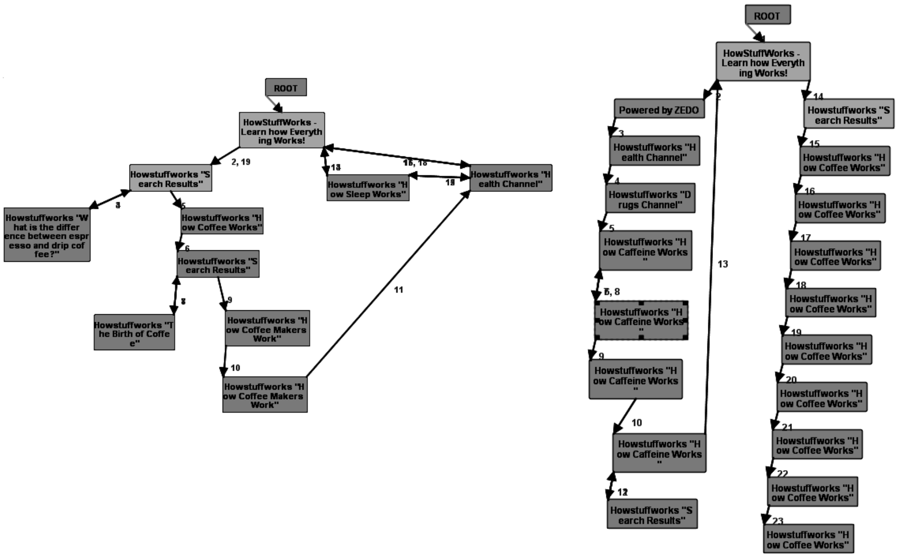

Figure 5.5: *In the picture on the left the navigation path of a participant working on a task without link suggestions is displayed. The randomness of page revisits is clearly visible. In the picture on the right the navigation path of a participant working on the same task, but provided with link suggestions, is displayed.*

by the CoLiDeS+ algorithm, were suboptimal due to the nature of similarity measurement by latent semantic analysis (Juvina & Van Oostendorp 2004). This provides evidence that the main impact of link suggestions is that they provide additional context.

There is no significant interaction between the variables gender and support with respect to any of the navigation measures considered. In other words, navigation patterns are basically the same for men receiving link suggestions as for women receiving link suggestions.

Winners win even more - gender differences

In general, when all participants from both conditions were pooled together, no differences between men and women were found in this study, at least with respect to the variables we have considered here - i.e. navigation measures and user perceptions.

However, although in the control condition women and men declare about the same level of disorientation, in the support condition men declare a much lower level of disorientation than women. Women do not seem to benefit from being provided with navigation support, their perceived disorientation levels are about the same in the 2 conditions. In the support condition, men and women also differ with respect to perceived usability. Men receiving navigation support perceive the websites more usable than men not receiving support (t = -2.66, p¡0.05). The difference is not significant for women. Therefore, there seems to be a gender gap indeed, but in this study we were able to find it only with respect to how much the two genders are able to benefit from being provided with navigation support and how they perceived the usability of the systems they used when such support was offered.

5.3.4 Discussion

In this study we explored the impact of providing link suggestions on user navigation behavior and on user perceptions. In general, highlighting links that are relevant to the task at hand is a well-received navigation support. Link suggestions make navigation path more linear, more structured and less redundant - a style that is associated with low degrees of perceived disorientation. Users provided with link suggestions were expected to perceive websites as more usable and themselves as less disoriented, but this expectation was confirmed only for men. Women seem to profit from suggestions only objectively - their navigation path becomes more structured, but not subjectively - their perceptions do not improve when receiving link suggestions. A possible explanation for this effect is that men use the link suggestions in addition to the available orientation clues.

There are a number of limitations to this study. First, we have not yet evaluated the impact of link suggestions on task performance. So far we only know that link suggestions make users' navigation more structured, reduce task execution time and improve user perceptions, but we don't know whether they actually help users. However, the lack of task performance results is not that important, as the impact of link suggestions on task performance varies per usage context and the way link suggestions are generated. Second, the number of female participants exceeded more than twice the number of male participants (22 female, 10 male) due to the way participants were recruited.

In our study we explicitly attempted to let users carry out real-life tasks. This aim is obviously violated by the fact that the subjects were given predefined tasks - or scenarios - in a laboratory setting. Although we have the impression that the impact of the artificial context is quite low, it is an open question what the effect of link suggestions will be in real-life situations in which users work on multiple tasks simultaneously. Nevertheless, before being able to observe the effects 'in the wild' it is necessary to first study them in more controlled settings.

In conclusion, link suggestions relevant to users' tasks have in general a positive impact on users. However, due to individual differences between users, the effects might not be the same for all users. Moreover, these differences might lead to other - possibly counter-intuitive - effects than anticipated. More research and user studies on the effects of personalization techniques are needed to find out what techniques are best suitable for various personalization goals.

5.4 Concluding remarks

In this chapter two laboratory studies are presented, focused on how to recognize and how to alleviate well-known problems associated with users experiencing disorientation while looking for information on the Web. In the first study we found two navigation styles that are closely related to a user's self-reported degree of lostness: *flimsy navigation* is a weak navigation style, characterized by slow navigation, unstructured backtracking and frequent returns to the site's home page. *Laborious navigation* is an exploratory navigation style with frequent backtracking, mainly to navigational hubs; these activities typically take place at the beginning of a task and serve to quickly build a mental model of the site structure. In the second study we investigated the impact of link suggestions on user navigation styles. Users who were provided with task-related link suggestions scored significantly lower on a navigation style that bears many similarities with the flimsy navigation style found in the first study. This result is strengthened by a decrease in task execution time and better user perceptions on disorientation and site usability.

The results from the two studies confirm the observation that disorientation leads to a decrease in performance. This decrease in performance cannot merely be explained by an increased amount of page revisits, as suggested by Smith (1996), but rather by revisit patterns and the resulting navigation complexity. Unstructured navigation paths with apparently random page revisits indicate that users are unable to keep track of their positions (Thüring et al. 1995), which may lead users to unexpectedly returning to pages visited before and to return to the site's home page for another trial. By contrast, purposeful revisits to navigational hubs and periods of heavy exploration show that users build and make use of a mental model of the site structure (McEneaney 2001) (Otter & Johnson 2000).

From a theoretical perspective, the time spent on learning the site structure by exploration can be explained in terms of *enrichment activities*. In chapter three we explained the concept of information foraging (Pirolli & Card 1999), which explains user navigation in terms of following one's 'information scent'. With little information on the environment, users have to rely solely on the available proximal cues, such as link anchors. Orienteering behavior (Teevan et al. 2004) provides context for interpreting the proximal cues and allows more distant cues to propagate to the current user position. Further, the cognitive overhead needed for keeping track of one's position (Thüring et al. 1995) will be highly reduced and more attention can be given to the actual task. This situation may be compared with people moving to another city: the best way to find out 'where things are' is by first spending some time on walking or driving around with no apparent goal in mind. In terms of utility, the time that users will spend on orienteering will be a trade-off between effort and expected increase of results in terms of navigation efficiency. We believe that the empirical findings of our studies call for the integration of *enrichment by orienteering* in existing theories on information foraging and Web navigation. It has been argued earlier in this thesis that the assumption of the CoLiDeS model (Kitajima et al. 2000) that backtracking activities may be considered as 'recovering from a failure' does not reflect the knowledge gained by trails that did not lead to the actual goal. A first step to improve the model on this matter is carried out by (Juvina et al. 2005), who presented a model called CoLiDeS+, which we have used for generating the link suggestions used in our second study. In addition to the direct expected relevancy of the proximal cues in relation to the user goal, the model reflects the expected contribution of each alternative option to the information gained thus far, the *path adequacy*.

Most reported studies in the field of adaptive hypermedia assess actual systems, whereas exploratory studies as described in this chapter, are rare (Weibelzahl 2003). In most cases, studies in the former category show whether a system satisfies the goals is was designed for, yet they offer little insight in the impact of adaptive hypermedia techniques nor do they bridge the gap between empirical observations and theory on user navigation. In our studies we combined quantitative measures on user navigation with qualitative data from questionnaires, cognitive tests and self-assessments to relate our findings with both earlier observations and theoretical insights. The study on the impact of link suggestions was aimed to show the effects of one specific adaptation technique without referring to a specific domain or system (Weibelzahl 2003). Our strategy was to fix other relevant variables, such as the user population, the tasks and the information sources used, to ensure that the differences measured would most likely be caused by the adaptation technique itself. However, it would require a large amount of such limited studies to investigate the impact and usefulness of all varieties of knowledge inference, adaptation decisions and presentation techniques. Yet, in order to move toward an integrative model of hypermedia usage and Web navigation, and associated theoretical measures for the impact of adaptive systems, this kind of background knowledge is required.

In this chapter we provided a theoretical assessment measure for adaptation techniques that aim to reduce problems associated with disorientation in hypermedia systems, while looking for information in websites not visited before. We specifically aimed to identify navigation styles as an intermediary representation, instead of directly applying regression techniques to explain disorientation in terms of measures on navigational complexity. The navigation styles were identified and interpreted without relating them to the issue we were actually interested in, which means that they can be considered on themselves and not as a combination of measures that explains lostness. Both for the development of predictive cognitive models - which aim to simulate these patterns - and the development and evaluation of adaptive hypermedia techniques and systems, these patterns are highly functional as a point of reference.

Due to the focus of the studies, the discovered patterns are mainly applicable for within-patch navigation - navigation within a Web site by forward navigation actions such as link following, and backtracking activities. However, switching between patches by search activities has become an important part of information finding strategies. Including Web search in our studies would have introduced several problems, as this would most likely have resulted in many different Web sites visited, which would hinder the comparison of the navigation paths. Further, due to its laboratory setting, the navigation patterns that we found are representative for information finding activities, but most likely not for recurrent activities, which form a major part of user Web navigation (Tauscher & Greenberg 1997). In order to gain more insight in recurrent activities and the relation between search, forward navigation and backtracking activities, we conducted a long-term study, which is reported in the next chapter.

Chapter 6

Long-term study

6.1 Introduction

Both for the design of Web sites and the design of user-adaptive navigation support, insight is needed on how users navigate the Web. In chapter 3 several laboratory studies, field studies and theoretical models have been discussed. Information gathering activities were considered the most important activity on the Web, leaving still about 40% of activities that were considered general browsing, transacting, housekeeping and communicating. Searching, browsing and backtracking are the main activities while gathering information. In general, recurrent behavior is predominant in Web usage, with a strong focus on a small set of highly popular pages and most recently visited pages. Part of this recurrent behavior can be considered as backtracking activities as a result of *orienteering* behavior, another part as recurrent activities.

The last long-term study on Web usage behavior dates back to 2000 (Cockburn & McKenzie 2001). Together with earlier studies (Catledge & Pitkow 1995) (Tauscher & Greenberg 1997) they provide insights on patterns in Web navigation. Yet, they only provide a limited and mainly quantitative view. Furthermore, given the fast development of the Web and its usage in the past few years, these insights might already been outdated.

In this chapter we present results from a long-term study, conducted in cooperation with Harald Weinreich, Matthias Mayer and Hartmut Obendorf from the University of Hamburg, Germany. Results have been reported in (Herder 2005) (Weinreich, Obendorf, Herder & Mayer 2006, forthcoming) (Herder, Weinreich, Obendorf & Mayer 2006, forthcoming). The chapter is organized as follows. In the first section we describe the data collection process. We continue with general statistics on Web navigation; how much time do users spend on the Web, how long do they stay on pages and what activities do they carry out. In the third section we deal with several aspects of page revisits on the Web; the amount of revisits among the navigation actions is reconsidered, different categories of revisits are identified and the extent to which these revisits can be supported. In section four we analyze user search activities on the Web; besides general statistics on the amount of search, different search activities and query lengths, we also identify individual differences and describe the relation between search and recurrent activities. In section five the use of browser windows, tabs and interface widgets is described; from the results it becomes apparent that Web navigation does not take place in a single window and that a significant amount of activities involve interacting with tools rather than navigating between content - this has several consequences for Web navigation support. In section six we investigate differences in navigation styles among navigation sessions. We conclude with a summary and discussion of the results.

6.2 Data collection

The participant pool consisted of 25 participants, who were recruited by personal invitation and were not paid for participation. Nineteen participants were male and six female. The average age was 30.5, with a range from 24 to 52 years. All participants were experienced internet users, browsing the Web for 3 to 12 years - 8 years on average. Seventeen participants were German - of which three lived abroad -, the remaining eight participants were Dutch and living in the Netherlands. While all eight participants from the Netherlands (32%) worked as university employees in computer science, nine participants had a different background: two worked in psychology, and one each in sociology, geology, electrical engineering, trading, coaching, history, and photography. Seven additional participants began the study, but dropped out for personal or technical reasons and had to be excluded from the analysis.

Whereas the participant pool is relatively heterogeneous, still the participants' average background is academic and technology oriented. This implies that the results from this study most likely cannot be generalized to *all* users. However, as earlier long-term Web usage studies made use of similarly composed or even more uniform participant pools (Catledge & Pitkow 1995) (Tauscher & Greenberg 1997) (Cockburn & McKenzie 2001), we can use these results as a base for comparison.

Given the importance of the Web, it might seem surprising that only a small number of quantitative long-term studies analyzed the browsing behavior of Web users. This may be explained by the issues we encountered during the preparation of the study. Today, browsing is considered a personal activity, even if logging only takes place at the workplace. The Web is now used for many confidential tasks, such as online banking, shopping or writing e-mails. After initial informal surveys, it became clear we had to make use of a capturing system that did not record usernames or passwords and ignored secure connections. The participants were also given the option to screen the content of all log files before transmitting them to us.

Some prospective participants were also concerned that the installation of such 'spyware' might have negative influence on the stability or performance of their personal computer. These concerns were not unfounded: in one of our pilot studies instrumenting the Internet Explorer to record user actions and page requests led to stability issues when different Explorer versions were used or new plug-ins were installed - unacceptable for a long-term study where the browser is used daily as a production tool.

6.2.1 Data capturing procedure

The data was collected using a proxy server, which is part of the Java-based framework Scone, described in more detail in chapter 4. Basic data included the times at which the page requests took place, the Web address, title and size of requested pages and the time spent on the pages. Javascript events were inserted into the Web pages to capture additional information, such as the opening and closing of browser windows and tabs. As the participants were required to register themselves in the framework and were instructed to turn off their browsers' caching, no heuristics were needed to separate users or to reconstruct navigation paths (Cooley et al. 1999).

15 of the 25 participants used an instrumented version of Firefox that was modified to record the interaction with all user interface widgets. These participants were either already using Firefox as their preferred browser or took the opportunity to switch. Using exact timestamps, this second clickstream log could be merged with the navigation log of the intermediary to gain more detailed and accurate data.

The participants were logged for some period between August, 2004 and March, 2005. The average time span of the actual logging periods was 104 days, with a minimum of 51 days and a maximum of 195 days. Participants were logged in their usual contexts - seventeen at their workplace, four both at home and at work, four just at home. For 22 participants the focus of their Web activities was mainly work related, for 3 it was private activities. Several participants used other, non-logged, computers or browsers during the study period; some due to technical problems, others mentioned privacy reasons - as an example, they did not want their online bank transactions to be recorded. All participants mentioned that between 50% and 100% of all their Web activities were logged during the study period, eight participants mentioned 90% or more. As a comparison, the Web logs used by Cockburn & McKenzie (2001) consisted of four months of Web usage data of 17 participants in total, during which 84.841 page requests were recorded.

During the logging period 152,737 page requests in total were recorded. A significant number of these page requests turned out not to be initiated by the users themselves but to be automatically generated as a side-effect of user actions. The major categories of these artifacts are:

- *advertisements*, typically embedded into other pages using iframes;
- *automatic reloads*, mainly news sites which refreshed after some time period;
- *automatic redirects*, mainly on dynamically generated Web sites;
- *frame sets*, various single files that together constituted one page view.

Various heuristics were used for identifying these artifacts, including server exclusion lists, patterns in Web addresses and temporal aspects. These heuristics are explained in more detail in chapter 4. As an example, automatic reloads typically generated peaks in the otherwise power law distribution of time spent on Web pages. In total 10.1% of all page requests were identified as artifacts - with individual percentages of over 28% for participants who did not use ad-blockers -, leaving 137,737 page requests that were used for analysis.

6.3 General statistics

On average, the participants issued 5491 page requests during the logging period, which would translate to 47 page visits per day - a number that is in line with the 42 page visits per day reported by Cockburn & McKenzie (2001). However, the average does also include days on which participants did not access the Web from their workplace - most importantly weekends and holidays. Therefore, we recalculated the average to represent the number of visits per *active day* - an active day is defined as a day on which at least one page request was issued. Given this definition, the average number of page visits per day was 89 ($\sigma = 63.42$), with a minimum of 25 and a maximum of 284 visits per day.

This study confirms the rapid interaction behavior with heavy tailed distribution already reported by previous studies (Cockburn & McKenzie 2001). Our participants stayed only for a short period on most pages: 25% of all documents were displayed for less than 4 seconds and 52% of all visits were shorter than 10 seconds (median: 9.4 seconds). However, nearly 10% of the page visits were longer than two minutes. Figure 6.1 shows the distribution of stay times grouped in intervals of one second. The peak value is located at stay times between 2 and 3 seconds; they contribute 8.6% of all visits.

For splitting the participants' navigation data into sessions, we used the common 25.5 minutes session time-out heuristic, as established by Catledge & Pitkow (1995). The average number of sessions per active day is 3.25 ($\sigma = 1.19$) and in total 77 minutes ($\sigma = 49.91$) were spent on these sessions per day. From figure 6.2 it can be observed that short sessions are more prevalent than longer sessions; however, there is a long tail of longer sessions. The median session length is 13 pages, but the fast majority of page visits are part of a session of 34 pages or longer - see figure 6.3.

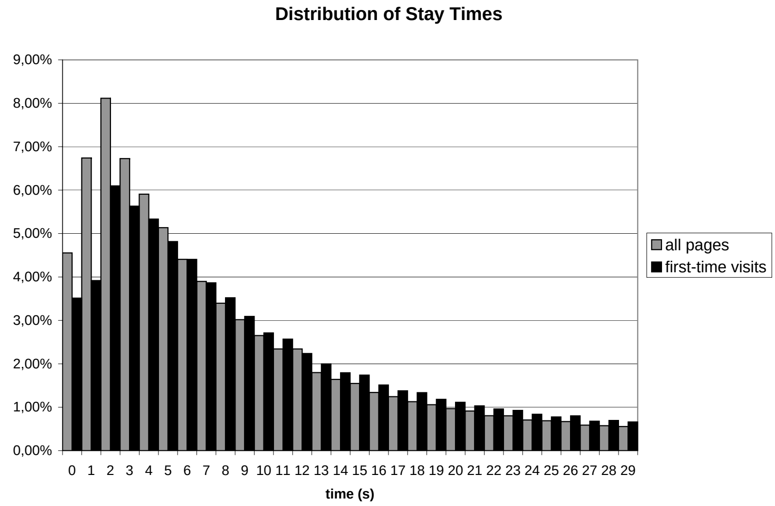

Figure 6.1: *Distribution of stay times for all participants.*

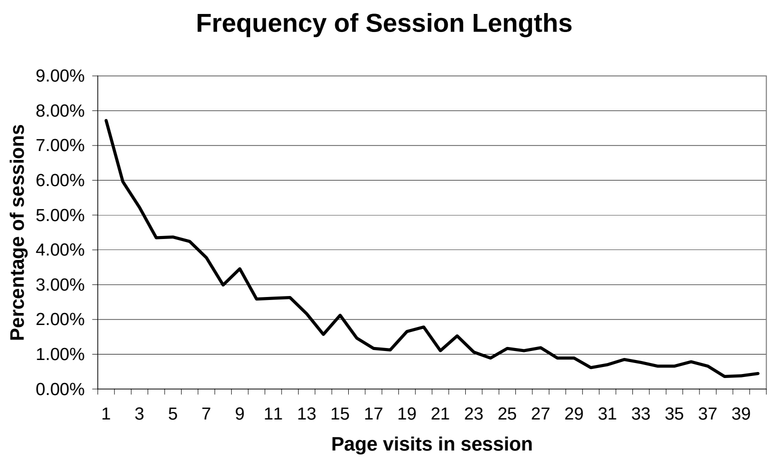

Figure 6.2: *Distribution of session lengths; although in number sessions with less than 10 page visits are more common than longer sessions, the tail of longer sessions is quite long*

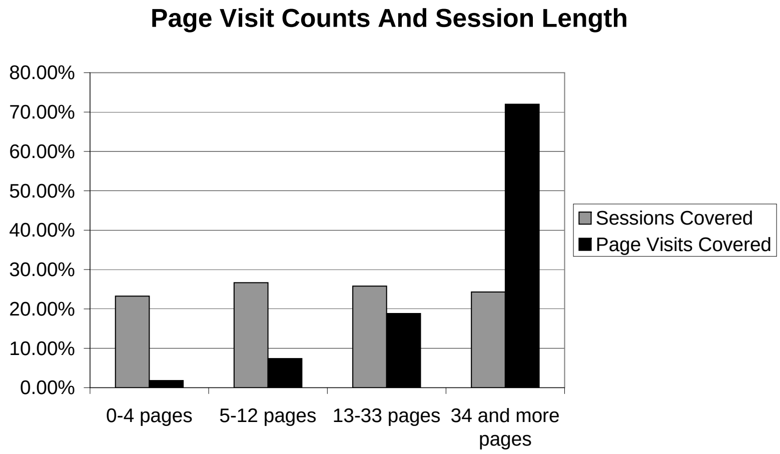

Figure 6.3: *Although only a quarter of the sessions is longer than 34 pages, these sessions cover over 70% of all page visits*

From these general results several observations can be made. First, the frequently reported and assumed power law distribution of session lengths (Baldi et al. 2003) may be valid for site visits - which unfortunately are often called sessions as well -, but actual user sessions typically involve a large number of page visits. Moreover, most Web activity takes place in sessions longer than 34 pages. Further, on average our participants spent over an hour per day actively navigating the Web. These numbers show that, at least for our participants - who may be qualified as 'knowledge workers' - Web usage can well be called a time-consuming activity.

6.3.1 Activities carried out on the Web

To find out what activities our participants mainly carried out on the Web, we classified the top 300 most frequently visited sites, which amounted for 65% of all page requests. This list is headed by the Google search engine - including the localized German and Dutch Google sites -, several news sites, some participants' shared institutional Web sites, a well-known online marketplace and a German-English dictionary - which was heavily used by most German participants. Sites at the bottom part of the top 300 list were visited less than 50 times by one or two users.

In order to categorize the sites, we followed a two-step process. In the first step, 54 subcategories were identified by one member of the research team, and verified and corrected by a second member. Subsequently, both researchers independently classified these 54 subcategories in eight meaningful categories. Both team members came up with eight almost identical categories - except for minor differences in the labels given. In the merging process, a number of sites was

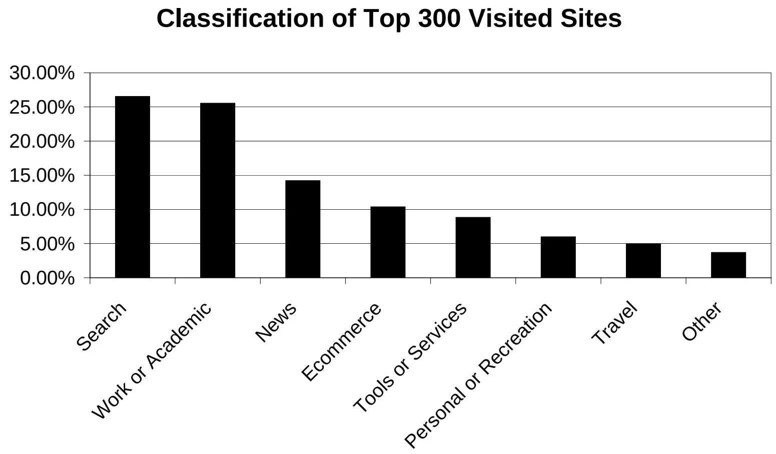

Figure 6.4: *Main categories of Web activities as deduced from the participants' Web logs.*

agreed on to be moved to a more appropriate category. The major categories we found, are:

- *Search*: general search engines, literature search, other specific search sites;
- *Work or University*: institutional Web sites, project Web sites, work-related applications;
- *News*: news, weather, sports, tech news;
- *Ecommerce*: online stores and marketplaces;
- *Tools or Services*: dictionaries, online reference material;
- *Personal or Recreation*: bulletin boards, club sites, general interest;
- *Travel*: public transport, location maps, route descriptions, hotel bookings;
- *Other*: purpose not clear or not belonging to one of the above categories.

The associated visit percentages are listed in figure 6.4.

In addition, in a pre-study interview we asked our participants what they would use the Web usually for. Different numbers of comments per participant were possible. Of the 171 comments many were related to search and research, looking for specific information, literature search, people and product information. The Google search engine and the German-English dictionary were explicitly mentioned several times. Less frequently mentioned activities included blogs, self-edited sites and the specific use of Google for spell checking purposes. Similar

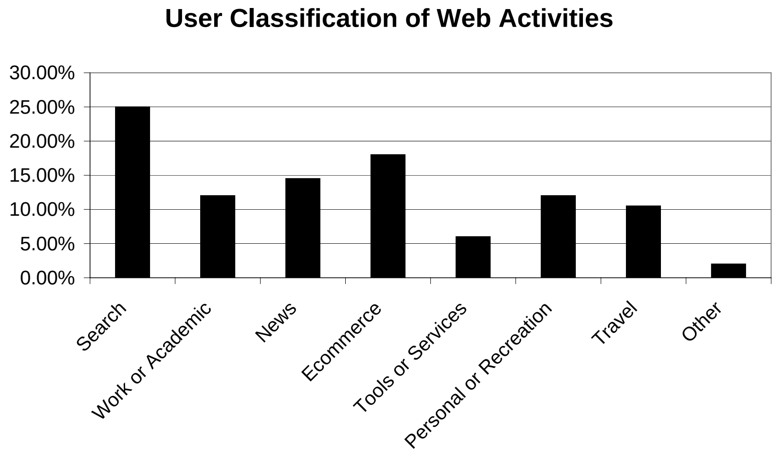

Figure 6.5: *Main categories of Web activities as indicated by the users themselves.*

to the top 300 domains, we classified all comments in the eight major categories derived before - see figure 6.5.

When comparing the two figures, it becomes apparent that our participants were less involved in ecommerce, recreational browsing and travel planning than they indicated themselves, and more involved in work-related activities. An obvious explanation for this bias is that the majority of the participants was only logged on their workplace, an effect that needs to be taken into account while interpreting the results. However, it may also be that our participants were overestimating the time spent on these activities, which may reoccur frequently, but which are less time-consuming than work-related information gathering activities.

6.4 Revisitation on the Web

While browsing the Web, users frequently return to pages already visited before. Earlier studies (Catledge & Pitkow 1995) (Tauscher & Greenberg 1997) (Cockburn & McKenzie 2001) have shown that the majority of page requests involves requests to pages visited before. Of these page revisits, a majority is covered by a small number of very popular pages that are visited far more often than any other page. Another interesting observation is that most page revisits involve pages visited only very recently in the past.

As recurrent behavior is predominant in user navigation actions, Web browsers provide various navigation tools to support page revisits. The most well-known and most frequently used tool is the browser's back button, which allows for revisiting pages visited earlier in the same session. Tools for revisiting pages from earlier sessions are bookmarks and the temporally ordered history list - which is hardly used (Tauscher & Greenberg 1997). All of these three revisitation tools have several problems that undermine their usability; perhaps the most serious problem is the lack of integration of these tools (Kaasten & Greenberg 2001). Various research efforts on the design of novel revisitation tools are reported in the literature (Tauscher & Greenberg 1997) (Mayer 2000) (Milic-Frayling et al. 2004). This research is typically guided by empirical observations such as the frequencies, distributions and probabilities of various forms of page revisits.

In the past decade three long-term studies have been carried out to characterize users' page revisit behavior (Catledge & Pitkow 1995) (Tauscher & Greenberg 1997) (Cockburn & McKenzie 2001). Results from the most recent study, carried out by Cockburn and McKenzie, were based on data from late 1999 and might need to be updated to better reflect current usage patterns. Also, the studies did not separate within-session page revisits from cross-session page revisits. As will be explained in more detail in this section, it is useful to discern these two activities, both for the design and evaluation of novel or adaptive revisitation support.

In the first part of this section we return to the question whether page revisits really accounts for 81% of all navigation actions - as reported by (Cockburn & McKenzie 2001); this number is likely to be an overestimation. We continue with an overview of different motivations for page revisits. In the third subsection we describe two characteristic distributions found in user page revisit behavior, separate within-session revisits from cross-session revisits and characterize the frequencies with which they occur. We continue with an analysis to what extent the two characteristic distributions may contribute to these different kinds of revisits. In the second-last part of this section the page popularity distribution in frequently visited sites is characterized. We conclude with a discussion on support for revisiting less frequently visited pages.

6.4.1 How often do people revisit Web pages

Earlier studies have shown that page revisits are very common in Web browsing behavior (Tauscher & Greenberg 1997). The common formula used for calculating the percentage of revisits among navigation actions is

$$R = 100\% \times (1 - \frac{individual\,pages\,visited}{total\,page\,visits}) \tag{6.1}$$

Interestingly, the average percentages shown in these studies indicate that page revisits have become more common in the past few years. From the log data from 1994 and 1995 (Catledge & Pitkow 1995) (Tauscher & Greenberg 1997) revisit rates of respectively 58% ($\sigma = 9\%$) and 61% ($\sigma = 9\%$) were reported; Cockburn & McKenzie (2001) report a significantly higher percentage, 81% ($\sigma = 10\%$), of revisits in their Web logs of late 1999. According to (Baldi et al. 2003) this higher percentage might indicate that the usage of the Web may have evolved 'from a more exploratory mode in 1994-1995, to a more utalitarian mode by 1999', a mode

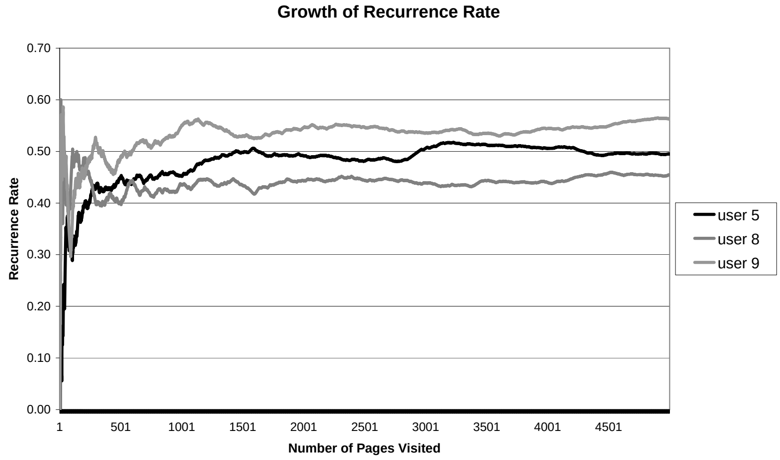

Figure 6.6: *From the picture above it can be observed that the page revisit rate stabilizes after a short period*

where regular visits to sites as general news, travel planners and bulletin boards are predominant. From the 2004 ranking of top 50 websites (comScore Media Metrix 2004) it can be observed that these sites - entertainment and ecommerce oriented sites in particular - are highly popular indeed compared to 1996.

Based on the above observations, we expected the average percentage of page revisits in our data to be as high or even higher than the 81% reported by Cockburn & McKenzie (2001). Much to our surprise, the average revisit rate for our participants was only 49% ($\sigma = 11\%$). Per-subject revisit rates ranged from a minimum of 22% to a maximum of 79%. As our participant pool is very similar to the participant pools used in the earlier studies, this significant difference either indicates a dramatic change in Web usage or it might be caused by differences in the way the data is analyzed.

As an alternative reason for the differences in revisit rates between the studies, Baldi et al. (2003) suggested that the estimation of the revisit rate from a finite time-window may be an underestimation, as long-term revisits may not be captured during the logging period. However, as indicated in the previous section, the time window used in our study is about the same as the time window used byCockburn & McKenzie (2001) - 104 days on average - and exceeds the time window used in the other studies. We analyzed the effect of the length of the logging period on the page revisit rate and found that the revisit rate stabilizes after about 1000 page views, a number that is reached by each participant in about 10 active logging days, see figure 6.6.

The rather quick stabilization of the recurrence rate can be explained by Tauscher & Greenberg (1997)'s finding that the ratio between the number of page requests and the number of unique pages visited thus far - the *url vocabulary*

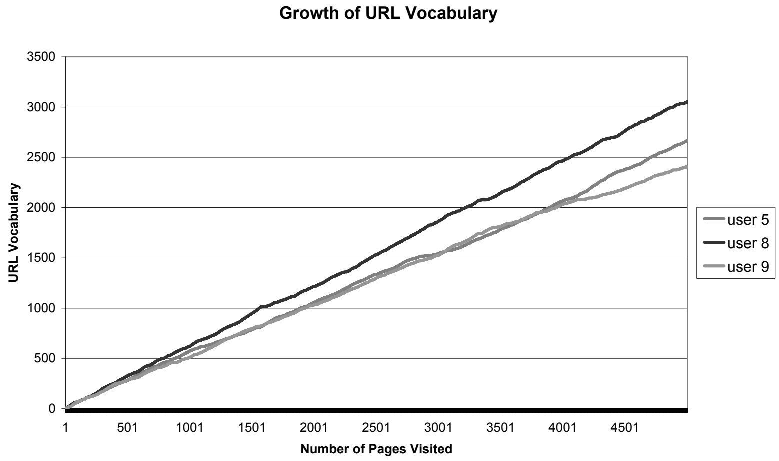

Figure 6.7: *The ratio between the number of page requests and the number of unique pages visited thus far is roughly linear. The slight drop in vocabulary growth of user 5 at the end of the logging period translates to a slight increase in the recurrence rate.*

growth - is roughly linear, see figure 6.7.

We found that the reported increase in page revisits is in fact due to differences in data preprocessing. Cockburn & McKenzie (2001) mentioned several data cleaning steps that were taken before analysis. One particular step appeared interesting: URLs involving search queries were truncated to remove suffixes of the form *?name=value&name=value...*, which implies that all queries to search engines, as well as various dynamically generated pages, were generalized into visits to just one page. The authors indicated that this cleaning step did not distort their results and that the characteristics reported were similar to the 'unclean' data. We reanalyzed our data after removal of the query terms, which resulted in a revisit rate of 70,8% ($\sigma = 8,2\%$). This percentage is even higher on the unprocessed data - 73,7% ($\sigma = 8,5\%$), which is the value we used for comparison with the Cockburn and McKenzie study, as they did not report any removal of artifacts caused by adservers, frame sets or automatic reloads. The 95% confidence interval is between 70,2% and 77,2%, which is most likely to have an overlap with the confidence interval on the Cockburn and McKenzie data, which they did not report in their paper.

We leave it up to the reader which of the estimates represents best the actual amount of page revisit behavior. Personally, we prefer the lower estimate, as the query part of a url 'serves to identify a resource within the scope of the URI's scheme and naming authority' (Network Working Group 2005). In practice this means that search engine queries or dynamic page locators often result in different Web pages to be loaded. However, from the above it has become clear that the reported increase in page revisits between 1995 and 1999 is most likely due to

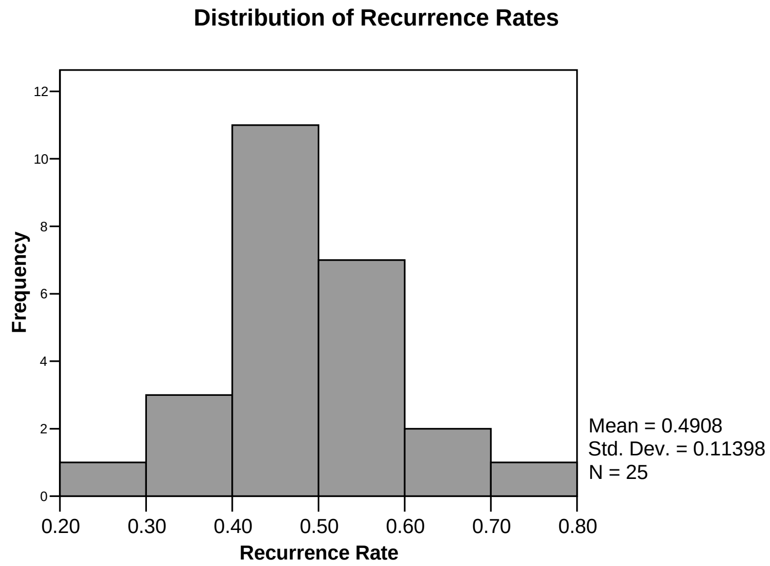

Figure 6.8: *While the average recurrence rate is 49%, large differences can be observed between users.*

a difference in data preprocessing rather than a change in general Web usage; the average percentage of page revisits has remained fairly stable during the past decade.

What may even be a more important observation is the wide variation of recurrence rates between users - see figure 6.8. Extensive support for recurrent behavior may be beneficial for users with an average to high recurrence rate, but users who revisit pages to a smaller extent will most likely take less profit from this support and might better be helped with emphasis on support for finding new information. Rather than designing for the average, designers should take these differences into account.

6.4.2 Within-session and cross-session revisits

Whereas recurrence rates vary between users, in general page revisits are very common in Web browsing behavior. Tauscher & Greenberg (1997) identified the following main reasons for revisiting pages:

- the information contained by them changes;
- they wish to explore the page further;
- the page has a special purpose;

- they are authoring a page;
- the page is on a path to another revisited page.

This frequently cited list of reasons applies to page revisits in general. There is one important subcategory of page revisits that we would like to point out explicitly: revisits to pages visited before in the same navigation session. As the Web is a non-linear medium, users typically navigate in a non-linear manner (McEneaney 2001). Within the user navigation paths several pages can be recognized from which multiple alternative paths are initiated. These pages are generally called *hubs* (Kleinberg 1999) Milic-Frayling et al. (2004). Likely candidates for pages to become hubs are sites' home pages and index pages that serve to navigate users to a number of pages - such as tables of contents and lists of search results (Pirolli et al. 1996).

Whereas recurrent behavior is heavily reported in empirical studies, most theoretical models of Web navigation take only forward navigation into account (Pirolli & Card 1999) (Kitajima et al. 2000). The CoLiDeS model (Kitajima et al. 2000) regards backtracking as an activity that takes place when forward navigation fails, which does not match the empirical observations mentioned in the previous paragraph. In (Herder 2004) we proposed a model that separated three different categories of navigation actions:

- *searching*, the process of locating information by issuing queries in a search engine;
- *browsing*, the process of viewing and navigating between Web pages;
- *backtracking*, the process of reviewing pages visited before, either for reference or as a starting point for an alternative path.

As motivated by Teevan et al. (2004), the combination of these three navigation activities can be regarded as *orienteering behavior*, which allows users to specify less of their information need explicitly and provides a context in which to interpret the information found. Results indicate that orienteering is a common strategy for relocating and revisiting information from earlier sessions as well. This finding suggests that cross-session revisits usually appear in chunks, in which the same process of searching, browsing and backtracking can be observed.

In an earlier study (Juvina & Herder 2005) we found that certain patterns of within-session page revisits indicate that users understand and exploit the site's navigation structure; users that displayed these patterns used within-site navigation support for page revisitation rather than the browser's back button and showed more confidence that they could relocate the pages to be revisited at a later point. Recently, an enhanced implementation of the back button has been proposed that explicitly supports backtracking to hub pages, including sites' home pages, search results and bookmarked pages (Milic-Frayling et al. 2004).

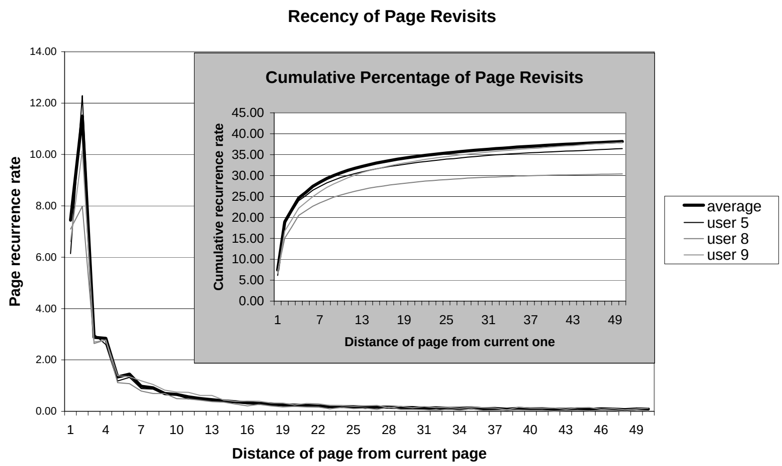

Figure 6.9: *Percentage of page revisits as a function of distance from the current page. The four most recently visited page account for about 50% of all revisits*

6.4.3 Characterizing page revisits

Both Tauscher & Greenberg (1997) and Cockburn & McKenzie (2001) reported two important distributions of revisited pages. First, there is the *recency effect*; from figure 6.9 it can be observed that the majority of page revisits involves revisits to pages visited very recently. Second, there is the dominance of favored pages; as can be observed from figure 6.10, only a small number of frequently revisited pages accounts for the majority of all page revisits. The top n lists of our participants' most popular pages showed an interesting variety, but the majority of these pages could be categorized as search engines, news sites, their own personal or institutional Web sites, or individual interest sites - in particular bulletin boards.

Although there is likely to be an overlap between these two distributions, they characterize two different forms of page revisits: respectively, visits to pages visited before in earlier sessions and visits to pages visited before in the same session. Based on the two distributions alone it is hard to find out which of the two is the most predominant. In order to explore this in more detail, we identified and annotated the page requests with the following revisit categories:

1. visits to pages not visited before;
2. visits to pages visited before in the same session;
3. visits to pages visited before in one or more earlier sessions;
4. visits to pages visited before in the same session and in earlier sessions.

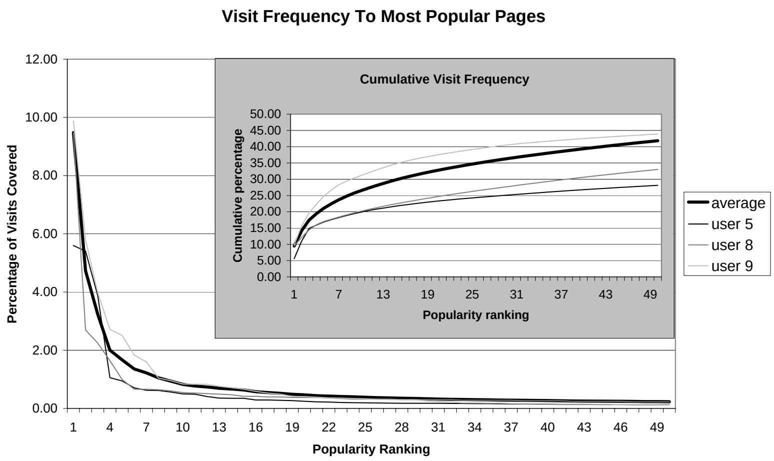

Figure 6.10: *Page popularity rank versus the percentage of revisits covered. The fifteen most popular pages account for about 30% of all revisits*

Following (Tauscher & Greenberg 1997) we used a 25.5 minute time-out mechanism for detecting session boundaries.

In figure 6.11 the distributions of page visit categories are displayed. From the leftmost bar, which conveys the overall distribution, it can be observed that within-session page revisits represent the most common form of revisitation, covering 73,5% of all revisits. 44,2% of the page revisits involves revisits to pages not visited before in earlier sessions. Given the predominance of within-session revisits, it is not surprizing that the back button is the most commonly used revisitation tool.

The four remaining bars show another interesting aspect of recurrent behavior: first-time visits, within-session revisits and cross-session revisits tend to occur in chunks. First-time visits are the most common type of visits and if users look for new information or services, they will most likely continue to do so. Backtracking activities are likely to be followed by either another backtracking activity or by a visit to a new page. This confirms the observation that users frequently backtrack to explore new paths from pages visited before (Pirolli et al. 1996) (Tauscher & Greenberg 1997) (McEneaney 2001). Users involved in recurrent activities will most likely continue to do so as well, backtracking to a similar extent as in first-time visit situations. However, there is still a fair chance of a little more than 30% that they will leave the already visited pages for a page not visited before.

From figure 6.12 it can be observed that within-session revisits and cross-session revisits may occur at any point in a navigation session. Obviously, the very first page visits are unlikely to be within-session revisits, although the ratio stabilizes pretty quickly. For the same reason, short sessions have relatively few within-session revisits. Interestingly, the amount of revisits in general is a little

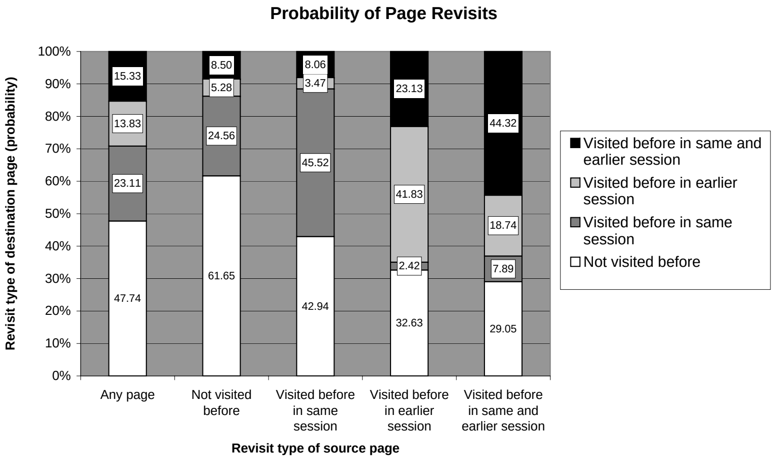

Figure 6.11: *Transition probabilities from one visit category to another. The leftmost bar depicts the overall distribution of categories. It can be observed that users are most likely to remain in the same visit category.*

higher in shorter sessions, mainly due to a large number of cross-session revisits. This confirms the intuition that shorter sessions are more utilitarian in character than longer sessions.

6.4.4 Support for recent and frequent revisits

Intuitively, the popularity and recency distributions, as described in the last section, can be put to use in a straightforward manner: the lists of the top n most popular pages and the top n most recently visited pages will most likely cover the majority of the pages that will be revisited. Tauscher & Greenberg (1997) showed that the list of most recently visited pages performed significantly better than the stack-based back button - which is stack-based, as described in chapter 3, and of which the list is cleared upon closing the browser window. The list of most popular pages performed far worse than both the recency list and the back button.

The bad performance of the list of most popular pages can be explained by the observation from the last section that the majority of page revisits (73,5%) are backtracking activities - revisits to pages visited before in the same session. This type of revisits is typically short-term and therefore quite well supported by the recency-based list. However, the remaining 26,5% of cross-session revisits - recurrent activities - will most likely be better supported by the popularity-based list.

In order to estimate the performance of the two lists for these different kinds of situations, we calculated for each revisit category the relative amount of pages

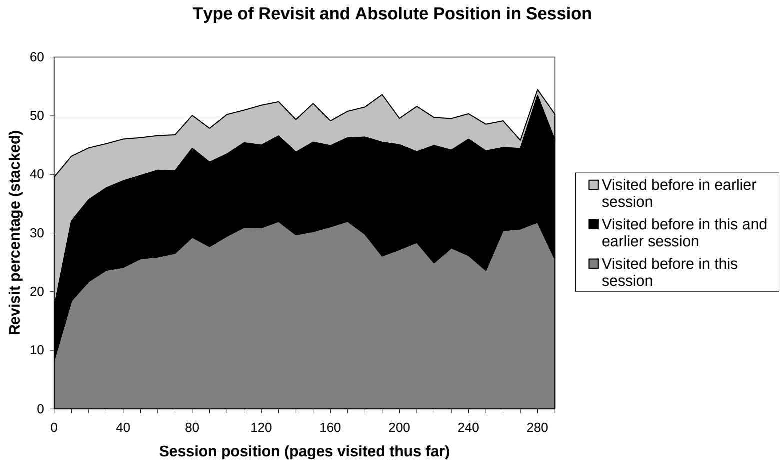

Figure 6.12: *Page revisit probabilities throughout sessions. The spiky behavior in the chart for longer sessions is caused by the low sample rate.*

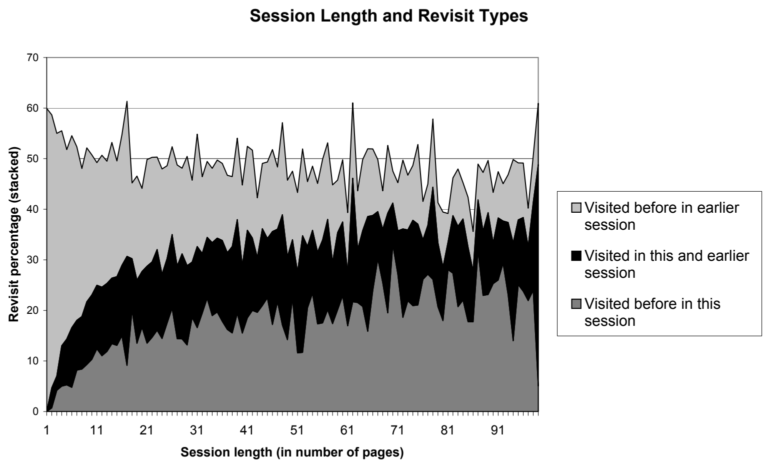

Figure 6.13: *Page revisit rates for different session lengths.*

that were present in the list of 15 most popular or most recently visited pages. The results are listed in the tables displayed below.

revisit type	*average*	*st.dev*
same session	94.56%	2.97
earlier session	21.48%	9.74
same and earlier	87.88%	6.25

Table 6.1: *Revisit support of the list of 15 most recently visited pages.*

revisit type	*average*	*st.dev*
same session	42.28%	31.05
earlier session	51.24%	13.05
same and earlier	71.60%	14.10

Table 6.2: *Revisit support of the list of 15 most frequently visited pages.*

The tables confirm the intuition that the recency-based list supports backtracking activities - which involves revisits to pages visited before in the same session -, and it supports these activities to a remarkably well extent: almost 95% of all backtracking activities are covered. Revisits to pages visited in the same session and in earlier session are still quite well covered, but to a slightly lesser than pages not visited before in earlier sessions. This reflects that the category does not only cover pure backtracking activities while looking for information found before, but also a number of highly popular pages that are revisited several times in the majority of sessions. The popularity-based list performs far worse. This can be explained by the long tail of pages that are revisited only a couple of times. However, revisits to pages from earlier sessions that haven't been visited yet in the current session, are far better supported by the list of most popular pages.

As a comparison, we ran a simulation to find out to what extent within-session revisits are covered by the list of pages behind the back button. As explained by Cockburn & Jones (1996), the list of pages behind the back button is organized as a *stack*. The number of pages that can be directly accessed via the back button is fifteen in most browsers. When using the back button, the user moves down in the stack; the forward button can be used to go to the top of the stack. If a user decides to make use of other navigation tools than the back and forward buttons, all pages that are located above the current location are removed from the stack. We simulated all navigation sessions in the Web logs and reconstructed the contents of the back button stack for each page request. We assumed that each session started with an empty stack and filled or popped the stack with each page request. The average percentage of within-session revisits that was supported by the back button for each user was 51,7% ($\sigma = 10, 8$). This confirms

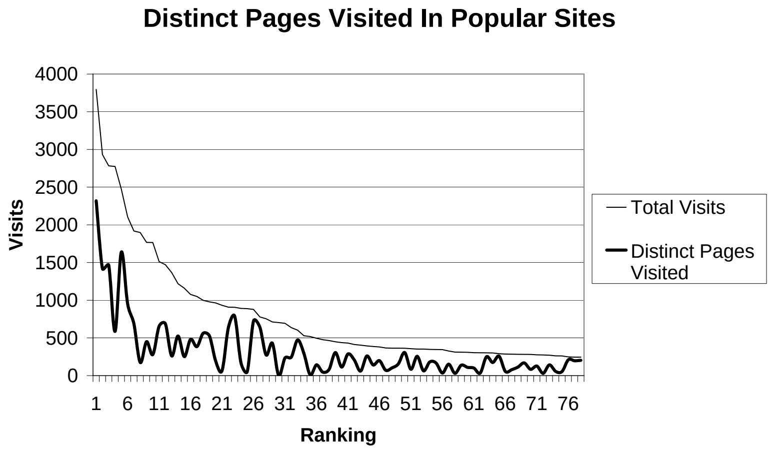

Figure 6.14: *Comparison of the total number of visits to popular sites and the number of distinct pages visited.*

the observation of Tauscher & Greenberg (1997) that the stack-based approach may be suitable for very short-distance revisits, but that it is outperformed by a recency-ordered history list. The percentage of within-session revisits that were actually initiated with the back button is 42,7% ($\sigma = 13,8$).

6.4.5 Popular sites and pages in popular sites

From figure 6.14 it can be observed that when a site is visited more frequently, more pages are visited within the site (r=.903, $p < 0.01$). However, there seems to be a distinction between sites in which only a small number of pages is visited frequently - the plunges in the graph - and sites in which users continue to visit new pages - the peaks in the graph.

An explanation for this effect is given in figure 6.15. Search engines and dictionaries provide a home page that is visited frequently; from this home page a query can be issued that results in an individual result page. Hence, by their very nature, these sites have only one 'popular' page and a long tail of pages that is visited only once or twice. By contrast, whereas institutional Web sites and project Web sites do have a portal page that is visited most frequently, there is a range of other pages that are visited frequently as well; these pages may contain information on a certain topic or department, or may provide an application that is used on a regular basis. News sites provide a different pattern of three to four frequently visited pages; these relate to portal sites of the news categories the user is interested in.

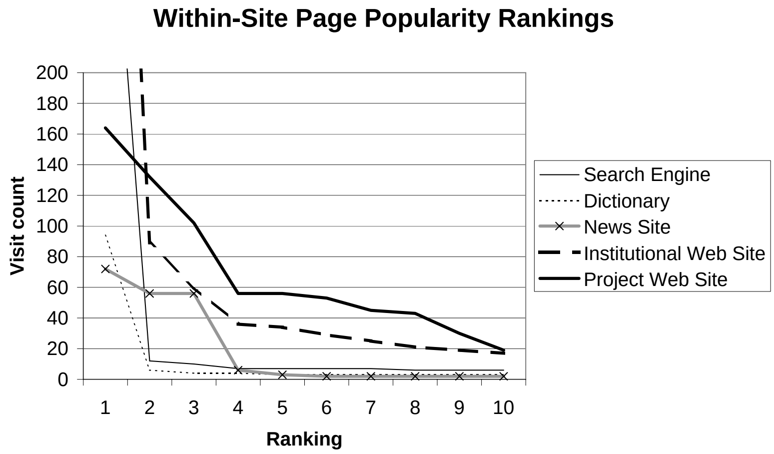

Figure 6.15: *Page popularity ranking in popular sites.*

6.4.6 Support for less frequent revisits

Whereas the recency-based list supports the far majority of within-session revisits, only just half of the cross-session revisits are supported by the popularity-based list. This can be explained by the long tail of pages that have been revisited less often. From figure 6.16 it can be observed that the interval between two subsequent revisits grows with its popularity ranking; for the pages with a popularity ranking lower than 10 it is not uncommon that more than a week has passed between two visits.

In order to find differences in how users revisit pages depending on the time span between two revisits, we analyzed the browser tools used for initiating the revisits for four different interval categories: within an hour, within a day, within a week, and longer than a week. From figure 6.17 it becomes apparent that for short-term revisits the back button is the most commonly used tool. For medium-term revisits the page address is directly typed into the address bar; from the interviews it became apparent that the address bar's automatic url completion function, which is offered by most browsers, has become a highly popular alternative 'history' tool that allows users to relocate pages with only partially remembered addresses. However, after a certain period the page is removed from the url completion list. In these situations, if a user does not remember the exact address and if the address has not been bookmarked, users need to rely on *waypoints*, from which a trail to the desired page can be followed. This effect can be observed in figure 6.17 for the revisits after an interval longer than a week.

During an interview session we asked our participants about specific situations in which it was difficult to relocate a specific page. To find possible difficult

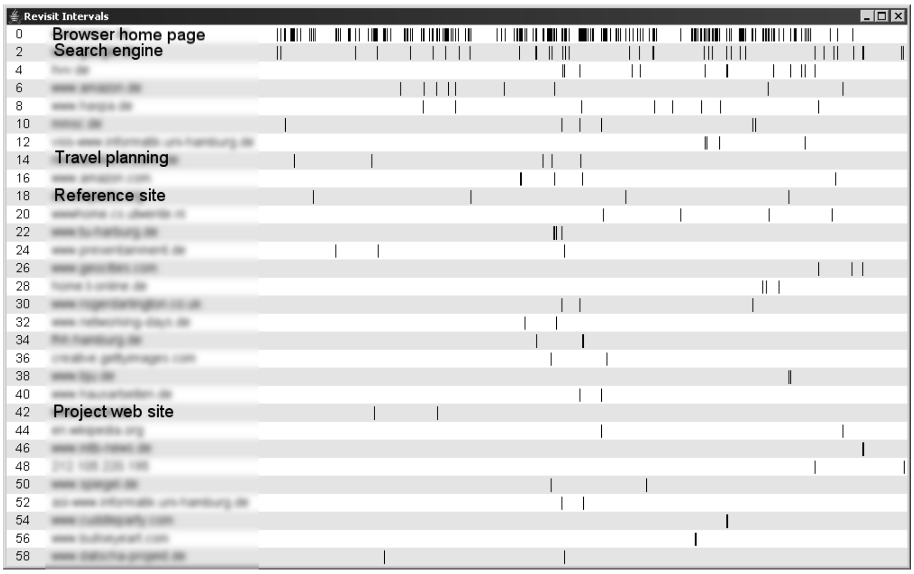

Figure 6.16: *Distribution of revisits over the logging period, for one participant's 60 most popular pages. The time span from left to right is about six months. Page urls have been blurred for privacy reasons.*

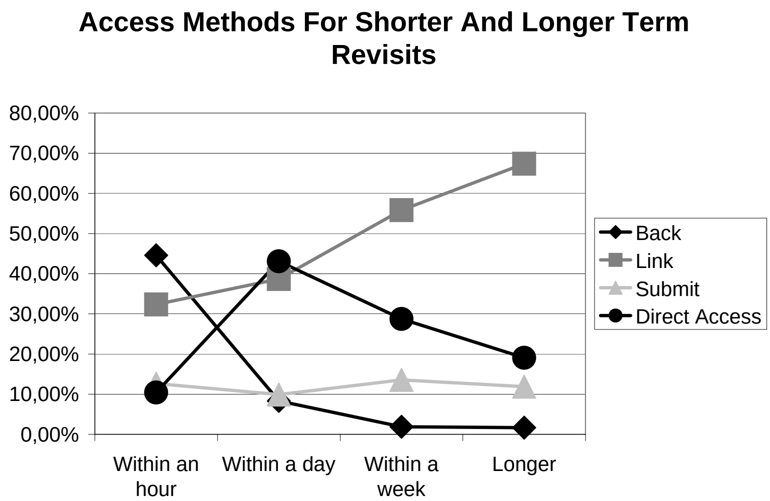

Figure 6.17: *Navigation actions used for page revisits, for different categories of time intervals since the last page revisit.*

situations, we marked pages that were revisited only a limited number of times and with at least a number of days elapsed between the page revisits. We ignored revisits that were likely not to cause any problems, such as pages close to a site's home page and pages that were easily retrievable with a search engine. It turned out that several of these situations were not considered that difficult. Most pages that were considered difficult to relocate provided information to be reviewed rather than a specific tool. Below a list of these kinds of revisits, as mentioned by our participants:

- relocating a physician's personal homepage in a hospital's Web site; the participant did not remember the correct spelling of the person's name;
- a researcher's home page somewhere on a crowded institutional Web site that was not easily retrievable via a search engine;
- a list of soccer results located within an unstructured club site;
- locating newly listed advertisements; the participant missed the possibility to mark to recognize already visited items;
- revisiting certain auctions on an online marketplace; two participants expressed to be annoyed by having to reformulate exactly the same query repeatedly.

Several users reported to repeatedly click through well known paths in order to revisit specific pages. In particular if the target page is used on a semi-regular basis, this strategy of finding a *waypoint* (Capra & Pérez-Quiñones 2003) and following a path from this waypoint is considered inefficient; even if the path is well-known, time and attention is required to reconstruct the path. One participant stated in this context: 'If there would be a faster way, I would use it'.

From the above observations it becomes clear that, even though recurrent behavior is not well supported by a list of most popular pages, it is not very productive to focus on these very popular pages; as these pages are visited on a frequent - often on a daily - basis, users are likely to remember the address at least partially and can easily relocate the page, with help from the url autocompletion function. Ironically, the longer the interval between two subsequent revisits - and the more likely that the user does not correctly remember the exact address - the more likely it is that the address is not present anymore in the browser history.

6.5 Search

Search has become an important activity on the Web. 25% of all Web activities carried out by our participant involved visits to sites that provided search functionality - general search, literature search or other kinds of specialized search.

In this section we analyze our participants' general search activities using the Google search engine. Earlier click-through studies did not consider this aspect of Web navigation; Web search analysis has been conducted separately, making use of search engine's server logs. These analyses have as a disadvantage that they cannot relate the recorded search activity to browsing activities in other sites.

In the first subsection we provide general statistics on user search behavior. We continue with a categorization of the different search actions that users can execute and the query lengths. In the second-last subsection individual differences in search behavior between users are described. We end this section with an analysis of search and recurrent activities.

6.5.1 General statistics

All our participants used Google almost exclusively as main search engine (21,157 visits in total to Google, 405 to Yahoo, 50 to Altavista, 30 to search.msn). In our data analysis we ignored visits to the Google homepage, and only considered visits to the Google result pages, as the strategies to invoke Google queries varied per user:

- some users had Google as their home pages, which would lead to an overestimation of the number of Google visits

- some users accessed Google via the Google toolbar, which would lead to an underestimation of the number of Google visits

- at least one user issued Google queries from a self-edited browser homepage.

As we are mainly interested in how often our participants made use of keyword search and not how often they visited Google, we ignored the visits to the Google home page in order to reduce noise.

12.8% ($\sigma = 6.91$) of the page requests issued by our participants was a visit to any Google search result page. This includes new queries, returning to result lists, requests for additional results and query modifications. Our dataset consisted of 16690 result pages. In 1999 Lau & Horvitz (1999) analyzed data from 4690 queries to the Excite Internet search engine.

Google Web search is by far the most popular search function of Google: 87.5% of all visits to result pages were for www search. Google Image Search accounted for 10.2% of the visits. The remaining 2.3% consisted of visits to Google Groups, Google News, Google Scholar and Froogle.

6.5.2 Query formulation and modification

While searching for information, users make use of different search actions; besides entering new queries they request additional results and refine queries. Following (Lau & Horvitz 1999), we categorized queries along the following categories:

- *New:* a query for a topic not previously searched for by the user within the scope of the dataset;
- *Continuation:* A query on a topic searched on earlier by a user; one or more keywords appeared in earlier queries, but not in the last query;
- *Modification*: one of the following changes have been applied to the query:
  - *Generalization:* A query on the same topic as the previous query, but seeking more general information than the previous query; one or more keywords have been deleted from the query;
  - *Specialization:* A query on the same topic as the previous query, but seeking more specific information than the previous query; one or more keywords have been added to the query;
  - *Reformulation:* A query on the same topic that can be viewed as neither a generalization nor a specialization;
- *Request for additional results*;
- *Return to results*;

Note that, in contrast to Lau & Horvitz (1999), we grouped the three types of query modifications in one general category. The latter category, return to results, was not taken into account by Lau & Horvitz (1999). However, from figure 6.18 it can be observed that returns to the result set occur as often as new queries.

In contrast to the observations by Lau & Horvitz (1999), our participants often modified queries - 21.39% of all visits were a result of modifying an existing query. Actually, this result is not in disagreement with their findings, as splitting the query modification category into its three subcategories - see figure 6.19 - would have resulted in the lower values reported in their study. 29% of all modifications involve specialization actions; query generalization is less popular with 13%. Our participants tended to prefer query modification above requesting additional results - the latter action covered 13.3% of all visits.

6.5.3 Query length

From earlier studies it is known that the average query length is short. Our participants' average query length was 2.57 ($\sigma = 0.54$), which is similar to the earlier findings. Figure 6.20 shows that the most common number of keywords is

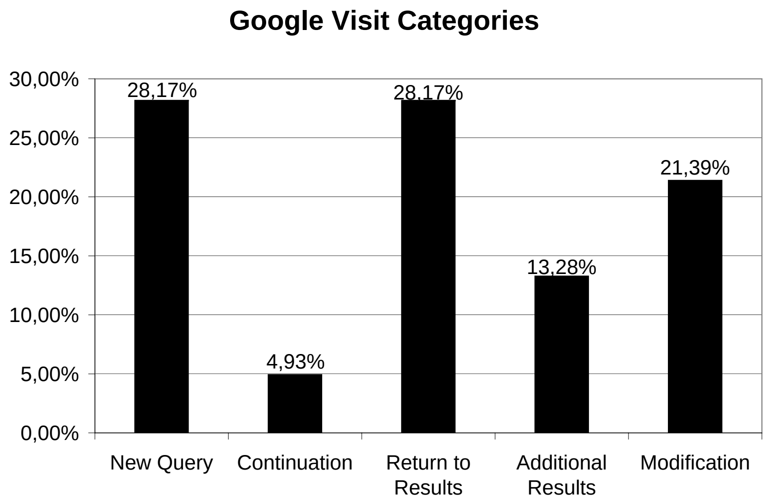

Figure 6.18: *Types of visits to Google result pages.*

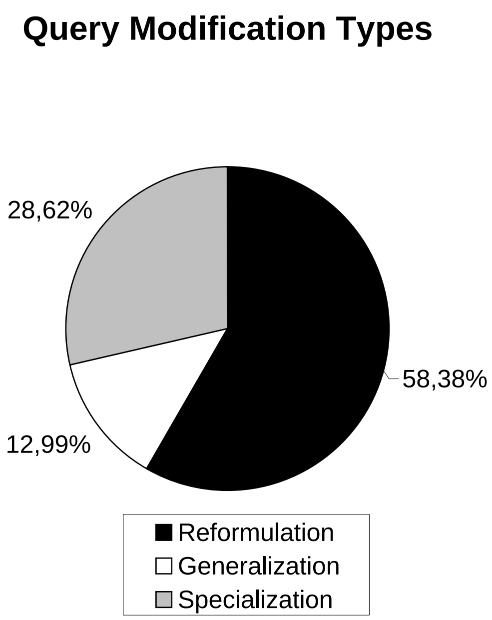

Figure 6.19: *Categorization of query modifications.*

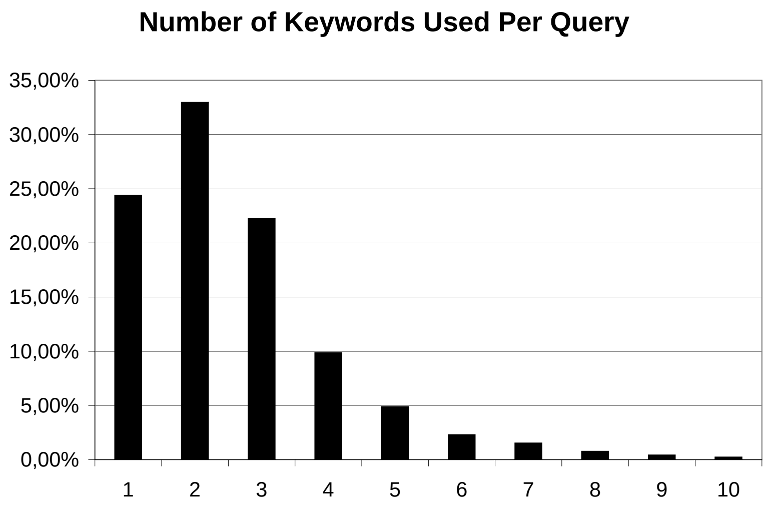

Figure 6.20: *Distribution of query lengths*

two; queries with one or three keywords are very common as well. The maximum number of keywords used by our participants is 24 - this query involved the complete title of a publication and its authors; in fact all searches with more than twelve keywords were either titles or sentences.

The average length of a new query is 2.14 ($\sigma = 0.38$) for our participants, slightly shorter than the average query - see figure 6.20. By their very nature, query specialization and reformulation lead to longer queries. The relatively high average length of generalizations indicates that in most cases a query has already been refined before one starts generalizing. Interestingly, the average query length in the 'additional result' class is below average. This might indicate that users request for additional results mainly in the early phases of their searches - presumably to explore the context and to find candidate keywords for query refinement.

6.5.4 Differences in search strategies

On an individual level we can see interesting differences in Google usage. First, there is a spread in the intensity of search engine use - varying from 2.14% of all page requests to 25.62%. Second, the average query length also differs between users - varying from 1.29 to 2.94 for newly formulated queries. There is no correlation between the intensity of Google usage and the average length of the first (new) query.

Users with longer initial queries tend to modify their queries to a larger extent (r=0.625; $p < 0.01$), in particular by general reformulation (r=0.560; $p < 0.01$).

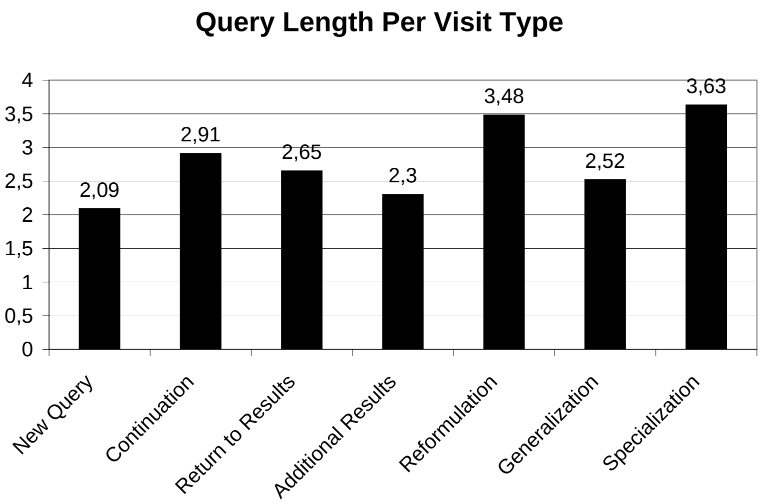

Figure 6.21: *Query length per visit type. All listed values are global averages, not averages per participant.*

At first sight, this is counterintuitive; searchers who already spend more effort on their initial query, might be expected to find what they need without less modifications. We think this reflects a group of users that carefully formulates their queries and rethinks the results; whether this is a personal preference or caused by the nature of the queries, is not clear.

Whereas no correlation between the intensity of Google usage and the average length of the initial query was found, intensive Google users specialize more frequently than less intensive searchers (r=0.628; $p < 0.01$). Intensive searchers also issue more queries (r=0.795, $p < 0.01$), which suggests that there is a group of users that frequently issue a new search and rely on the quick-and-dirty refinement technique of adding keywords.

6.5.5 Search and recurrent behavior

As our participants indicated to use repeated search queries as waypoints, we considered the revisit category of pages navigated to from search result pages. The far majority, 79%, of all page visits following a result page were first-time visits and only a modest amount of 9% search results were followed by a revisit to a page visited in earlier sessions but not yet in the current session - in 24% of these cases the exact query had been used in earlier sessions as well.

In general, repeated queries were rare; only 2% of all queries were queries issued in earlier sessions. The modest contribution of repeated queries to less frequent revisits is confirmed by the fact that only 8% of page revisits with an

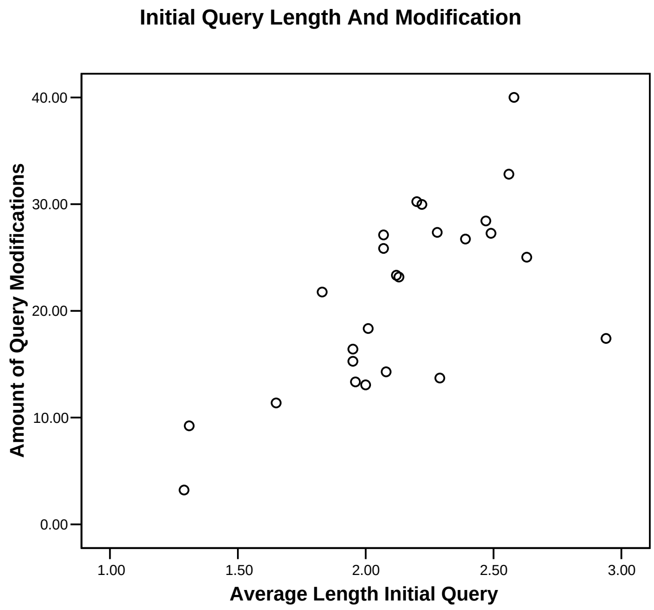

Figure 6.22: *Users with longer initial queries tend to modify their queries more often*

interval longer than a week since the last visit were preceded by a search result page.

The above observation is confirmed by the correlations between the number of revisits and the extent of searching in navigation sessions. The percentage of page revisits is significantly lower in sessions with many visits to result pages (r=-0.359; $p < 0.01$). Interestingly, the percentage of within-session revisits is higher in search-intensive sessions (r=0.133; $p < 0.01$). As indicated in section 2, this is most likely due to users returning to hubs in order to follow an alternative path. These hubs are mainly the result pages themselves, as 31% of all visits to Google involved returns to the result page - a smaller percentage of 27% were the results of a new query.

From the results it can be observed that repeated searches are used only to a limited extent for revisiting pages. In the majority of cases when users do use the search engine for recurrent behavior, a different query than the original has been used. This confirms our participants' remarks that they often could not remember the original query or the exact spelling of a proper name. Further, users often reach less frequently revisited pages by other means than search, which supports the observation that refinding information often involves finding a waypoint and following a path from this waypoint (Capra & Pérez-Quiñones 2003).

In general, our participants did not use many keywords repeatedly in different sessions; the number of keywords used in more than 10 sessions was about 5 for all

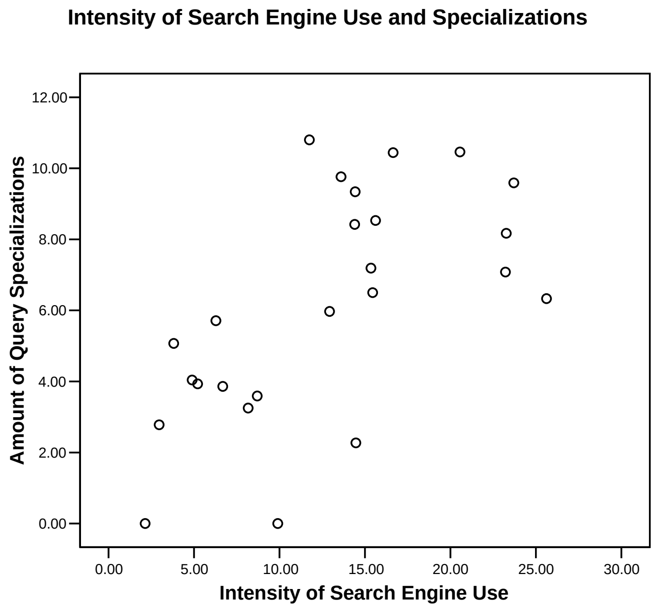

Figure 6.23: *Users who search more often tend to often refine their queries by adding keywords*

participants. In addition, most keywords used in more than 15 individual sessions were either stopwords (of, in, for) or general interest areas (php, agent, hamburg, 2005). However, from the individual top n lists of frequently used keywords, the main activities or interests of the participants became clear.

- Participant 8: hamburg, history, kunst, hypertext, kreativität;
- Participant 19: meeting, evaluation, summarization, text, language;
- Participant 25: xquery, xml, integration, probabilistic, database;
- The author: pirolli, chi, Web, java, user.

6.6 Browser tools used

In this section, we examine the actions that our participants employed to initiate page visits. We call these events *navigation actions*. Apart from selecting links, users can trigger navigation actions in different ways: entering urls directly into the address bar of the browser, using different browser history mechanisms to revisit pages seen before, or submitting information via forms to interactive Web services, such as search engines.

6.6.1 General statistics

The latest reported distributions between the various navigation actions that are based on long-term data date back to studies from 1995 and 1996. The comparison chart - table 6.3 shows some major differences, which reflect both the changed nature of the Web and the way users interact with browser interfaces.

Link following has remained the most common navigation action, accounting for about 45% of all page transitions. Direct access to pages via the bookmark menu, bookmark toolbar - which was not present in previous studies, see figure 6.24 -, home page button or the address bar has remained stable at about 10% as well. The detailed Firefox log as well as the interviews revealed, however, that our users had different preferences to access frequently used pages: some mainly used the bookmark menu, others solely preferred the bookmark toolbar and a few had the custom to type the urls of their favorite pages into the address bar, using its auto-completion function when available. These different behaviors show that customization of the interaction with the browser is necessary. On the other hand it also indicates that none of the current revisitation tools is entirely satisfying (Abrams, Baecker & Chignell 1998) (Jones et al. 2001).

The comparison chart shows a major increase in navigation actions that lead to opening a new browser window. In the mid-1990s, this event accounted for less than 1% of all navigation actions, compared to over 10% nowadays. However, while formerly only the explicit action of opening a new window using the associated menu item was registered , in this study other actions could also result in opening a new browser window. These actions include following hyperlinks with *target="_blank"* as an attribute, starting the browser manually or from other applications, and the use of the 'open link in new window' or 'open link in new tab' entries of the browser's context menu - figure 6.25. Nevertheless, this confirmed that it has become common behavior to open more than one window while browsing the Web.

Accounting for over 15% of all navigation actions, form submission has become a key feature of user navigation as well, as it is a required interaction mechanism with service-oriented Web sites. 43% of all form submissions involve queries to search engines, followed in popularity by an online dictionary and travel planners.

By contrast, the share of back button actions has dropped from over 30% in the mid-1990s to less than 15% presently. This number includes backtracking multiple steps via the back button's pull-down menu, which contributed less than 4% to all backtracking actions. The browser history is not specifically listed in the comparison chart, as it is hardly used - merely 0.2% of all page requests were initiated from the browser history. Only two of our twenty five participants

Figure 6.24: *Bookmark toolbar and browser history in sidebar.*

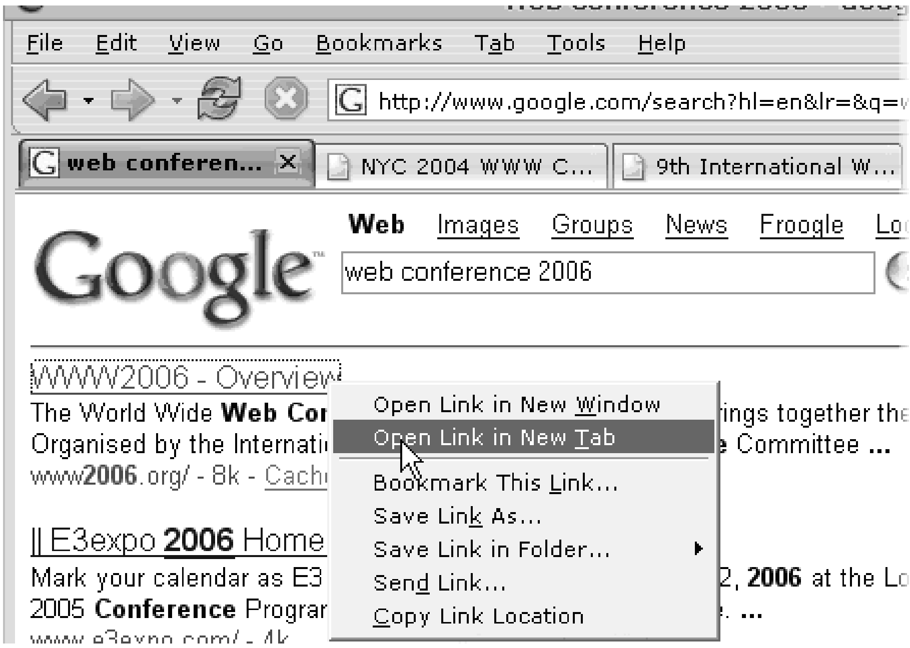

Figure 6.25: *Tabbed browsing: opening a new tab from a link.*

stated to use it from time to time, while ten participants even weren't aware of the function at all.

	Catledge & Pitkow	Tauscher & Greenberg	This Study
Time of study	1994	1995-1996	2004-2005
No. of users	107	23	25
Length (days)	21	35-42	52-195,
No. of visits	31,134	84,841	137,272
Recurrence rate	61%	58%	52.2%
Link	45.7%	43.4%	43.5%
Back	35.7%	31.7%	14.3%
Submit	-	4.4%	15.3%
New window	0.2%	0.8%	10.5%
Direct access	12.6%	13.2%	9.4%
Reload	4.3%	3.3%	1.7%
Forward	1.5%	0.8%	0.6%
Other	-	2.3%	4.8%

Table 6.3: *Comparison chart of three long-term studies*

The reduced usage of the back button, in combination with an increase of 'forward navigation actions' - following links, submitting forms and opening new windows - might indicate that users return less frequently to previously visited pages. However, the recurrence rate - the percentage of page revisits (Tauscher & Greenberg 1997) - decreased to a much lesser extent; from about 60% to 49%. One explanation is that most sites now offer structural links on every page that allow returning to the home page or a landmark page without using the back button. However, there might also be a relation with the increased amount of submit and new window actions. This issue will be explored in the following two subsections.

6.6.2 Form submission and backtracking

The increased number of form submissions confirms another change of the Web: the move from a hypertext information system with mainly static pages to a hybrid between 'classical hypertext' and service oriented interactive systems, such as search engines, dictionaries and travel planners. The latter category of sites is often more similar to desktop applications than to information-centered hypertext: whereas navigation in hypertext involves orienteering behavior with frequent backtracking, interactive applications are mainly used for 'getting things done'. This would imply that backtracking is less prominent during these activities. In order to confirm this hypothesis, we compared the backtracking usage of the top third form submitters of our participants with the remaining participants. The frequent submitters used the back button less frequently (9.2%) than the others (16.2%), a difference that is marginally significant (t=2.715, p=0.012).

Dynamic, interactive pages pose several challenges to the browser's history mechanisms. First, the browser history does not take form submissions using the http POST method into account. Unless the URL of the resulting page explicitly contains the issued parameters, users cannot revisit earlier created documents such as travel plans without going through the process of entering the data again. Once the browser window is closed, the travel plan is lost. The same problem arises for documents that users might want to keep for future reference, such as order confirmations and flight reservations. Unlike in static hypertext, these pages are volatile, even if they contain information that will remain most relevant in the near or more distant future. Whereas users can save or print the information, most participants reported that they almost never did so. The Internet Explorer for Mac OS X features the Scrapbook, which provides an integrated interface for storing an exact copy of the Web page as it appears in the browser window. With the advent of service-oriented sites and volatile pages, similar functionality might be needed in other Web browsers as well. Travel plans, flight reservations and order confirmations should be treated as documents; context-sensitive functionality for storing, retrieving, opening and printing - like in regular office applications - appear to be essential in these situations.

To follow the analogy with office applications even further: as mentioned earlier in this section, the concept of hypermedia navigation is often replaced by a concept of interaction with an application. Hence, navigation tools such as the back button lose their original meaning in these contexts. While using an interactive Web service, the back button bears more similarity to the undo button: users press back to correct errors. To avoid potential disruption of the interaction, some online services disable the use of the back button by opening a pop-up window without navigation toolbars for sequences of interactive forms, or they explicitly advice the user not to use the back button.

In conclusion, browser interfaces miss appropriate functions for service-oriented sites, although these sites play a prominent role in Web usage. We think that one major challenge for the next generation of Web browsers is to reconcile the two different Web usage contexts - hypermedia navigation and interaction with Web-based services.

6.6.3 Multiple windows and the back button

Tauscher & Greenberg (1997) found that most revisits occur within a relatively small distance (measured in page visits). They argued that the back button is responsible for a characteristic pattern that also surfaced in our data - see figure 6.9 -, a high amount of revisits ('spikes') for even distances, generated by the user navigating back to a previously seen page using the back button:

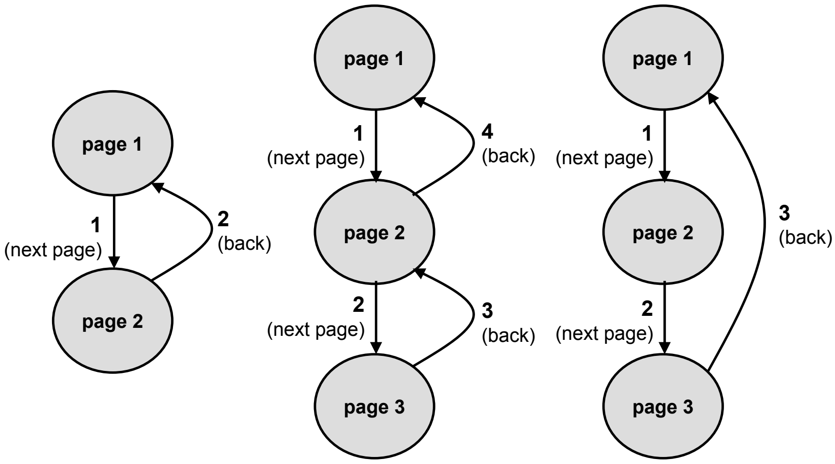

Figure 6.26: *The back button generates even revisit distances. The pull-down menu - on the right - may generate odd revisit distances.*

navigating back a single step generates a revisit with distance 2, navigating back two steps generates two revisits with distance 2 and 4 (fig. 6.26).

These effects did not surface immediately in our data: when we analyzed the click-stream in time order, the spikes were less visible. Our first hypothesis was that the relatively low usage of the back button observed, combined with an increased use of the back button pull-down menu might be responsible, as backtracking over several nodes would be possible (fig. 6.26). Searching for pull-down events, we found that this browser feature was practically not used - only 3.1% of all back events indicated a use of the pull-down menu for our Firefox users. A possible explanation for this neglect of a supposedly useful feature is that it is 'often simpler to just click several times on the back button', as one user reported, than to wait for the pull-down menu to appear and choose the mechanically more difficult activity of selecting from a pull-down menu. Also, the concrete implementation of the back menu that shows the page titles, which is often not very useful for determining page content, could be partly responsible.

In figure 6.27 the contribution of browser interface widgets to revisits on distances one to ten is shown. A first observation is that a large amount of revisits is initiated by following links. The pattern of a relatively large amount of back button actions responsible for revisits at even distances, and a relatively small amount at odd distances, is clearly visible. At odd distances the use of multiple windows explains relatively more of these page revisits. This confirms the observation that traditional hub-and-spoke navigation is partially being replaced by changing between tabs or windows to 'revisit' previously reached pages.

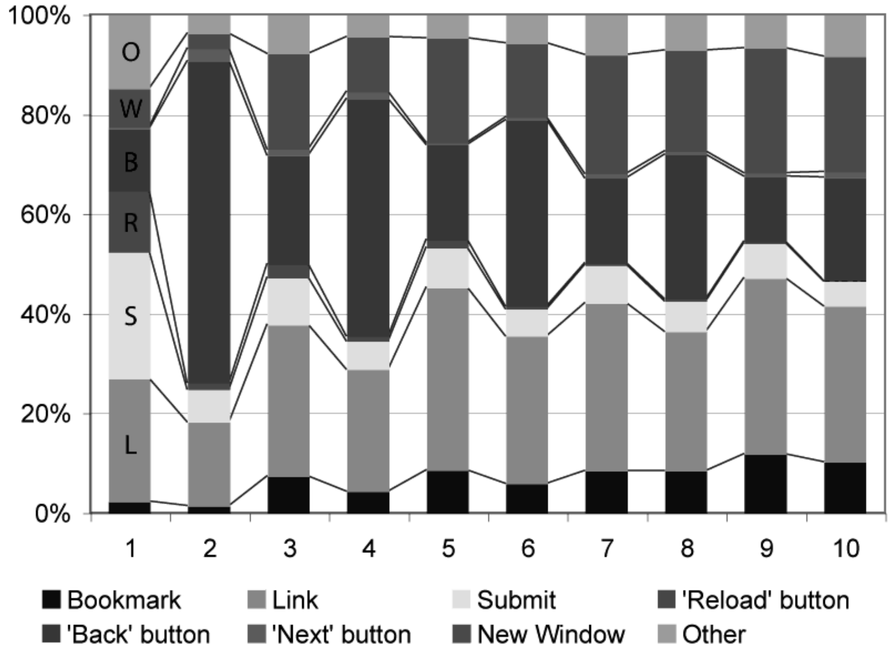

Figure 6.27: *Contribution of interaction types to recent revisits.*

6.6.4 Multiple windows

As reported before, our participants tended to use multiple windows to a large extent. The strategy of opening new windows seems to offer several advantages. Our users reported that multiple windows allowed them to 'compare search results side by side' and that 'pages can be loaded in the background' while they continue their navigation activities. Keeping search results and resulting navigation trails in separate windows also reduces the risk of losing the path to a decisive page because of backtracking to the result page. Several participants also had the habit of keeping a browser window with their favorite news site open in the background.

Ten participants said they frequently use tabs. Several reported that they usually have more tabs than windows open and to use tabs for breadth search - following different search results or comparing items - and prefer multiple windows for non-related parallel tasks. Seven did not know tabs in Web browsers; seven knew them but would not use them - some since they used the Internet Explorer browser - which did not support tabs - others since they prefer the windows concept. One participant mentioned to use both Netscape and Internet Explorer for different parallel tasks in order to easily separate them according to different task bar icons. However, one participant also refrained from using multiple windows since they would clutter the task bar.

By combining the clickstream data - which provides information on which page was loaded at what time - with the additional information on page view

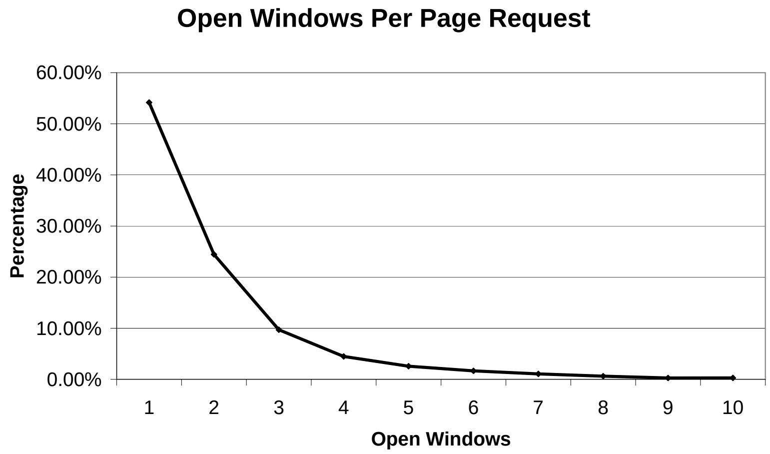

Figure 6.28: *Distribution of windows opened per page request.*

times and the window identifiers, we were able to reconstruct how many windows and tabs were opened at each page request, and which page requests led to the opening of new windows.

Despite the reported increased usage of multiple windows and tabs, still at 54% of all page requests only one window or tab was opened, at 41% of all page request two to five windows were opened - see figure 6.28. However, individual preferences vary. The average number of simultaneously opened windows varied from 1.02 to 4.20, with an average of 1.91 per user ($\sigma = 0.78$).

As mentioned earlier, 10.5% of all navigation actions involved the opening of a new window or tab. From the qualitative observations from the last subsection one would expect that users would typically open new tabs or windows from a search engine result page or when changing to a different Web site. This is confirmed by the higher percentages of new window/tab actions in the log for page requests that follow a Google result - 22.78% - and page requests to a different site than the last site accessed - 23.23%. The negative correlation between the percentage of window switch actions and the amount of backtracking within a session (r.566, $p < 0.01$) confirms that multiple windows are used as an alternative strategy for backtracking. Interestingly, sessions in which users visit more pages that they haven't visited before - supposedly information gathering activities - show a higher amount of window switching (r=.407, $p < 0.05$).

6.6.5 Windows, tabs and backtracking

A comparison of the frequency of backtracking with the usage of multiple windows or tabs showed a correlation between these two navigation actions. The group of participants with the top third of new window events employed the back button to a lesser extent (10.2%) than the bottom third (16.4%); it is plausible that multiple windows are used as an alternative to backtracking (t=2.509, p=0.026). In addition to multiple windows, Firefox provides 'tabbed browsing' - several pages can be opened simultaneously in different browser tabs. Six participants who used tabs frequently, were backtracking less often (9.9%) than the remaining seven Firefox users (18.3%) who hardly opened any tabs (t=2.311, p=0.038). One participant explained he used 'new tabs for closely related tasks and new windows for parallel tasks'.

A more disturbing consequence of the use of multiple windows is that it disrupts the concept of the back button. Its principle functionality is to return to recently visited pages. If users followed trails in multiple windows or tabs, the recent visit history is split into separate stacks, with no temporal relation. Moreover, each individual stack is does not include actions from the originating window. Hence, users need to remember what actions they performed in which window or tab in order to relocate a previously visited page. This places a high cognitive burden on the user, in addition to the already demanding task of keeping track of their location in the Web (Thüring et al. 1995). Handling multiple windows in information systems was already reported to cause disorientation in pre-Web studies (Halasz 1988). The above concerns were confirmed by our participants; several said that they find many open Web documents hard to manage, in particular because the page titles displayed in task bar and tabs are often not helpful.

The more prominent use of multiple windows requires a major rethinking of the history mechanisms of browsers. It provides new challenges to the often criticized (Cockburn & Jones 1996) (Kaasten & Greenberg 2001) yet often used back button. A linear history of most recent revisits, as proposed by Tauscher & Greenberg (1997), does not reflect the character of parallel trails and the unrelated back button stacks do not take the temporal relations between the trails into account.

6.7 Categorization of navigation sessions

Our participants indicated that they seldom ended up without having found what they were looking for; only five participants stated that this happened regularly; of these five participants, two indicated that they are often looking for things that 'no one else is interested in'. The most often mentioned reason for lack of success was that 'the information is not there'. The participants typically identified information gathering activities with search actions; unsuccessful sessions were regularly explained by the fact that they did not know the correct search terms and that search engines provided too many results. Usually, a session will be left uncompleted if a user thinks it is not worth the effort to spend more time on it.

In the previous chapter we showed different navigation styles employed by users in data gathering activities. We found a *flimsy* navigation style that was associated with issues on perceived disorientation, and a *laborious* navigation style that allowed users to build a mental model on the Web site structure. From the extensive log files and the wide variety of tasks that our participants were involved in, it is impossible to find specific episodes in which our participants experienced these kinds of issues. As our participant pool mainly consists of experienced Web users who state to be generally quite successful in finding what they need, we can assume that the majority of navigation sessions can be labeled as 'successful' and use it as a common base for quantitatively identifying differences between users and their sessions. In this analysis we only took into account sessions with a minimum amount of ten page visits, to avoid short sessions, with mainly recurrent behavior, to skew the results.

As mentioned earlier, longer navigation sessions show a higher amount of backtracking and a lower amount of visits to pages from earlier sessions. Based on these impressions, it would be likely to assume that longer sessions are associated with information gathering activities. However, longer sessions are not associated with more intensive use of search engines and there is a tendency to stay longer within a particular site (r=-.305, $p < 0.01$).

A higher amount of first-time page visits is associated with a more intensive use of search engines (r=.350, $p < 0.01$). This suggests that a distinction can be made between search-intensive (information gathering) sessions and exploratory (browsing) sessions can be made. In order to make this distinction, we applied k-means clustering (Mitchell 1997) to the sessions on all navigation measures except the path length with different values for k - with a total of three clusters, two clusters emerged that separated search-intensive sessions from browsing sessions - between 24 and 34 pages in length -, leaving a third cluster of very long sessions. 53% of the sessions were categorized in cluster two. This cluster is characterized by a high percentage of Google result pages (18.4) and a low percentage of revisits (36.04). 28% of the sessions were categorized in cluster one. This cluster is characterized by a low percentage of Google result pages (4.3) and a high percentage of revisits (73.87). Both clusters were similar in terms of navigation complexity measures (see chapter 4 for an overview of these measures): the values for path compactness and stratum were average, with values between 50% and 60% and the average return rates were similar as well. The average number of pages visited in one site was higher in cluster one (t=4.187, $p < 0.01$). The remaining sessions that were categorized in cluster three contained far more page visits than the first two clusters; as expected, these longer sessions are less linear and show more backtracking activities. Due to the low sampling rate, no differences in these longer sessions can be found. The navigation measures of the three categories are summarized in table 6.4.

Looking for individual differences, we compared the session means between users. The results confirm that users who regularly switch between windows

Cluster	1	2	3
Session length	24.5	33.5	123.7
Average pathlength in site	4.90	4.09	5.81
Percentage of site switch actions	22.39	25.85	14.33
Percentage of window switch actions	18.81	21.73	10.87
Percentage Google result pages	4.3	18.4	10.7
Within-session revisits	37.9	25.7	46.2
Cross-session revisits	35.94	10.37	9.73
Total amount of revisits	73.87	36.04	55.92
Percentage back button used	10.2	11.6	18.0
Percentage first-time visits	40.65	85.23	80.01
Fan degree (Rauterberg)	1.81	1.40	1.97
Cycles (Rauterberg)	10.8	9.8	56.8
Path density (Rauterberg)	0	0	0
Path compactness	55.5	50.8	77.0
Path stratum	59.37	60.45	21.47
Average return rate	2.10	2.07	2.97
Average connected distance	3	5	8

Table 6.4: *Navigation measures of three session clusters*

backtrack to a lesser extent than users who do not switch windows very often (r=-.566, $p < 0.01$). Further, users who backtrack more often tend to make use of navigational hubs, as expressed by the *average return rate* (r=.704, $p < 0.01$).

From the results it becomes apparent that sessions of average length can be categorized as either focused on search activities to find new information or on browsing activities. In the first category - representing 58% of all sessions - users make intensive use of the search engine and visit more pages that they haven't visited before. In the latter category - representing 28% of all sessions - users revisit pages to a large extent, which are both recurrent activities and backtracking. These numbers correspond to a large degree with the clusters of activities identified by (Sellen et al. 2002) from interviews with knowledge workers. Information gathering and information finding activities, which accounted for 59% of all user activities, were considered as the most important activities. The remaining activities involved leisure browsing, transacting, housekeeping and communicating - activities that are likely to account for the patterns found in the 'browsing' cluster.

6.8 Discussion

The results from this long-term study show a wide variety of patterns in Web usage. Perhaps the most important observation is that there is no such a thing as a 'typical Web user' or a 'typical Web navigation session'. The Web is used for many purposes, varying from work-related information gathering to recreational browsing and usage of Web applications. This large spread of activities translates into a huge variety of navigation behavior.

In contrast to popular belief, most Web activities take place in long navigation sessions, covering 34 page requests or more. This means that in most situations the user context can be estimated from the activities earlier in the session. From the last section it has become clear that sessions can be categorized as focused on information gathering or focused on browsing. The latter category is likely to involve leisure browsing and commonly reoccurring tasks. Shorter sessions show a large amount of recurrent activities, focused on a small set of frequently visited pages.

An important observation is that the Web has evolved from a hypermedia system to a 'hybrid' between hypermedia and interactive tools and applications. This has several consequences for user-adaptive support. When users fill in forms to obtain a service or a product, hypermedia browsing tools lose their original meaning; the back button bears more similarity to the undo button: users press back to correct errors. In addition, pages that are served in response to the submitted forms - such as travel plans, reservation confirmations and receipts - are volatile, whereas it is likely that users would want to store it for future reference.

In the remainder of this section we will concentrate on hypermedia navigation activities. In particular, implications of our results for the design of tools that support backtracking and recurrent activities will be discussed. We will end this chapter with some concluding remarks.

6.8.1 Support for backtracking

Currently, the back button is the most frequently used tool for backtracking activities. In addition, users often backtrack to pages visited before by using the links in the Web page. Most backtracking activities aim to return to a navigational hub, which is predominantly either an index page within the same Web site or a list of search results. As an alternative strategy, users open a new window or tab to follow several trails from a hub. This has as an advantage that users do not need to backtrack and that they can compare the trails side by side. A major disadvantage of this strategy is that the back button stack is split into several shorter stacks, which requires users not only to remember the past activities that led to the current page, but also what activities were carried out in which window. This places a high cognitive burden on the users.

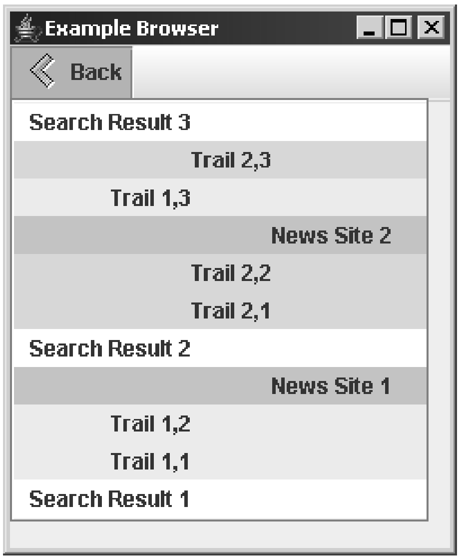

Figure 6.29: *Example branching history view for four browser tabs. Page visits are ordered by time, from bottom till top. Each tab has a unique associated color, and indenting of the page title. This approach allows for scanning of navigation both activities in all tabs, and of activities in one tab.*

Whereas the back button only would support about 50% of all backtracking activities, a simple list of the 15 most recently visited pages would contain about 90% of all revisited pages. This would call for replacing the back button by such a list that shows the most recently visited pages in temporal order, regardless in which window these pages were visited. However, this does not do justice to the (semi-)parallel character of the trails followed in different windows. An alternative solution would be a *branching* history, which shows the trails in temporal order, but which separates the activities in different windows. An example branching history is shown in figure 6.29. However, from the laboratory studies in the previous chapter it has become clear that users who backtrack using the site's link structure are often more successful than users who simply 'push back' until they reach the page they want to revisit. This can be explained by the fact that users who effectively backtrack using the within-site links, have a better mental model of the site structure. Therefore, rather than to encourage using a separate list of most recently visited pages, it might be better to use adaptive hypermedia techniques such as link annotation to visually identify the links that point to recently visited pages, in particular those pages that serve as hubs. Various heuristics can be used to identify these hubs (Milic-Frayling et al. 2004), such as the number of outgoing links and the opening of a new window or tab from the page.

To summarize, theoretically backtracking is well-supported by the very simple mechanism of most recently visited pages. The main issue is how to visualize this list. Based on the findings in our study, we propose a combination of a branching history and integration of within-page links and the recent history list by means of link annotation, to support a user's building and exploiting an effective model of the site structure.

6.8.2 Support for recurrent behavior

Users have a large variety of pages that they visit on a regular basis, of which a small number very frequently. Whereas the majority of our participants had bookmarks as a shortcut to the small set of very popular pages - often positioned in the browser toolbar -, many relied on the browser's automatic url completion, which is shown to be an effective tool for supporting these revisits. An explanation for this discrepancy is that our participants indicated that the list of bookmarks often was incomplete and outdated. This discrepancy raises the question whether manually managed bookmarks as a shortcut to frequently visited pages is the right solution for supporting frequent revisits, if these frequent visits can be identified automatically. We found out that the list of 15 most frequently visited pages is fairly stable; of the 15 most popular pages in the first quartile of the study, on average 11.16 ($\sigma = 2.78$) are in the list of the last quartile as well.

For less frequent revisits, users often need to rely on finding waypoints that lead to the desired page, in particular if the address is not present anymore for automatic completion in the address bar. Web search appears to be an ineffective manner to find these waypoints, in particular as users seem not to be able to replicate the exact query. Our participants expressed their annoyance if they have to follow a well-known trail to a page that is used on a semi-regular basis. Given the share amount of these kind of pages, a manually organized list of bookmarks will most likely be incomplete - let alone the time it would take to manage such a list. A strategy that seems to be more effective to support orienteering behavior for relocating information, would comprise:

- a search history that recognizes queries similar to queries issued before, and provides these related queries and result pages;
- annotation of trails followed before from identified waypoints, such as search results, which would allow users to recognize the steps followed in this trail more easily;
- for frequently followed trails, a shortcut to the destination page would save much effort from the user. According to (Tauscher & Greenberg 1997) the number of longer repeated trails is very limited.

This would call for a long-lasting browsing history, either at the user's computer or on a remote server. This may cause privacy issues - users who share the same computer most likely don't want to share their complete Web usage history. It is beyond the scope of this thesis to find acceptable solutions in this respect, but it is clear that a trade-off is needed between what is being stored and functionality provided. From the growing popularity of Web-based email providers it is likely that users will become less sensitive on this matter. This is exemplified by a remark on a bulletin board regarding the mail functionality provided by Google: 'you may be monitored, but there is no one watching'.

6.8.3 Concluding remarks

While the Web and Web usage has seen dramatic changes in the past few years, the user interface of current Web browsers and their integrated navigation support tools closely resemble those of browsers from the early days of the Web. Most of the changes are of technical nature, such as the integration of plug-ins and JavaScript, allowing new content types and a greater degree of interaction, but not directly relating to browsing support. This is at least partially due to lack of empirical studies; many aspects of how people use their Web browsers today are either unknown or unproven. This study was aimed to fill this gap and to provide a substantial amount of background knowledge for interface design, user modeling and Web personalization.

Though the recorded data of this study is extensive and detailed, it has its limits, as it misses contextual information. The extent of logging was restricted technically, as data capture was limited to the browser, ignoring related software that was used in conjunction with the Web client, such as word processors, e-mail agents and other office applications. Furthermore, the data of click-stream logs have a limited expressiveness; user tasks and user goals often stay below the surface. This makes their contextual interpretation inherently difficult and additional qualitative information is needed to support a detailed task-related evaluation of the data. The two 90-minute interviews conducted at the beginning and at the end of the study could only deliver scarce data for a substantial qualitative analysis.

Although we tried to recruit participants with very different backgrounds, all were frequent computer users with long internet experience. Still, the variance in the captured data was fairly large for almost all aspects of navigation: the number of visits, page vocabulary and use of search engines differed between the individuals. Also, the applied navigation habits, especially to directly access documents and to revisit pages varied. However, the remarkable differences in this small participant group did already reveal that Web browsers are used with various personal preferences and that individual users have particular demands.

Chapter 7

Conclusions

The research reported in this thesis was aimed to improve our knowledge on user interaction with the Web. Current navigation support - as offered by the link structure of Web sites, and browser tools for backtracking and recurrent activities - is known to have several limitations. Several lines of research target these issues with new concepts for Web site adaptation, personalized search, relevance feedback, backtracking mechanisms and history overviews, to name a few. However, none of these concepts has been integrated in current browser interfaces or Web site designs. This is partly caused by technological issues and privacy considerations; we think the main obstacle for the development of enhanced navigation support concepts is insufficient insight in how users currently interact with the Web.

In this thesis, we adopted a research approach closely related to approaches seen in the field of adaptive hypermedia; however, we did not immediately assume that personalization would provide the answer to all usability issues - there is no clear border between improved designs, such as support for backtracking, and 'real' user adaptivity. From the literature overview on Web navigation studies in chapter 3 it became clear that Web navigation is a complex of searching, browsing and backtracking activities, which together establish a behavior that has been termed *orienteering*. The gathering of information in order to compare, choose, decide about, or use for some purpose, is generally considered more important than leisure browsing. In addition, the Web is often used for recurrent activities, which are not very well supported by current browsers. From the literature it remained unclear what patterns in information gathering can be found, and what patterns are more likely to lead to better results; likewise, it remained unclear what recurrent activities users carry out on the Web, and how users interact with their Web browsers.

In this chapter we summarize the findings from the literature research and user studies carried out to fill this gap. We start with returning to the research questions, as formulated in the introduction. We continue with a discussion on the theoretical insights from our research. In the third section we provide pointers for future work on Web navigation support that better addresses the users and their contexts. We end this chapter with a discussion and some concluding remarks.

7.1 Answers to research questions

In this section we summarize our findings on user navigation patterns on the Web. We concentrate on the new findings and findings that were different from earlier research. For the presentation of the results we follow the five challenges as formulated in the introduction of this thesis.

What kinds of generalizations can we draw from patterns of user interaction with the Web?

From the long-term study it has become clear that the Web evolved from a static hypermedia system to a dynamic 'hybrid' between hypermedia and interactive tools and applications. Whereas the metaphor of Web navigation still applies to a large amount of activities, interactive services like travel planning bear more similarity to interacting with desktop applications. This observation might seem obvious from introspection of Web use, but the lack of research on support for these interleaving activities indicates that it has not received sufficient attention from the adaptive hypermedia and World Wide Web communities [1].

Users often revisit pages, but to a lesser extent than suggested by earlier studies. The overall majority of these revisits (74%) involves backtracking activities to pages visited before in the same session. A list of the 15 most recently visited pages covers these activities in about 90% of all these cases, far more than the coverage of the back button. Revisits to pages from earlier sessions were considered recurrent activities. The page popularity distribution follows a power law distribution; the most popular pages are portals for sites such as search engines and news sites. In contrast to these sites, frequently visited institutional and project Web sites have a broader range of pages that are visited multiple times. Access to the most popular pages is quite well supported, as the addresses are often quite easy to remember and visited on a frequent basis; most browsers offer URL completion functionality in these situations. Revisits to a broad range of pages that are visited on a less frequent basis are poorly supported; users often do not remember the exact address, and ironically browsers do not 'remember' the address either. Users often resorted to clicking through a range of pages to relocate the page. Due to its sheer number, an overly long list of bookmarks would be needed to solve this situation.

Information gathering activities are characterized by more intensive use of the search engine. In contrast to earlier findings, users often modified their queries.

[1]This remark is given by an anonymous reviewer of (Weinreich et al. 2006, forthcoming)

This bias might be caused by our participant pool; however, there is evidence that it is more likely that the server-based approach used in Web search studies did not succeed in capturing this effect; the query categorization used by (Lau & Horvitz 1999) masked the importance of query modifications. Queries are seldom repeated, and only a small amount of keywords is repeated.

From server-side Web usage studies it was concluded that the far majority of sessions is very short. Our client-side data shows a different pattern; the median length of a session is 13, and the far majority of pages is visited in sessions longer than 34 pages. This is good news for adaptive hypermedia researchers, as this means that in many cases a great deal of the current context can be derived from activities carried out before in the same session.

What individual differences in user access patterns can we find in these interaction patterns?

From the laboratory studies different navigation styles emerged, which were closely associated with usability issues related to disorientation. These issues were reflected in a weak, *flimsy* navigation style, characterized by a frequent use of the back button, frequent returns to a site's home page, and long trails. This reflects situations in which users were not able to effectively make use of the site structure. Periods of *laborious* exploration appear to help in constructing a mental model of the site structure. With additional context information, user navigation becomes more structured - page revisits are more often initiated by within-page links, and more focused on a small number of navigation hubs. Although we could not analyze this effect in our long-term study, we expect that the same differences can be found in real-life Web usage.

The long-term study showed a wide difference in recurrent behavior and search engine use. Several participants had their own 'sticky' sites, such as their club's bulletin board or their favorite news site, other participants had not. Users who made more intensive use of the search engine, tended to prefer the quick-and-dirty approach of keyword addition for query refinement. Users with longer initial queries tended to refine their queries more often, in particular by careful reformulation.

The use of multiple windows or tabs is adopted by most users. However, the extent to which multiple windows are used varies dramatically between users. Multiple windows are often used as an alternative for backtracking, and to compare several trails next to one another. A further reason for using multiple windows is that users are working on parallel tasks - or keeping their favorite news site opened in the background.

What user actions and user goals can we predict from these patterns?

As most page requests take place in a longer session, it is well possible to determine to what extent users are looking for new information or carrying out recurrent tasks, such as leisure browsing, housekeeping and transacting. Information gathering activities are often associated with an increased number of search actions, including a substantial percentage of query modifications. This information can be used to create a current user interest profile.

From the spread of page popularity distribution and the low amount of query and keyword repetitions it appears that long-term user interests are hard to retrieve in an automated manner; the variety of Web activities is too large to extract more than some general main interests (section 6.5.5). In addition, recent research has shown that even if a user interest profile is known, personalization of search often leads to decrease in performance (Teevan, Dumais & Horvitz 2005).

As discussed before, frequently reoccurring activities can be distilled quite easily from the page popularity ranking. However, predicting less frequent page visits is likely not to be much more than a wild guess without additional information from the user's real life context. By contrast, revisits that involve backtracking activities can be covered to a large degree by a list of the most recently visited pages.

What kinds of changes could we make by design improvements and personalization techniques?

The above results suggest that inference of long-term user interests, based on their Web activity, will result in overly general profiles that only cover a limited part of the users' Web activities. Additional user modeling techniques, as described in chapter 2, would be needed for this purpose. However, it is well possible to take the current user context into account for enhanced browser support - situations in which the user is looking for new information, carrying out recurrent tasks, interacting with Web services, or making use of multiple windows.

In addition, the current user session, as well as earlier sessions, provide a wealth of context information that can be used for enhanced navigation support. Less frequent visited pages can be easier located using the concept of *waypoints*, and information gathering can be supported by more sophisticated search and browsing history concepts. Many of these concepts require little more than rather straightforward inferences from the user's interaction history. However, this does not mean that the concepts should remain unexplored. On the contrary, straightforward inferences have the benefit that little uncertainty is introduced. Furthermore, context-aware adaptations that directly address issues in Web navigation are likely to have a higher impact than adaptations based on rather indirect evidence. Later on in this chapter we will formulate several concepts for future navigation support, based on the findings in our studies.

Can we evaluate the impact of these techniques from changes in user interaction with the Web?

In the second laboratory study we explored the impact of task-based link annotations on user navigation styles. The results indicate that these kinds of added support have their impact on user navigation styles, and suggested an approach to detect these changes in laboratory settings. The changes found in navigation styles were related to, and found consistent with, existing knowledge of Web navigation and usability issues related to disorientation.

Further, we have seen that the use of browser interface elements can be analyzed and interpreted in a long-term study. As an example, we showed that the use of multiple windows has a measurable impact on the use of the back button. We also showed the theoretical benefits of the recency-ordered history list, based on the long-term Web usage data. As mentioned before, a drawback is that some results might appear obvious and merely confirmations of common intuition. On the other hand, in order to motivate the need to address these issues 'hard' evidence is required instead of intuitions.

7.2 Theoretical insights

In this section we relate our results to existing insights in Web navigation. First, we discuss how we can benefit from a broad corpus of various different perspectives on user interaction with the Web. We continue with a discussion on the use of navigation styles for the evaluation of hypermedia systems. The section concludes with several observations on client-side analysis of Web navigation: issues that need to be addressed, and the relation with server-side analysis of navigation in one particular Web site.

7.2.1 Models of Web navigation

From the literature study a wide spectrum of aspects associated with Web navigation emerged. This is not surprising, given the wide variety of activities on the Web that we found in our study. This confirms the common belief that there is no such thing as a 'typical Web user' or a 'typical Web navigation session'. This means that in Web site design and browser design these differences, as well as general patterns, should be taken into account. The same yields for theoretical models of Web navigation, which mainly take goal-oriented information gathering activities into account. Results from our long-term study indicate that different usage situations can be derived from simple statistics on the amount of revisits and search engine queries.

Theoretical models, such as information foraging and its associated SNIF-ACT cognitive architecture, aim to model the average Web user, who is assumed to react on the relevance of the navigation options provided by the current page.

From the comparison of the theoretical models with empirical studies it appeared that the role of backtracking in user navigation is mostly ignored in these models, or even considered as recovering from a failure. This does not correspond to the high amount of backtracking observed in empirical studies. Our laboratory studies showed that various patterns in backtracking activities can be distinguished, of which in particular focused, hub-oriented patterns are related to successful navigation.

A particularly interesting finding from the first laboratory study is the role of *laborious* exploration; episodes of apparently random and meaningless back and forth navigation that help in building a mental model of the information environment. Existing theoretical frameworks do not take this behavior into account, or consider it as 'failure'. Exploration is a natural element of Web navigation, even in goal-directed activities, and essential for getting to know 'what's out there'.

The abundance of query modifications observed in our long-term study support the concept of *orienteering*, as identified by Teevan et al. (2004) from semi-structured user interviews. Orienteering involves a series of steps of gradually further specifying the actual goal, based on the context gathered during the process. Orienteering suggests that the actual user goal may well be a moving target. In addition, we observed that users often worked on several related or unrelated tasks in parallel.

Theoretical models are useful as a guide for hypermedia research, for the interpretation of empirical findings and the development of implicit user models. However, they are necessarily simplified abstractions of the variety of user actions on the Web. Instead of striving to one integrated theory of Web navigation, we think it is more desirable to strive toward a comprehensive body of empirical data and theoretical insights that each provide different views on Web navigation. For empirical data it is important to understand the context in which it was derived - the participants and their tasks, the study setup and data capturing methods, the technology used, the goals of the study - which might bias the interpretation of the results -, to name a few (Hawkey & Kellar 2004). Similarly, for theoretical frameworks it should be made explicit in what situations they are applicable, by specifying at least:

- the type of Web navigation tasks it covers; information finding, information gathering, recurrent activities;
- the variety of users and navigation strategies that can be represented by the declarative and procedural knowledge;
- whether the model covers parallel tasks and moving goals;
- whether the model incorporates the user navigation history;
- what assumptions are made on the browser's interface;

Conflicting results and predictions are unavoidable when using a variety of reference frameworks. We believe this is far from disadvantageous; it provides indicators where further investigation is necessary, or what assumptions should not be made in a user model for adaptive systems, or in a general interaction design process. In order to pinpoint these situations, overviews and discussion of these frameworks is needed.

A particular issue that needs to be addressed is the physical metaphor of Web *navigation*. This metaphor clearly does not take the parallel character of multiple windows, and the interaction with Web applications into account. In many situations the concepts of moving back and forward in an information space are not sufficient to describe the interaction with the Web. Further theoretical work is needed to relate hypermedia navigation to WIMP (Windows, Icons, Menus and Pointers) interfaces (Dix et al. 2004), and to provide metaphors that better reflect the interleaved navigation and interaction activities.

7.2.2 Navigation styles as evaluation measures

A general goal of user-adaptivity and improved navigation support for information gathering and knowledge acquisition is to reduce the complexity of interaction (Jameson 2003). The provision of additional context may not directly help in improving users' knowledge on the domain, but reduction of the cognitive overhead that is required for the interaction process will allow the user to concentrate on the task itself. As we have seen in chapter 2 and 3, several techniques can be used for this purpose, including personalized history mechanisms, collaborative filtering, overview maps, and link annotation based on the user's level of knowledge. Many of these techniques have been explored in adaptive hypermedia systems, but it is still an open question which technique works best in what situations. Due to the open-ended character of information gathering, adaptation success often cannot be defined as error reduction, or speed improvement. Subjective ratings have been shown to have severe limitations as well, as users have no baseline for comparison (Weibelzahl 2003).

By lack of objective measures, one has to resort to *theory assessment measures*. In our laboratory studies we explored the use of second-order *navigation styles*, based on a 'battery of measures' (Otter & Johnson 2000), to characterize differences in user disorientation while *orienteering* (Teevan et al. 2004). The approach is quite similar to the *behavioral complexity* measures proposed by Weibelzahl (2003) for general, system-independent assessment of adaptive systems. We identified a number of navigation styles, of which the *flimsy* navigation style was highly associated with disorientation; episodes of *laborious* exploration helped in improving orientation. We showed that the additional context provided by task-based link annotations made user navigation less flimsy, and more purposeful.

The concept of navigation styles has several benefits over (first-order) complexity measures. They reflect more detailed aspects than, for example, the number of cycles or the amount of revisitation. Based on the components from which they are built, their meanings can be interpreted; this interpretation step adds some subjectivity to the navigation styles, but with proper methodology - various independent interpretations, visualization techniques, references to empirical and theoretical models - this effect can be minimized. The fact that we consistently found the same four styles in both laboratory studies, suggests that flimsy and laborious navigation can be used as baselines for comparing systems for information gathering. For other kinds of activities the baselines are likely to differ: Gwizdka & Spence (2005) adopted our approach in a laboratory study on indirect assessment of Web navigation success. Their results suggest that success in more goal-directed information finding tasks are characterized by shorter, more linear navigation paths.

The results booked with the approach show that navigation styles are an effective concept to identify usability issues and to evaluate the effect of alternative designs. However, currently they are associated with an intensive interpretation process. More studies on different types of tasks are required to find effective sets of baselines for various kinds of tasks, to alleviate and possibly automate the interpretation process.

7.2.3 Client-side Web usage analysis

Given the high interest in server-side Web usage mining for the evaluation, improvement, and personalization of site designs, the number of reported client-side long-term studies is surprisingly low. This imbalance is reflected in the lack of changes in the browser interface, compared with the dramatic changes in server-side technology, services provided and site designs. Currently, numerous extensions for the Firefox browser are developed by companies and hobby enthusiasts alike (Mozilla Foundation 2006), but most of the added functionality provided does not satisfy essential user needs - the extensions are merely technology-driven. For user-driven interface changes, empirical data is needed. From our experiences in the preparation phase of our long-term study, privacy issues are a severe hindrance for carrying out these kinds of studies; even at work, browsing the Web is considered a personal activity.

The results from our long-term study showed that a fair amount of Web traffic involves page requests not initiated by users, but as a side effect of another page request. The majority of these artifacts involved advertisements, commonly embedded in a page in the form of an *iframe*; the average percentage of advertisements recorded at participants who did not use an adblocker was over 28%. In addition, several sites embed code in their pages, to instruct them to reload after some period - for example to update the news content. In many cases, the browser window involved is left unattended, or placed in the background. These kinds of artifacts may severaly skew results from content-based analysis, and may cause pages or sites to appear far more heavily visited than they actually are. In particular proxy-based usage mining systems and Web intermediaries should take these issues into account. We presented several heuristics to identify the artifacts involved.

Our client-side data shows that server-side observed Web usage patterns cannot necessarily be generalized into 'user interaction with the Web'. As an example, we found that the length of Web sessions is considerably longer than often assumed (Baldi et al. 2003); also, users often modify their queries, an effect that is obscured in server-side log analysis. Further, Web usage patterns involve more than a click-through log can reveal: information on the browser interface elements used and the use of multiple windows is essential to interpret the findings. Earlier long-term studies mainly focused on patterns, generalized over the whole participant pool, and individual navigation sessions were not taken into account. The results from our study strongly indicate that the *diversity* between participants and sessions is equally important to investigate. Further, enrichment of the Web log by various annotations turned out to be essential during the analysis process; these annotations do not only facilitate the actual mining, but also stimulate curiosity and creativity.

7.3 Directions for navigation support

One thing that became apparent from the long-term study is the wild variety of queries issued by users, and the associated variety of pages and sites visited. Whereas general tasks and user interests could be detected, these were often topics that were either very obvious or very general. This implies that great care should be taken with pro-active techniques such as user notification. The same yields for constructing a user interest model for search activities.

7.3.1 User support for information gathering

Information finding and information gathering involve the combined activities of searching, browsing, and the reading of content. Search is used to create a selection of information patches, in which the user can navigate. Often, several query refinement steps are required to reformulate or refine the information goal, based on the context gathered while browsing. Research in the field of information retrieval aims to improve the search part of the gathering process.

Currently, the query refinement process is not supported by search engine interfaces. This is likely caused by the common misconception that users do not modify their queries very often. However, our - admittedly, advanced - long-term study participants often modified their queries, except for 'trivial' general interest search. Several lines of research are directed to automatic suggestions for query expansion, based on semantic similarity or collaborative mechanisms. In addition, we think the query modification history would help in supporting the refinement process: feedback on which keywords have been added, deleted or modified during the process, and the effect on the search results, is likely to be effective query refinement support.

Search engine results show a list of pages that best match the given query. When leaving the search engine, the only thing the user knows is that the page of arrival is supposed to be relevant. But would there be any surrounding pages that also match the query? How profitable would it be to further explore a site? The only way to find this out is by actually spending time on navigating the site. The decision to do so, or to leave the site, is largely uninformed and based on the proximal cues provided. One approach to improve the information scent is explored in the ScentTrails system (Olston & Chi 2003), which provides link annotations based on the destination's relevance to the query. A disadvantage of this approach is that it does not provide a general indication of the relevancy of complete site or subparts of the site. This could be solved by search engines not only providing relevancy indicators of one particular page, but also indicators of the relevance of the surrounding pages, site parts, and the site in general. In particular in the initial phase of orienteering, a higher site relevance is likely to be more profitable than the high relevance of one particular page within the site.

7.3.2 Support for backtracking

Backtracking - revisiting pages from the current session, not necessarily initiated by the back button - is an important aspect of Web navigation. Backtracking allows users to revisit content and to follow several alternative trails from one point - this is called hub-and-spoke navigation. Due to the stack-based behavior of the back button, and due to the fact that the stack is cleared upon closing the browser window, only a little more than 50% of all backtracking activities would be supported by the back button. Our long-term study indicates that the list of 15 most recently visited pages covers about 90% of all backtracking activities.

This observation supports the design concepts of the recently developed SmartBack button (Milic-Frayling et al. 2004). SmartBack lists pages in order of their last visit, with duplicates removed. In addition, it marks pages that form the beginning of a trail, navigation hubs, and search result pages for quicker access. A user study indicated great savings in navigation effort required for backtracking. It should be noted that the SmartBack behavior is quite similar to the traditional browser history list, which is hardly used. The unpopularity of the history list is likely caused by he misconception that it is mainly useful for distant past activities - which is not the case; by contrast, the history list would be very practical for instant page access in backtracking activities.

As shown in the last chapter, the use of multiple tabs and windows undermines the functionality of the back button, and would also undermine the functionality of the SmartBack button. When users carry out several tasks in parallel windows, the history list is split into several lists, each associated with the corresponding window. This causes additional cognitive overhead: in order to revisit a page, users need to remember in which window it was visited. A temporally ordered list that integrates all page visits in all windows would solve this problem, but introduces another one: it does not do justice to the (semi-)parallel character of trails. A branching history, which shows the trails in temporal order, but that visually distinguishes activities in separate windows (figure 6.29, page 189) would solve this issue. It is an open question how an integrated multiple-window 'back' or 'history' should behave exactly. Should it open the page in the same window as it was opened before, or should it be opened in a new window? Several alternatives should be explored.

A further consideration is the observation from our laboratory studies that more successful users backtrack using within-page links rather than using the back button. Most likely, these users better understand the site structure. This would mean that comprehension of the site structure would be enhanced by, in addition to the 'back list', marking the corresponding links and pages:

- mark already visited links - this is already part of the common 'web vocabulary', but is violated by many sites, in particular in graphical menus;

- mark links that lead to already visited pages - links that point to pages already visited are currently treated similarly to already visited links; they are different;

- mark links to trail beginnings, hubs, and search results.

7.3.3 Supporting recurrent behavior

In addition to backtracking, users also revisit pages for recurrent activities, such as reading the news or booking a trip. In particular those activities that take place on an infrequent basis are often not supported by a browser history mechanism, whereas in particular in these situations support is needed. We saw that the number of less frequently occurring activities is too long to be manageable with a bookmark list; also, the variety is too large to make reliable predictions.

In the Stuff I've Seen project (Dumais, Cutrell, Cadiz, Jancke, Sarin & Robbins 2003) an advanced search interface for desktop search is developed, which is used for retrieving documents, images, music, and any other kind of file that is created, edited, saved and opened on the computer. The interface provides free keyword search in combination with setting parameters on time period, possible file types, authors. Desktop search relies on a long-term interaction history, and we expect that the paradigm will particularly be useful for relocating documents (pages) on the Web.

Teevan et al. (2004) and Capra & Pérez-Quiñones (2003) found that a common strategy to relocate items, is to rely on *waypoints*, pages that are known to lie on the path to the desired page, and of which they remember the address or a query that leads to the page. As explained in the last chapter, we believe that in addition to desktop search functionality, annotation of earlier followed trails from these waypoints would be a most welcome browser history mechanism. Like annotation of recently followed trails for backtracking, a color-coding would be sufficient - with details on demand.

7.3.4 Adapt to the user's task context

The Web has evolved from an information-oriented static hypermedia system to a hybrid of hypermedia and interactive tools and services, such as travel planning and online banking. Hypermedia navigation tools - back, forward and home - loose their meaning in the latter kinds of applications, and the use of these tools is likely to disrupt the interaction with the service. The outcomes of these services - such as travel plans and receipts - are volatile, even though it is likely that users want to keep them for future reference. Many sites solve this issue by sending a copy of the receipt via email. These interactive applications ask for an office-like interface, with toolbars for the opening and saving of documents, and an undo button. On the other hand, for information-oriented hypermedia navigation this functionality would not be very useful.

We think one of the major challenges in Web interface design is to reconcile these two situations. A hidden *save page* option in a menu is not sufficient, as it is not obvious for users to use it. Moreover, as it is no integrated functionality, users would have to use a different application to relocate a stored page on their hard disk; a travel planner site would not be able to recognize that a user saved a travel plan, in order to provide a pointer to this plan upon the next visit. Server-side solutions are not ideal either, as this results in a myriad of different solutions. We think the following two steps are needed to solve the matter:

- an integrated toolbar with both support for hypermedia navigation and interactive applications; irrelevant options should be disabled in situations when they are not relevant or disruptive, by greying them out or perhaps even by hiding them. This can be achieved by increasing site control on the browser functionality;
- there should be a protocol for document management, which supports manual and automatic local saving of pages such as travel plans and receipts; like in regular office applications, there should be one point that a user can go to for referencing a saved travel plan or to make a new one;

In summary, the navigation and document management options as provided by sites and browser should be integrated into one interface concept, which allows for consistent interaction protocols. Due to the variety of needs associated with different usage situations, the support should be adaptive to the task context. This will require rethinking the metaphor of Web navigation into an interleaving process of navigation and interaction.

7.3.5 Summary - the future browser

To summarize the design implications presented in this chapter, and to provide a more vivid image of how the various aspects of the user's recent and past interaction history will improve user interaction with the Web, we sketch an imaginary 'future browser'.

An unmistakable trend is the move from an information-oriented hypermedia repository toward a dynamic, interactive network of information and services. In order to support the interleaving activities of hypermedia navigation and interaction with services, future browsers should extend the Web navigation interface metaphor with document management tools. The future browser should be sensitive to the current user context: during periods of interaction with Web-based services, such as an order check-out, hypermedia navigation tools - like the back button - are disabled or hidden, as they are essentially meaningless. Alternatively, the functionality of the back and forward button will be slightly different in these situations, and used for switching between consecutive forms. Further, the future browser will have an intuitive interface for storage, printing and retrieval of documents. When a user visits a travel planning site, references to the last requested schedules are available on the home page; in addition, earlier schedules can be retrieved using mechanisms such as file browsers from this point.

The future browser provides radically different history support tools. The current back button is replaced by a strictly temporally organized back button, which also supports cross-window backtracking activities. In addition, within Web pages links that point to items visited in the same session are differently visualized than other links; recently followed links are marked as a trail. Frequently reoccurring activities are supported by a simple top n list, accessible from a menu or a toolbar. Less frequent visits are supported by annotated trails and shortcuts from known waypoints. An important advantage of the use of trails is that they support the users in understanding the link structure on the Web, and the paths that they have paved.

In addition to the search button, future search engines should provide the option to explicitly search for 'something I have visited before'. As users often have issues in exactly repeating an earlier query, results from related past queries can be taken into account. The future search engine should particularly support the information gathering process by showing the user's query modification history, and by providing relevancy indicators for a site as a whole; this would enable the user to better estimate the benefits of further digging into an information patch.

The above implicitly suggests that we expect that future interaction with the Web will continue to involve extensive navigation, in particular in information gathering activities. With the growing importance of Web-based information and services in everyday life, orienteering behavior and additional context is needed to ensure that one gets what one needs, and not what one asked for.

Given the inherently rather lengthy character of information gathering activities, it is expected that many Web interactions will continue to take place from a desktop setting, with comfortable input devices and screens, a comfortable chair, and coffee at hand. However, with the proliferation of mobile devices, most likely certain Web activities will also be carried out in mobile settings. During a recent event of the World Wide Web Consortium (2005*b*), main players in the field identified the following promising areas, based on the current - still limited - mobile use: *location-based service discovery* using yellow pages and maps; *quick reading* of regular sources and *infotainment*; and *sharing current events* using mobile blogs and photo albums. Due to the inherent challenges of mobile contexts, the W3C's Mobile Web guidelines strongly advice to design for quick transfer, minimal user input, various screen sizes, and goal-directed usage (World Wide Web Consortium 2006). To reach these goals and to enable efficient navigation, adequate support for backtracking and review of earlier visited pages - or stored pages, like travel plans and maps - is needed, possibly even more than in desktop settings.

It should be noted that in the above image we purposely left out many other future developments. It is likely that our knowledge on proper site design will continue to improve, and that site designs will become more standardized. New applications will emerge. On a local, site level, targeted personalization techniques - like social navigation, recommendations, localization, and learner-adaptive courses - will become more developed. However, in addition to these developments, we believe that improvement of the browser interface is of extreme importance, as well as more sophisticated usage of the user's interaction history.

7.4 Concluding remarks

In this thesis we explored many aspects of user interaction on the Web, to enhance our knowledge on the matter, and to provide pointers for better support for these activities. We found that Web interaction varies wildly between users, and even more wildly between different tasks. We did not find concrete evidence that sophisticated personalization or user profiling techniques would help in supporting Web navigation in general. Instead, we found that the user's recent and long-term interaction history can be put into use for supporting information gathering, recurrent behavior, and backtracking in fairly straightforward ways. Several directions have been suggested in this chapter. The current challenge is to experiment with various alternative design concepts, and to eventually integrate them in the browser interface.

At this point it is unclear which solutions would work best. In an initial phase the effect of navigation support for information gathering or relocating information can be evaluated in controlled studies, and changes in the user navigation styles can be used as a theory assessment measure, in combination with qualitative observations and feedback. The next phase would comprise investigating the actual adoption and use of these concepts in real-life. Currently, this is mainly done by recruiting a fairly limited pool of participants from the own research institute. We believe that it would be beneficial to involve the general audience in the process, for example by developing and distributing extensions for open source browsers (Mozilla Foundation 2006). This provides the opportunity to receive a wealth of feedback, and statistics on the adoption of the concept do not only indicate its usability, but also whether it will actually be used (Dix 2005).

Research in the field of adaptive hypermedia is mainly restricted to closed environments, typically one web site serving a particular goal. The benefits of course adaptation, movie recommendations, and many other personalizations based on inference from a user profile have been shown in various research projects. In closed environments, with knowledge on the document structure and the user goals, sophisticated personalization techniques are an area well worth exploring. We believe that general Web navigation does not provide enough evidence to put these techniques into use. This does not mean that less sophisticated adaptation mechanisms or straightforward use of the interaction history should not be explored. We believe that the development and evaluation of personalized - or enhanced context-aware - support concepts for general Web interaction is an area that definitely deserves attention from the adaptive hypermedia community, as well as from user interface researchers in general.

This leaves the question whether improvement of Web interfaces, based on rather straightforward inferences from interaction patterns may be called 'personalization'. It does fit Brusilovsky's definition of adaptive hypermedia:

> By adaptive hypermedia we mean all hypertext and hypermedia systems which reflect some *features of the user* in a *user model* and apply this model to *adapt various noticeable aspects of the system* to the user.

Perhaps the question is not that important at all. Offering users personalized interfaces that better support their individual needs is just one strategy for dealing with a large variety of users and their tasks. A one-size-fits-all design for diversity might reach the same goal as well.

Bibliography

Abrams, D., Baecker, R. & Chignell, M. (1998), Information archiving with bookmarks: Personal web space construction and organization, *in* 'Proc. CHI98'.

Ahuja, J. & Webster, J. (2001), 'Perceived disorientation: An examination of a new measure to assess web design effectiveness', *Interacting with Computers* **14 (1)**, 15–29.

Anderson, C. (2002), A Machine Learning Approach to Web Personalization, PhD thesis, University of Washington.

Anderson, C. & Horvitz, E. (2002), Web montage: A dynamic personalized start page, *in* 'Proc. 11th WWW Conference', pp. 704–712.

Anderson, J. & Lebiere, C. (2000), *The atomic components of thought*, Lawrence Erlbaum Associates.

Andrews, K., Kappe, F. & Maurer, H. (1995), 'The hyper-g network information system', *J. UCS* **1 (4)**.

Baldi, P., Frasconi, P. & Smyth, P. (2003), *Modeling the Internet and the Web*, John Wiley & Sons.

Barrett, R., Maglio, P. & Kellem, D. (1997), How to personalize the web, *in* 'Proc. Conf. on Human Factors in Computing Systems', pp. 75–82.

Benyon, D. (1993), 'Adaptive systems: a solution to usability problems', *User Modeling and User-Adapted Interaction* **3 (1)**, 65–87.

Berendt, B. & Bernstein, E. (2001), 'Visualizing individual differences in web navigation: Stratdyn, a tool for analyzing navigation patterns', *Behavior Research Methods, Instruments & Computers* **33 (2)**, 243–257.

Berners-Lee, T. (1999), *Weaving the Web - The original design and ultimate destiny of the World Wide Web by its inventor*, Orion business Press.

Bieber, M., Vitali, F., Ashman, H., Balasubramanian, V. & Oinas-Kukkonen, H. (1997), 'Fourth generation hypermedia: Some missing links for the world wide web', *Int. J. Human-Computer Studies* **47**, 31–65.

Blackmon, M., Polson, P., Kitajima, M. & Lewis, C. (2002), Cognitive walkthrough for the web, *in* 'Proc. CHI 2002', pp. 463–470.

Botafogo, R., Rivlin, E. & Shneiderman, B. (1992), 'Structural analysis of hypertexts: Identifying hierarchies and useful metrics', *ACM Trans. Information Systems* **10 (2)**, 142–180.

Broder, A. (2002), 'A taxonomy of web search', *SIGIR Forum* **36 (2)**, 3–10.

Broder, A., Kumar, R., Farzin, M., Raghaven, P., State, R., Tomkins, A. & Wiener, J. (2000), Graph structure in the web, *in* 'Proc. 9th WWW Conference', pp. 247–256.

Brown, P. (1987), Turing ideas into products: The guide system, *in* 'Proc. Hypertext 87', pp. 33–40.

Browne, D., Totterdell, P. & Norman, M. (1990), *Adaptive User Interfaces*, Academic Press.

Brusilovsky, P. (1996), 'Methods and techniques of adaptive hypermedia', *User Modeling and User Adapted Interaction* **6 (2-3)**, 87–129.

Brusilovsky, P. (2001), 'Adaptive hypermedia', *User Modeling and User-Adapted Interaction* **11**, 87–110.

Brusilovsky, P. (2004), 'Adaptive navigation support: From adaptive hypermedia to the adaptive web and beyond', *PsychNology Journal* **2**, 7–23.

Brusilovsky, P., Callaway, A. & Nürnberger, A., eds (2005), *Proc. PIA 2005 Workshop on New Technologies for Personalized Information Access, at User Modeling 2005*.

Brusilovsky, P., Karagianidis, C. & Sampson, D. (2001), The benefits of layered evaluation of adaptive applications and services, *in* 'Proc. Workshop Empirical Evaluation of Adaptive Systems (at UM2001)', pp. 1–8.

Brusilovsky, P. & Tasso, C. (2004), 'Preface to special issue on user modeling for web information retrieval', *User Modeling and User-Adapted Interaction* **14**, 147–157.

Bush, V. (1945), 'As we may think', *The Transatlantic Monthly* **July**.

Buyukkokten, O., Garcia-Molina, H., Paepcke, A. & Winograd, T. (2000), Power browser - efficient web browsing for pdas, *in* 'Proc. CHI 2000'.

Byrne, M. (2003), *Human-Computer Interaction Handbook*, Lawrence Erlbaum, chapter Cognitive Architecture, pp. 97–117.

Byrne, M., John, B., Wehrle, N. & Crow, D. (1999), The tangled web we wove: A taskonomy of www use, *in* 'Proc. CHI 99'.

Capra, R. & Pérez-Quiñones, M. (2003), Re-finding found things: An exploratory study of how users re-find information, Technical report, Computing Research Repository.

Catledge, L. & Pitkow, J. (1995), 'Characterizing browsing strategies in the world wide web', *Computer Networks and ISDN Systems* **27 (6)**, 1065–1073.

Center for the Digital Future (2005), 2005 digital future report - highlights, Technical report, USC Annenberg School.

Chen, C. (1999), *Information Visualization and Virtual Environments*, Springer-Verlag.

Chen, C. (2000), 'Individual differences in a spatial-semantic virtual environment', *J. American Society for Information Science* **51 (6)**, 529–542.

Chen, S. & Macredie, R. (2002), 'Cognitive styles and hypermedia navigation: Development of a learning model', *Journal of the American Society for Information Science and Technology* **53 (1)**, 3–15.

Chi, E. (2002), 'Improving web usability through visualization', *IEEE Internet Computing* **6 (2)**, 64–71.

Chi, E., Rosien, A., Supattanasiri, G., Williams, A., Royer, C., Chow, C., Robles, E., Dalal, B., Chen, J. & Cousins, S. (2003), The bloodhound project: Automatic discovery of web usability issues using the infoscent simulator, *in* 'Proc. Chi 2003'.

Claypool, M., Le, P., Waseda, M. & Brown, D. (2001), Implicit interest indicators, *in* 'Proc. Intelligent User Interfaces 2001'.

Cockburn, A., Greenberg, S., KcKenzie, B., Jasonsmith, M. & Kaasten, S. (1999), Webview: A graphical aid fro revisiting web pages, *in* 'Proc. OZCHI'99'.

Cockburn, A. & Jones, S. (1996), 'Which way now? analysing and easing inadequacies in www navigation', *Intl. J. Human-Computer Studies* **45 (1)**, 105–129.

Cockburn, A. & McKenzie, B. (2001), 'What do web users do: An empirical analysis of web use', *Intl. J. Human-Computer Studies* **54 (6)**, 903–922.

comScore Media Metrix (2004), 'Surfing down memory lane to january 1996: comscore media metrix revisits first-ever web site rankings'.
URL: *http://www.comscore.com/press/release.asp?press=434*

Conklin, J. (1987), 'Hypertext: An introduction and survey', *IEEE Computer* **20 (9)**, 17–41.

Cooley, R. (2003), 'The use of web structure and content to identify subjectively interesting web usage patterns', *ACM Trans. Internet Technology* **3 (2)**, 93–116.

Cooley, R., Mobasher, B. & Srivastava, J. (1999), 'Data preparation for mining world wide web browsing patterns', *Knowledge and Information Systems* **1 (1)**.

Davis, H., Hall, W., Heath, I. & Hill, G. (1992), Towards an integrated information environment with open hypermedia systems, *in* 'ACM ECHT Conference'.

De Bra, P., Aerts, A. & Rousseau, B. (2002), Concept relation types for aha! 2.0, *in* 'Proc. AACE ELearn'2002', pp. 1386–1389.

De Bra, P., Aerts, A., Smits, D. & Stash, N. (2002), Aha! version 2.0, more adaptation flexibility for authors, *in* 'Proc. AACE ELearn'2002 Conference', pp. 240–246.

De Bra, P. & Houben, G. (1998), Hypertext metrics revisited: Navigational metrics for static and adaptive link structures.

Dhyani, D., Ng, W. & Bhowmick, S. (2002), 'A survey of web metrics', *ACM Computing Surveys* **34 (4)**, 469–503.

Di Battista, G. (1999), *Graph Drawing: Algorithms for the Visualization of Graphs*, Prentice-Hall.

Di Iorio, A. & Vitali, F. (2004), 'Writing the web', *J. Digital Information* **5 (1)**.

Dill, S., Kumar, R., McCurley, K., Rajagopalan, S., Sivakumar, D. & Tomkins, A. (2002), 'Self-similarity in the web', *ACM Trans. Internet Technology* **2 (3)**, 205–223.

Dillon, A. & Vaughan, M. (1997), '"it's the journey and the destination": Shape and the emergent property of genre in evaluating digital documents', *New Review of Multimedia and Hypermedia* **3**, 91–106.

Dix, A. (2005), *Handbook of Human Factors in Web Design*, Lawrence Erlbaum Associates, chapter Human-Computer Interaction and Web Design, pp. 28–50.

Dix, A., Finlay, J., Abowd, G. & Beale, R. (2004), *Human-Computer Interaction, 3rd edition*, Prentice Hall Europe.

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R. & Robbins, D. (2003), Stuff i've seen: A system for personal information retrieval and re-use, *in* 'Proc. SIGIR '03', pp. 72–79.

Eirinaki, M. & Vazargiannis, M. (2003), 'Web mining for web personalization', *ACM Trans. Internet Technology* **3 (1)**, 1–27.

Eveland, W. J. & Dunwoody, S. (1998), 'Users and navigation patterns of a science world wide web site for the public', *Public Understanding of Science* **7**, 285–311.

Facca, F. & Lanzi, P. (2005), 'Mining interesting knowledge from weblogs: a survey', *Data & Knowledge Engineering* **53**, 225–241.

Field, A. (2005), *Discovering Statistics Using SPSS - 2nd Edition*, Sage Publications.

Frana, P. (2004), 'Before the web there was gopher', *IEEE Annals of the History of Computing* **March**, 20–41.

Furnas, G. (1997), Effective view navigation, *in* 'Proc. CHI 1997'.

Greer, J. & Philip, T. (1997), Guided navigation through hyperspace, *in* 'Proc. workshop Intelligent Educational Systems on the World Wide Web'.

Grossman, D. & Frieder, O. (1998), *Information Retrieval - Algorithms and Heuristics*, Kluwer Academic Publishers.

Gwizdka, J. & Spence, I. (2005), Indirect assessment of web navigation success, *in* 'Proc. CHI 2005', pp. 1427–1430.

Halasz, F. (1988), 'Reflections on notecards: Seven issues for the next generation of hypertext systems', *Communications of the ACM* **31 (7)**, 836–852.

Halasz, F. (1991), "seven issues': revisited', Closing Plenary at Hypertext 91.

Hawkey, K. & Kellar, M. (2004), Recommendations for reporting context in studies of web browsing behaviour, Technical report, Faculty of Computer Science, Dalhousie University, Halifax, Canada.

Heer, J., Card, S. & Landay, J. (2005), prefuse: a toolkit for interactive information visualization, *in* 'Proc. CHI 2005'.

Heer, J. & Chi, E. (2002), Separating the swarm: Categorization methods for user sessions on the web, *in* 'Proc. CHI 2002', pp. 243–250.

Herder, E. (2003), Utility-based evaluation of adaptive systems, *in* 'Proc. Workshop Evaluation of Adaptive Systems at UM2003'.

Herder, E. (2004), Sniffing around for providing navigation assistance, *in* 'Proc. German Workshop on Adaptivity and User Modeling in Interactive Systems ABIS2004'.

Herder, E. (2005), Characterizations of user web revisit behavior, *in* 'Proc. German Workshop on Adaptivity and User Modeling in Interactive Systems ABIS2005'.

Herder, E. & Juvina, I. (2004), Discovery of individual user navigation styles, *in* 'Proc. Workshop on Individual Differences in Adaptive Hypermedia, at AH2004', pp. 40–49.

Herder, E. & Van Dijk, B. (2003), From browsing behavior to usability matters, *in* 'Proc. 10th Intl. Conf. Human-Computer Interaction (vol. 3)', pp. 452–455.

Herder, E. & Van Dijk, B. (2005), *Adaptable and Adaptive Hypermedia Systems*, IRM Press, chapter Site Structure and User Navigation: Models, Measures and Methods., pp. 19–35.

Herder, E. & Weinreich, H. (2005), Interactive web usage mining with the navigation visualizer, *in* 'Proc. Chi 2005'.

Herder, E., Weinreich, H., Obendorf, H. & Mayer, M. (2006, forthcoming), Much to know about history, *in* 'Proc. Adaptive Hypermedia 2006'.

Herman, E., Melançon, G. & Marshall, M. (2000), 'Graph visualization and naivgation in information visualization: A survey', *IEEE Trans. Visualization and Computer Graphics* **6 (1)**, 24–43.

Hollink, V., Van Someren, M. & Ten Hagen, S. (2005), Discovering stages in web navigation, *in* 'Proc. User Modeling 2005'.

Höök, K. (2000), 'Steps to take before intelligent user interfaces become real', *Interacting with Computers* **12 (4)**, 409–426.

Horvitz, E., Breese, J., Heckerman, D., Hovel, D. & Rommelse, K. (1998), The lumiere project: Bayesian user modeling for inferring the goals and needs of software users, *in* 'Proc. Conf. Uncertainty in Artificial Intelligence'.

Ivory, M. & Megraw, R. (2005), 'Evolution of web site design patterns', *ACM Trans. Information Systems* **23 (4)**, 464–497.

Jameson, A. (2003), *Human-computer interaction handbook*, Erlbaum, chapter Adaptive Interfaces and Agents, p. 305330.

Jones, W., Bruce, H. & Dumais, S. (2001), Keeping found things found on the web, *in* '10th Int. Conf. Information and Knowledge Management', pp. 119–134.

Jungnickel, D. & Schade, T. (2000), *Graphs, Networks and Algorithms*, Springer.

Juvina, I. & Herder, E. (2005), The impact of link suggestions on user navigation and user perception, *in* 'Proc. User Modeling 2005'.

Juvina, I. & Van Oostendorp, H. (2004), Individual differences and behavioral aspects involved in modeling web navigation, *in* 'Proc. User Interfaces for All 2004'.

Juvina, I. & Van Oostendorp, H. (Submitted), 'Cognitive model generating goal-specific web navigation support', *Interacting with Computers* .

Juvina, I., Van Oostendorp, H., Karbor, P. & Pauw, B. (2005), Toward modeling contextual information in web navigation, *in* 'Proc. CogSci 2005'.

Kaasten, S. & Greenberg, S. (2001), Integrating back, history and bookmarks in web browsers, *in* 'Proc. Chi 2001', pp. 379–380.

Karagiannidis, C. & Sampson, D. (2000), Layered evaluation of adaptive applications and services, *in* 'Proc. Adaptive Hypermedia 2000', pp. 343–346.

Karger, D., Katz, B., Lin, J. & Quann, D. (2003), Stickey notes for the semantic web, *in* 'Proc. Intelligent User Interfaces 2003'.

Katz, M. & Byrne, M. (2003), 'Effects of scent and breadth on use of site-specific search on e-commerce web sites', *ACM Trans. Computer-Human Interaction* **10 (3)**, 198–220.

Kay, J. (1993), *Adaptive User Interfaces - Principles and Practice*, North-Holland, chapter Pragmatic User Modelling for Adaptive Interfaces, pp. 129–147.

Kay, J., Kummerfeld, R. & Lauder, P. (2002), Personis: a server for user models, *in* 'Proc. Adaptive Hypermedia 2002', pp. 203–212.

Kellar, M., Watters, C., Duffy, J. & Shepherd, M. (2004), Effect of task on time spent reading as an implicit measure of interest, *in* 'Proc. ASIS AM 04'.

Kelly, D. & Belkin, N. (2004), Display time as implicit feedback: Understanding task effects, *in* 'Proc. SIGIR 2004'.

Kelly, D. & Teevan, J. (2003), 'Implicit feedback for inferring user preference: A bibliography', *ACM SIGIR Forum* **37 (2)**, 18–28.

Kitajima, M., Blackmon, M. & Polson, P. (2000), 'A comprehension-based model of web navigation and its application to web usability analysis', *People and Computers* **XIV**, 357–375.

Kleinberg, J. (1999), 'Authoritative sources in a hyperlinked environment', *Journal of the ACM* **46 (5)**, 604–632.

Kobsa, A., Koenemann, J. & Pohl, W. (2001), 'Personalized hypermedia presentation techniques for improving customer relationships', *The Knowledge Engineering Review* **16 (2)**, 111–155.

Kobsa, A. & Pohl, W. (1995), 'The user modeling shell system bgp-ms', *User Modeling and User-Adapted Interaction* **4 (2)**, 59–106.

Lau, T. & Horvitz, E. (1999), Patterns of search: Analysis and modeling ofweb query refinement, *in* 'Proc. User Modeling 1999', pp. 119–128.

Lee, M., Montgomery, A., Shapiro, S., Shah, V. & Want, Q. (2000), The world wide web as an engineering paradigm, Technical report, MIT 6.933/STS.420.

Lieberman, H. (1995), Letizia: An agent that assits web browsing, *in* 'Proc. IJCAI-95'.

Linderholm, T. (2002), 'Predictive inference generation as a function of working memory capacity and causal text constraints', *Discourse Processes* **34 (3)**, 259–280.

MacGregor, S. (1999), 'Hypermedia navigation profiles: Cognitive characteristics and information-processing strategies', *J. Educational Computing Research* **20 (2)**, 189–206.

Mayer, M. (2000), Contextual web visualization: Browsingicons and browsinggraphs to support orientation and navigation in the world wide web., *in* 'Proc. 6. Tagung der Deutschen Sektion der Internationalen Gesellschaft fr Wissensorganisation', pp. 267–280.

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

McNee, S., Lam, S., Konstan, J. & Riedl, J. (2003), Interfaces for eliciting new user preferences in recommender systems, *in* 'Proc. User Modeling 2003', pp. 178–187.

Miles-Board, T., Carr, L. & Hall, W. (2002), Looking for linking: Associative links on the web, *in* 'Proc. Hypertext 2002', pp. 76–77.

Milic-Frayling, N., Jones, R., Rodden, K., Smyth, G., Blackwell, A. & Sommerer, R. (2004), Smartback: Supporting users in back navigation, *in* 'Proc. WWW 2004'.

Miller, C. (2004), Modeling information navigation: Implications for information architecture. Submitted Manuscript.

Miller, C. (2005), Modeling web navigation: Methods and challenges, *in* 'Proc. Workshop Intelligent Techniques for Web Personalization at IJCAI2003', pp. 37–52.

Mitchell, T. (1997), *Machine Learning*, The McGraw-Hill Companies, Inc.

Morrison, J., Pirolli, P. & Card, S. (2000), A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions, Technical report, UIR Technical report UIR-R-2000-17.

Mozilla Foundation (2006), 'Firefox add-ons'.
URL: *https://addons.mozilla.org/?application=firefox*

Mukherjea, S. & Foley, J. (1995), 'Visualizing the world-wide web with the navigational view builder', *Computer Networks and ISDN System* **27 (6)**, 1075–1087.

Nelson, T. (2000), 'Xanalogical structure, needed now more than ever: Parallel documents, deep links to content, deep versioning, and deep re-use', *ACM Computing Surveys* **37 (4es)**.

Network Working Group (2005), 'Uniform resource identifier (uri): Generic syntax'.
URL: *http://www.gbiv.com/protocols/uri/rfc/rfc3986.html*

Nielsen, J. (1998), 'Personalization is over-rated', Alertbox.
URL: *http://www.useit.com/alertbox/981004.html*

Nielsen, J. (2003), 'Information foraging: Why google makes people leave your site faster', Alertbox.
URL: *http://www.useit.com/alertbox/20030630.html*

Nielsen, J. (2005*a*), 'Mental models for search are getting firmer', Alertbox.
URL: *http://www.useit.com/alertbox/20050509.html*

Nielsen, J. (2005*b*), 'One billion internet users', Alertbox.
URL: *http://www.useit.com/alertbox/internet_growth.html*

Nielsen, J. (2005*c*), 'The power of default values', Alertbox.
URL: *http://www.useit.com/alertbox/defaults.html*

Nielsen, J. (2005*d*), 'Reviving advanced hypertext', Alertbox.
URL: *http://www.useit.com/alertbox/20050103.html*

Obendorf, H. & Weinreich, H. (2003), Comparing link marker visualiztion techniques - changes in reading behavior, *in* 'Proc. WWW Conference', pp. 736–745.

Obendorf, H., Weinreich, H. & Hass, T. (2004), Automatic support for web user studies with scone and tea, *in* 'Proc. CHI 2004'.

Olston, C. & Chi, E. (2003), 'Scenttrails: Integrating browsing and searching on the web', *ACM Trans. Computer-Human Interaction* **10 (3)**, 177–197.

Oppermann, R. (1994), 'Adaptively supported adaptability', *Int. J. Human-Computer Studies* **40**, 455–472.

Orwant, J. (1995), 'Heterogenous learning in the doppelganger user modeling system', *J. User Modeling and User-Adapted Interaction* **4 (2)**, 107–130.

Otter, M. & Johnson, H. (2000), 'Lost in hyperspace: Metrics and mental models', *Interacting with Computers* **13 (1)**, 1–40.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1999), The pagerank citation ranking: Bringing order to the web, Technical report, Stanford Digital Library Project, SIDL-WP-1999-0120.

Paramythis, A., Totter, A. & Stephanidis, C. (2001), A modular approach to the evaluation of adaptive user interfaces, *in* 'Proc. Workshop Empirical Evaluation of Adaptive Systems at UM2001'.

Park, J. & Kim, J. (2000), 'Contextual navigation aids for two world wide web systems', *Intl. J. Human-Computer Interaction* **12 (2)**, 193–217.

Pazzani, M., J., M. & Billsus, D. (1996), Syskill & webert: Identifying interesting web sites, *in* 'Proc. National Conference on Artificial Intelligence'.

Perkowitz, M. & Etzioni, O. (1997), Adaptive web sites: an ai challenge, *in* 'Proc. IJCAI 97', pp. 16–21.

Perkowitz, M. & Etzioni, O. (2000), 'Towards adaptive web sites: Conceptual framework and case study', *Artificial Intelligence* **118**, 245–275.

Pettijohn, T. (2002), 'Locus of control scale.', Web Survey.
URL: *http://www.ballarat.edu.au/ard/bssh/psych/rot.htm*

Pierrakos, D., Paliouras, G., Papatheodorou, C. & Spyropoulos, C. (2003), 'Web usage mining as a tool for personalization: A survey', *User Modeling and User-Adapted Interaction* **13**, 311–372.

Pilgrim, C. & Leung, Y. (1999), Designing www site map systems, *in* 'Proc. 10th Intl. Workshop on Database & Expert Systems Applications'.

Pirolli, P. (2004), 'The use of proximal information scent to forage for distal content on the world wide web', *Working with Technology in Mind: Brunswikian Resources for Cognitive Science and Engineering* .

Pirolli, P. & Card, S. (1999), 'Information foraging', *Psychological Review* **105 (1)**, 58–82.

Pirolli, P. & Fu, W. (2003), Snif-act: A model of information foraging on the world-wide web, *in* 'Proc. User Modeling 2003'.

Pirolli, P., Pitkow, J. & Rao, R. (1996), Silk from a sow's ear: Extracting usable structures from the web, *in* 'Proc. CHI 1996', pp. 118–125.

Pohl, W. (1999), 'Logic-based representation and reasoning for user modeling shell systems', *User Modeling and User-Adapted Interaction* **9**, 217–282.

Rashid, A., Albert, I., Cosley, D., Lam, S., McNee, S., Konstan, A. & Riedl, J. (2002), Getting to know you: Learning new user preferences in recommender systems, *in* 'Proc. Intelligent User Interfaces 2002'.

Rauterberg, M. (1992), A method of quantitative measurement of cognitive complexity, *in* G. van der Veer, M. Tauber, S. Bagnara & A. Antalovits, eds, 'Human-Computer Interaction: Tasks and Organisation', pp. 295–307.

Rich, E. (1979), 'User modeling via stereotypes', *Intl. J. Cognitive Science* **3**, 329–354.

Roy, M. & Chi, M. (2003), 'Gender differences in patterns of searching the web', *J. Educational Computing Research* **29**, 335–348.

Rozanski, H., Bollman, G. & Lipman, M. (2001), 'Seize the occasion: Usage-based segmentation for internet marketers', *eInsights (Booz-Allen & Hamilton)* .

Rumbaugh, J., Blaha, M., Eddy, F. & Lorensen, W. (1991), *Object-Oriented Modeling and Design*, Prentice Hall International Editions.

Russell, S. & Norvig, P. (1995), *Artificial Intelligence: a Modern Approach*, Prentice-Hall Inc.

Sellen, A., Murphy, R. & Shaw, K. (2002), How knowledge workers use the web, *in* 'Proc. CHI 2002', pp. 227–234.

Shahabi, C., Zarkesh, A., Adibi, J. & Shah, V. (1997), Knowledge discovery from users' web page navigation, *in* 'Proc. 7th Intl. Workshop on research Issues in Data Engineering on High Performance Database Management for Large-Scale Applications'.

Shneiderman, B. (1997), 'Designing information-abundant web sites: Issues and recommendations', *Intl. J. Human-Computer Studies* **47 (1)**, 5–29.

Sjolinder, M. (1998), *Towards a Framework for Design and Evaluation of Navigation in Electronic Spaces*, Swedish Institute of Computer Science, chapter Individual differences in spatial cognition and hypermedia navigation.

Smith, P. (1996), 'Towards a practical measure on hypertext usability', *Interacting with Computers* **8 (4)**, 365–381.

Smyth, B. & Cotter, P. (2002), The plight of the navigator: Solving the navigation problem for wireless portals, *in* 'Proc. Adaptive Hypermedia 2002'.

Spiliopoulou, M., Faulstich, L. & Winkler, K. (1999), A data miner analyzing the navigational behaviour of web users, *in* 'Proc. Workshop on Machine Learning in User Modelling at ACAI99'.

Spiliopoulou, M., Mobasher, B., Berendt, B. & Nakagawa, M. (2003), 'A framework for the evaluation of session reconstruction heuristics in web-usage analysis', *INFORMS J. Computing* **15 (2)**, 171–190.

Spink, A., Wolfram, D., Jansen, B. & Saracevic, T. (2001), 'Searching the web: the public and their queries', *J. American Society for Information Science* **53 (3)**, 226–234.

Srivastava, J., Cooley, R., Deshpande, M. & Tan, P. (2000), 'Web usage mining: Discovery and applications of usage patterns from web data', *SigKDD Explorations* **1 (2)**, 1–12.

Tauscher, L. & Greenberg, S. (1997), 'How people revisit web pages: Empirical findings and implications for the design of history systems', *Intl. J. Human-Computer Studies* **47**, 97–137.

Teevan, J. (2004), How people re-find information when the web changes, Technical report, MIT.

Teevan, J., Alvarado, C., Ackerman, M. & Karger, D. (2004), The perfect search engine is not enough: A study of orienteering behavior in directed search, *in* 'Proc. Chi 2004', pp. 415–422.

Teevan, J., Dumais, S. & Horvitz, E. (2005), Personalizing search via automated analysis of interests and activities, *in* 'Proc. SIGIR '05'.

Thüring, M., Hanneman, J. & Haake, J. (1995), 'Hypermedia and cognition: Designing for comprehension', *Comm. of the ACM 38 (8)* **38 (8)**, 57–66.

Ullrich, C. (2005), Course generation based on htn planning, *in* 'Proc. ABIS 2005'.

Van Ossenbruggen, J., Hardman, L. & Rutledge, L. (2002), 'Hypermedia and the semantic web: A research agenda', *Journal of Digital Information* **3 (1)**, No. 55.

Vitali, F. & Bieber, M. (1999), 'Hypermedia on the web: What will it take?', *ACM Computing Surveys* **31 (4es)**, Article No. 31.

Waern, A. (2002), 'User involvement in automatic filtering: An experimental study', *User Modeling and User-Adapted Interaction* **14**, 201–237.

Waterson, S. & Hong, J. (2002), What did they do? understanding clickstreams with the webquilt visualization system, *in* 'Proc. Advanced Visual Interfaces'.

Webb, G., Pazzani, M. & Billsus, D. (2001), 'Machine learning for user modeling', *J. User Modeling and User-Adapted Interaction* **11 (1-2)**, 19–29.

Weibelzahl, Lippitsch, S. & Weber, G. (2002), 'Advantages, opportunities, and limits of empirical evaluations: Evaluating adaptive systems', *Künstliche Intelligenz* **3 (02)**, 17–20.

Weibelzahl, S. (2003), Evaluation of Adaptive Systems, PhD thesis, Pedagogical University Freiburg.

Weibelzahl, S. (2005), *Adaptable and Adaptive Hypermedia Systems*, IRM Press, chapter Problems and Pitfalls in the Evaluation of Adaptive Systems, pp. 285–299.

Weibelzahl, S., Paramythis, A. & Totter, A. (2003), A layered framework for the evaluation of interactive adaptive systems, *in* 'Proc. 2nd Workshop on Empirical Evaluation of Adaptive Systems (at UM2003)', p. supplement.

Weinreich, H., Buchman, V. & Lamersdorf, W. (2003), Scone: Ein framework zur evaluativen realisierung von erweiterungen des webs, *in* 'Tagungsband Kommunikation in Verteilten Systemen - KiVS 2003', pp. 31–42.
URL: *www.scone.de*

Weinreich, H. & Lamersdorf, W. (2000), Concepts for improved visualization of web link attributes, *in* 'Proc. 9th WWW Conference', pp. 403–416.

Weinreich, H., Obendorf, H., Herder, E. & Mayer, M. (2006, forthcoming), Off the beaten tracks: Exploring three aspects of web navigation, *in* 'Proc. WWW 2006'.

Weinreich, H., Obendorf, H. & Lamersdorf, W. (2001), The look of the link: Concepts for the user interface of extended hyperlinks, *in* 'Proc. 12th ACM Conf. Hypertext and Hypermedia', pp. 19–28.

Wexelblat, A. (1999), History-based tools for navigation, *in* 'Proc. 32nd Int. Conf. on System Sciences'.

World Wide Web Consortium (2005*a*), 'Cascading style sheets home page', Internet.
URL: *http://www.w3.org/Style/CSS/*

World Wide Web Consortium (2005*b*), 'Mobile web initiative event', Online.
URL: *http://www.w3.org/2005/11/mwi-ukevent.html*

World Wide Web Consortium (2006), 'Mobile web best practices 1.0', Online Draft.
URL: *http://www.w3.org/TR/2006/WD-mobile-bp-20060113/*

Zukerman, I. & Albrecht, D. (2001), 'Predictive statistical models for usermodeling', *User Modeling and User-Adapted Interaction* **11 (1-2)**, 5–18.

Abstract

In a period of less than two decades, the World Wide Web has evolved into one of the most important sources of information and services. Due to the infancy of the Web and its rapid growth, our knowledge on how users interact with the Web is limited - knowledge which is likely to provide pointers for improvements in the design of Web sites and Web browsers. In this thesis, we aim to provide an integrative overview of theoretical insights and empirical findings, and to extend this body of knowledge with results from a number of user studies.

Research in the fields of user modeling and adaptive hypermedia, and the sub domain of Web personalization, is specifically aimed at understanding and supporting individual users' interests, needs, and interaction patterns. From the literature overview it becomes apparent that these techniques are useful in limited contexts, yet it is not evident whether general Web interaction carries sufficient evidence in order for these techniques to be useful. A large amount of Web activities has been shown to be associated with *recurrent tasks*; unfortunately, most data is limited to statistics that do not explain why and how users revisit pages. A further significant Web activity is *information gathering*. Theoretical models characterize this process of users following their *information scent*; empirical studies indicate that information gathering often involves *orienteering behavior* - an interleaving process of issuing queries and gathering context by navigating. A common usability problem in both activities is termed *disorientation*, which refers to situations in which users fail to understand their current position in a Web site, the way that led to this position, and where they can go to. A number of studies were carried out to shed a new light on the above matters. An analysis framework was developed for this purpose, based on first-order *navigation measures* and second-order *navigation styles*, *visualization* of user navigation paths, and the *cleaning* of client-side usage logs.

In our laboratory studies we targeted the issue of disorientation. Conflicting results exist on what interactions patterns indicate that users may experience such usability issues. In the first study we identified two aggregated measures

- navigation styles - that were closely associated with perceived disorientation. *Flimsy navigation* is a weak navigation style, mainly exhibited by inexperienced users who appear not to be able to reconstruct their navigation paths, and therefore are prone to get stuck. Task-related structural navigation support for both forward navigation and backtracking is expected to help users who navigate in a flimsy manner. By contrast, *laborious navigation* is an exploratory navigation style that emerges at the beginning of a task and serve to quickly build a mental model of the site structure. We believe that theoretical models of Web navigation should consider laborious *orienteering* as enrichment activities rather than failure, as it may result in increased performance at a later stage. The results from the first study were confirmed by the second laboratory study, in which we investigated the impact of task-based *link suggestions* on user navigation styles. From the results it became clear that users who were provided with this type of adaptive navigation support, navigated in a more structured manner, with more targeted backtracking and less returns to the sites' home pages. The link suggestions were positively received, improved user perceptions, and decreased task execution time.

From earlier user studies it is known that users often return to Web pages that they have visited before. In a long-term study, we aimed to distinguish the various kinds of revisits, and to find out to what extent browser support for these revisits may be improved. For this purpose, we tracked and analyzed the Web activities of 25 participants for an average period of three months. We distinguished cross-session revisits (*recurrent behavior*) from within-sesion revisits (*backtracking*) - the latter activity covered 74% of all revisits. Whereas backtracking is surprisingly well supported by a list of the 15 most recent pages, a list to cover a reasonable percentage of recurrent activities would need to be overly long. In particular infrequent and irregular page revisits are problematic, as the user is less likely to remember the exact addresses; ironically, it is also less likely that the address still resides in the browser history.

In contrast to the earlier studies, our participants often made use of multiple browser windows or tabs. This interaction style reduces the need for backtracking, and allows for side-by-side comparison of results. A disturbing consequence is that the concept of the back button is disrupted; as the history is split over several stacks, users need to keep track of what activities were carried out in which window; this places a significant cognitive load on the user. A further change is the highly increased percentage of form submissions, which indicates that the Web has become a hybrid between 'traditional' hypermedia and a dynamic, interactive system. This requires a radical rethinking of the browser interface.

From the results several directions for navigation support can be extracted. Based on the orienteering behavior that we observed in our studies, it is likely that navigation will remain an important paradigm in Web-based interaction; the combined process of search and exploration of information patches allows users to find what they need rather than a best match to some query. Currently, search and navigation are largely separated activities, and users need to actively keep track of their query refinement process. A visual query modification history, link relevancy indicators, and search result ranking that takes surrounding pages into account, are expected to be simple yet effective means to support this process.

As users often make use of multiple windows, users need to keep track of what activities where carried out in which window; branched history mechanisms that takes both the temporal order of user navigation, and the parallel character of multiple windows into account, are mechanisms that definitely need to developed. In addition, results from our laboratory studies indicate that annotation of recently followed paths and links that lead to recently visited pages may greatly reduce user disorientation.

Users often carry out recurrent activities on the Web. Whereas the larger part of these activities involve a small number of very popular pages, the greatest need of history support for these activities involves a large number of infrequently visited pages; the number is too large to be covered by a reasonably small history list or bookmark file. As an alternative, we propose explicit history search mechanisms, combined with visually annotated trails from earlier search results - which are often waypoints rather than the desired item to be revisited.

Perhaps the most important challenge for current Web browsers is the shift of the Web paradigm from a hypermedia system to a hybrid of hypermedia and interactive applications. Document management options are dearly needed to deal with volatile yet relevant dynamically generated pages, such as travel plans and order confirmations. Ideally, users should only need to access one point both for revisiting earlier travel plans, and for creating new ones. This will require a major rethinking of the metaphor of Web navigation into an interleaving process of navigation and interaction.

Several theoretical models and empirical studies have provided us with various perspectives on how users interact with the Web. It has become clear that there is no such thing as the ‘typical’ Web user or a ‘typical’ Web navigation session. It is most unlikely that the various aspects can be covered by one integrated model. Therefore, we think it is more desirable to strive toward a comprehensive body of empirical data and theoretical insights, with explicit indications in which situations they are applicable.

Personalization of the Web interface to the individual user has proven to be effective in limited environments. Given the sheer variety of tasks for which the Web is used, sophisticated user-adaptive mechanisms for general Web navigation are likely to be infeasible. Instead, we think a current challenge for the adaptive hypermedia community is to develop and evaluate straightforward support mechanisms, which may well turn out to be one-size-fits-all designs for diversity.

Samenvatting

In een kleine twintig jaar is het World Wide Web gegroeid tot een schier onmisbare bron van informatie en diensten. Omdat het Web nog een relatief jong medium is, en vanwege zijn snelle groei, is weinig bekend hoe het Web precies gebruikt wordt - kennis die belangrijk is om de interface van websites en webbrowsers te verbeteren. Het doel van dit proefschrift is om de huidige theoretische inzichten en empirische gegevens bij elkaar te brengen, en deze inzichten uit te breiden met een aantal gebruikersstudies.

Onderzoek op het gebied van gebruikersmodellering en adaptieve hypermedia, en het meer beperkte gebied van webpersonalisatie, is er op gericht de interesses, behoeften en interactiepatronen van individuele gebruikers te begrijpen en te ondersteunen. Uit het literatuuroverzicht komt naar voren dat deze technieken weliswaar hun nut hebben bewezen in beperkte domeinen, maar dat het nog zeker geen uitgemaakte zaak is of ze ook werken voor webnavigatie in het algemeen. Het is bekend dat veel activiteiten op het Web is gerelateerd aan *terugkerende taken*; helaas zijn er geen statistieken bekend waarom en op welke manier gebruikers terugkeren naar webpagina's. Tevens wordt het Web veelvuldig gebruikt voor het *verzamelen van informatie*. Theoretische modellen beschrijven dit proces als 'het volgen van je neus'. Gebruikersstudies laten zien dat het verzamelen van informatie veel meer inhoudt dan slechts het gebruik van zoekmachines; het *oriënterend* navigeren door sites levert informatie op om de zoekopdrachten te verfijnen. Een algemeen probleem hierbij wordt wel *desoriëntatie* genoemd: situaties waarin gebruikers niet meer weten waar ze precies zijn, hoe ze daar zijn gekomen, en welke richtingen ze uit kunnen. Een aantal gebruikersstudies is uitgevoerd om een nieuw licht op deze zaken te laten schijnen. Hiervoor is een raamwerk ontwikkeld dat gebruik maakt van eerste-order *navigatiemetrieken* en combinaties van deze metrieken, die we *navigatiestijlen* noemen, alsmede van *visualisatietechnieken* en methoden om *artefacten* te verwijderen uit logbestanden.

Middels twee gecontroleerde studies hebben we het fenomeen desoriëntatie onderzocht. Eerdere studies naar interactiepatronen die er op wijzen dat gebruikers met dit probleem kampen, laten tegenstrijdige resulaten zien. In de eerste studie hebben we twee navigatiestijlen gevonden die sterk met disoriëntatie te maken hebben. *Flimsy* (ongestructureerde) navigatie is een stijl die voornamelijk wordt waargenomen bij onervaren gebruikers, die moeite hebben met het onthouden van hun eerdere acties en daardoor sneller vastlopen. Taakgerelateerde navigatiehulpmiddelen die structuur bieden voor zowel voorwaarts navigeren als backtracken kunnen in dit geval soelaas bieden. Daarentegen is *laborious* (intensieve) navigatie een exploratieve navigatiestijl, die typisch aan het begin van een taak wordt gebruikt om snel een mentaal model van de sitestructuur te vormen. Naar onze mening zullen theoretische modellen van webnavigatie dit intensief *oriënteren* moeten interpreteren als een investering die later profijt kan opleveren, en niet als weinig zinvolle activiteiten. De resultaten van de eerste studie worden bevestigd door onze tweede studie, waarin we de invloed van taakgerelateerde *linksuggesties* op navigatiestijlen hebben onderzocht. Uit deze studie kwam naar voren dat deze vorm van gepersonaliseerde ondersteuning er toe leidde dat gebruikers op een meer gestructureerde wijze navigeerden, met een meer gefocuste manier van backtracken. De linksuggesties werden door de gebruikers gewaardeerd, verhoogden de tevredenheid met de interface, en leidden tot een vermindering van de tijd die men aan de verschillende taken besteedde.

Uit eerdere gebruikersstudies is bekend dat gebruikers regelmatig terugkeren naar webpagina's die men al eerder heeft bezocht. In een langetermijnstudie hebben we deze terugkerende paginabezoeken nader bekeken, en onderzocht in hoeverre webbrowsers hierbij ondersteuning bieden. Hiertoe hebben we de activiteiten van 25 gebruikers gevolgd, gedurende een periode van gemiddeld drie maanden. Daadwerkelijk *terugkerend gedrag* hebben we onderscheiden van herhaalde paginabezoeken als gevolg van *backtracken* - wat 74% van alle herhaalde bezoeken besloeg. Hoewel backtracken verrassend goed wordt ondersteund door een lijst van de 15 meest recent bezochte pagina's, zou men een lange lijst nodig hebben om een redelijk percentage van alle terugkerende activteiten te ondersteunen. Met name pagina's die slechts af en toe, of op onregelmatige basis, worden bezocht zijn problematisch, omdat de kans groter is dat de gebruiker het precieze webadres niet meer kan herinneren; ironisch genoeg is juist in deze gevallen het adres vaak ook niet meer aanwezig in het geheugen van de webbrowser.

In tegenstelling tot eerdere studies, observeerden we dat onze deelnemers veelvuldig gebruik maakten van meerdere vensters of tabbladen. Dit interactiepatroon vermindert de noodzaak om te backtracken, en maakt het mogelijk verschillende resultaten naast elkaar te vergelijken. Een groot nadeel is echter dat het concept van de back button wordt ondermijnd; omdat de navigatiegeschiedenis over meerdere vensters is verdeeld, moeten gebruikers onthouden welke activiteiten in welk venster hebben uitgevoerd. Een verdere verandering is een sterk verhoogd gebruik van formulieren; dit duidt er op dat het Web een hybride vorm van 'traditionele' hypermedia en dynamische, interactieve applicaties is geworden. De huidige browserinterface zal sterk moeten veranderen om hierop in te spelen.

Uit bovenstaande resultaten kan een aantal implicaties voor de interface worden afgeleid. Gezien het oriënterende gedrag dat we in onze studies hebben waargenomen, is het waarschijnlijk dat dit een belangrijk paradigma in webnavigatie blijft; de combinatie van het gebruik van zoekmachines en navigerend verkennen van de resultaten, verhoogt de kans dat men vindt wat men zoekt, in plaats van dat waar men naar vraagt. Op dit moment zijn zoeken en navigeren vrij onafhankelijke activiteiten, en gebruikers moeten zelf bijhouden hoe ze hun zoekopdrachten hebben veranderd. Een overzicht van recente zoekopdrachten, indicatie van de relevantie van links, en het aangeven van de relevantie van omliggende pagina's, zijn eenvoudige doch effectieve methoden om dit proces te ondersteunen.

Bij het gebruik van meerdere vensters, moeten gebruikers bijhouden welke activiteiten in welk venster hebben plaatsgevonden; daarom is het zeer wenselijk dat vertakte geschiedenismechanismen worden ontwikkeld, die zowel het temporele karakter van webnavigatie als ook het parallelle karakter van de meerdere vensters recht aan doen. Daarnaast suggereren onze gecontroleerde studies dat annotatie van eerder gevolgde paden en links die leiden tot eerder bezochte pagina's helpen bij het oriëntateren.

Wellicht de grootste uitdaging voor het ontwerp van webbrowsers is de ontwikkeling van het Web tot een hybride vorm van hypermedia en interactieve toepassingen. Browsers zullen functionaliteit voor documentbeheer moeten bieden voor dynamisch gerelateerde pagina's, die men na het sluiten van de browser niet meer terug kan vinden. Deze pagina's kunnen echter belangrijke informatie bevatten, zoals routebeschrijvingen en orderbevestigingen. Idealiter zouden gebruikers vanaf één centraal punt zowel oude routebeschrijvingen kunnen opvragen als nieuwe routes berekenen. De metafoor van webnavigatie zal moeten worden uitgebreid tot een afwisselend proces van navigatie en interactie.

Uit de verschillende perspectieven op webnavigatie komt naar voren dat er niet zoiets is als de 'typische gebruiker' of 'typisch webgebruik'. Het is onwaarschijnlijk dat de variëteit aan patronen door één geïntegreerd model kan worden verklaard. Daarom lijkt het ons raadzamer om te streven naar een corpus van empirische gegevens en theoretische inzichten, waarin expliciet wordt aangegeven in welke situaties ze gebruikt kunnen worden.

Personalisatie van de webinterface heeft zijn nut bewezen in beperkte domeinen. De diversiteit aan activiteiten die op het Web worden uitgevoerd, maakt het onwaarschijnlijk dat geavanceerde adaptieve technieken nuttig kunnen worden toegepast voor webnavigatie in het algemeen. Op dit moment ligt de grootste uitdaging voor de adaptieve-hypermediagemeenschap in het ontwikkelen en evalueren van eenvoudige concepten voor navigatiesupport; het is niet onmogelijk dat deze concepten algemeen toepasbare ontwerpen voor diversiteit zullen zijn.

SIKS Dissertation Series

Since 1998, all dissertations written by Ph.D. students who have conducted their research under the auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series. Below is an overview of the most recent publications.

2004

1. Virginia Dignum (UU)
   A Model for Organizational Interaction: Based on Agents, Founded in Logic
2. Lai Xu (UvT)
   Monitoring Multi-party Contracts for E-business
3. Perry Groot (VU)
   A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving
4. Chris van Aart (UVA)
   Organizational Principles for Multi-Agent Architectures
5. Viara Popova (EUR)
   Knowledge discovery and monotonicity
6. Bart-Jan Hommes (TUD)
   The Evaluation of Business Process Modeling Techniques
7. Elise Boltjes (UM)
   Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes
8. Joop Verbeek(UM)
   Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politile gegevensuitwisseling en digitale expertise
9. Martin Caminada (VU)
   For the Sake of the Argument; explorations into argument-based reasoning

10. Suzanne Kabel (UVA)
    Knowledge-rich indexing of learning-objects
11. Michel Klein (VU)
    Change Management for Distributed Ontologies
12. The Duy Bui (UT)
    Creating emotions and facial expressions for embodied agents
13. Wojciech Jamroga (UT)
    Using Multiple Models of Reality: On Agents who Know how to Play
14. Paul Harrenstein (UU)
    Logic in Conflict. Logical Explorations in Strategic Equilibrium
15. Arno Knobbe (UU)
    Multi-Relational Data Mining
16. Federico Divina (VU)
    Hybrid Genetic Relational Search for Inductive Learning
17. Mark Winands (UM)
    Informed Search in Complex Games
18. Vania Bessa Machado (UvA)
    Supporting the Construction of Qualitative Knowledge Models
19. Thijs Westerveld (UT)
    Using generative probabilistic models for multimedia retrieval
20. Madelon Evers (Nyenrode)
    Learning from Design: facilitating multidisciplinary design teams

2005

1. Floor Verdenius (UVA)
   Methodological Aspects of Designing Induction-Based Applications
2. Erik van der Werf (UM))
   AI techniques for the game of Go
3. Franc Grootjen (RUN)
   A Pragmatic Approach to the Conceptualisation of Language
4. Nirvana Meratnia (UT)
   Towards Database Support for Moving Object data
5. Gabriel Infante-Lopez (UVA)
   Two-Level Probabilistic Grammars for Natural Language Parsing
6. Pieter Spronck (UM)
   Adaptive Game AI
7. Flavius Frasincar (TUE)
   Hypermedia Presentation Generation for Semantic Web Information Systems

8. Richard Vdovjak (TUE)
   A Model-driven Approach for Building Distributed Ontology-based Web Applications
9. Jeen Broekstra (VU)
   Storage, Querying and Inferencing for Semantic Web Languages
10. Anders Bouwer (UVA)
    Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments
11. Elth Ogston (VU)
    Agent Based Matchmaking and Clustering - A Decentralized Approach to Search
12. Csaba Boer (EUR)
    Distributed Simulation in Industry
13. Fred Hamburg (UL)
    Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen
14. Borys Omelayenko (VU)
    Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics
15. Tibor Bosse (VU)
    Analysis of the Dynamics of Cognitive Processes
16. Joris Graaumans (UU)
    Usability of XML Query Languages
17. Boris Shishkov (TUD)
    Software Specification Based on Re-usable Business Components
18. Danielle Sent (UU)
    Test-selection strategies for probabilistic networks
19. Michel van Dartel (UM)
    Situated Representation
20. Cristina Coteanu (UL)
    Cyber Consumer Law, State of the Art and Perspectives
21. Wijnand Derks (UT)
    Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics

2006

1. Samuil Angelov (TUE)
   Foundations of B2B Electronic Contracting
2. Cristina Chisalita (VU)
   Contextual issues in the design and use of information technology in organizations
3. Noor Christoph (UVA)
   The role of metacognitive skills in learning to solve problems

4. Marta Sabou (VU)
   Building Web Service Ontologies
5. Cees Pierik (UU)
   Validation Techniques for Object-Oriented Proof Outlines
6. Ziv Baida (VU)
   Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling
7. Marko Smiljanic (UT)
   XML schema matching - balancing efficiency and effectiveness by means of clustering
8. Eelco Herder (UT)
   Forward, Back and Home Again - Analyzing User Behavior on the Web
9. Mohamed Wahdan (UM)
   Automatic Formulation of the Auditor's Opinion