Development of a Cognitive Model for Navigating on the Web

Ion Juvina

Ion Juvina Development of a Cognitive Model for Navigating on the Web

Development of a Cognitive Model for Navigating on the Web

Ontwikkeling van een Cognitief Model voor Navigatie op het Web
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. W.H. Gispen, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op donderdag 19 oktober 2006 des ochtends te 10.30 uur

door

Ion Juvina

geboren op 16 oktober 1967
te Cetatea, Roemenië

Promotor: Prof. Dr. J. van den Berg

Co-promotor: Dr. H. van Oostendorp

Reading committee:

Prof. Dr. Lynda Hardman
Prof. Dr. Mark Neerincx
Prof. Dr. Jean-Francois Rouet
Prof. Dr. Arno Siebes
Prof. Dr. Gerrit van der Veer

SIKS Dissertation Series No. 2006-23
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

Juvina, Ion
Development of a Cognitive Model for Navigating on the Web
Doctoral Dissertation, Utrecht University, 2006
Includes bibliographical references

Printed by: ZuidamUithof drukkerijen, Utrecht, The Netherlands
ISBN-10: 90-393-4382-9
ISBN-13: 978-90-393-4382-1

This book is dedicated
to my grandparents

Contents

Preface

Dear reader, here you are! There comes a time when a scientist has to deliver something. For the last four years, I have enjoyed doing the most exciting things I could think of: reading, doing experiments, teaching, writing, and meeting peers. It is your turn now to benefit and hopefully enjoy the product of my work.

Oh, have I just said "my work"? Please take it as in "I am the one lucky enough to report this teamwork". First of all, there is a great deal of input here from my daily supervisor Dr. Herre van Oostendorp. Not only has he masterminded the whole project from the very beginning, he has been there all the time. If one wonders what the expression "daily supervisor" really means, one should meet Herre. At times, I was quite challenged by his insistence on details, but this helped me sharpen my mind and my writing hand. My promotor, Prof. Dr. Jörgen van den Berg strategically guided me and, most importantly, gave me confidence. I have always known that he would have supported me if I had been in need of anything.

I had profitable exchanges of ideas with my colleagues in our group – Cognitie en Communicatie. In particular, I worked pretty close with Stacey Nagata, Joris Graaumans, and Christof van Nimwegen. Thank you all for your collaboration, and for making me feel a bit at home between yourselves!

I have been involved in teaching and supervising projects from the beginning to the end of my stay at Utrecht University. I owe thanks to Herre for assigning me not only assistant duties but also high-level responsibilities such as course development, coordination of research internships and assessment. Needless to say, I learned a great deal from teaching and interacting with other teachers. In particular, I advanced my knowledge of statistics by assisting Dr. Richard Starmans in his course "Advanced Research Methods". Thank you Richard, for making it smooth and, to use one of your favorite words, robust!

A number of master's students contributed directly to my research, most of the time with data collection, but also with brilliant ideas, or just with tough questions. Here they are: Ellert van den Broek, Vincent van der Linden, Koen Buurman, Poyan Karbor, Brian Pauw, and Arnaud Lek. Thank you, guys! I hope you benefited from our collaboration at least as much as I did.

I have collaborated with a PhD student from Twente University, now Dr. Eelco Herder. Together we did a couple of experiments, data analysis, and writing. It has been beneficial and most of the time enjoyable to work with a brilliant computer scientist. Thank you, Eelco, for everything, and especially for presenting our award-winning paper!

Working and living in Utrecht, The Netherlands, made it possible for me to successfully complete this project. Utrecht is a perfect city for studying and doing research and Utrecht University provides everything a scientist would need to do top-level research. I wish I could mention all the nice people from the management and administration of the Institute of Information and Computing Sciences, but there are just too many to be listed here. I do thank them though from all my heart! I owe special thanks to Ir. Wilke Schram, Monique Dixon and Floor Jansen for welcoming me in the institute and helping me with personal and professional matters.

Thanks to Wim de Jonge for being a friend and sports-partner all these years. Florentina Pena, my housemate, made me realize that sharing facilities can be not only practical but also enjoyable. Thank you Florentina for letting me drink the leftover coffee! Thanks to Dr. Hein van Vliet, my landlord, for being a gentleman and for keeping things under control.

It is such an honor to have the manuscript of this thesis proofread by Abigail G. Matthews, a Harvard PhD and Yale post-doc. Thank you Abigail, for tolerating my peculiar writing style!

Last but not least, I would like to thank Rosa de Vries, who has been my friend and girlfriend for most of the period of my studies in Utrecht. Thank you Rosa, for understanding my crush on science, downsized only by my crush on you!

Chapter 1. Introduction

The focus of the PhD project reported here is Web navigation. The idea that inspired this project comes from a study conducted by Van Oostendorp (2002). He started from a practical desideratum: given a website for the general public, how could one make sure it is visited by its intended users and visits are followed by re-visits? In other words, how could one make sure the users' experience with a particular website is satisfactory and therefore the website is perceived as worth re-visiting by its intended users? As the main result of this study, *Interestingness of provided information* and *Ease of navigation* were found to be the main factors determining users' satisfaction. These two factors, as well as users' satisfaction were measured with the aid of questionnaires. Thus, how interesting the content is perceived to be and how easy to use the structure of the website seems to the users determine whether and to what extent users are satisfied with using the website. It is interesting to note as early as now that it was in this study for the first time when *structure-related aspects* (ease of navigation) were found to be complementary to *content-related aspects* (interestingness of provided information). Later on this dichotomy will reoccur and be referred to as *syntax vs. semantics*.

Throughout this report the term "Web navigation" is used as a central metaphor for using the Web (Instone, 2002). While we agree with those who argue that using the Web is much more than navigation, we adhere to the main stream of researchers who see the Web as a world-wide hypertext and navigation as the major part of user experience on the Web (Lazar, 2003). We use the term "Web navigation" in a broad sense referring to users' orientation in an information space, locating information and progressing from one information source to another. Other terms frequently used in this field, such as *surfing*, *foraging*, *browsing*, or *searching,* are considered synonyms with or enclosed in *Web navigation*, unless otherwise specified. Searching in strict sense – as in using a search engine – although an important part of Web use, is not our focus here, and is extensively treated elsewhere (Van Zwol & Van Oostendorp, 2004).

An interesting common connotation of the above metaphors is the suggested challenge involved in using the Web. The Web has brought us not only the opportunity of nonlinear access to information sources but also the challenges of *cognitive overload* and *disorientation* (Conklin, 1987; Edwards & Hardman, 1988).

Since this field of study is relatively new, researchers have to rely on more established domains to provide concepts and methods. In our case, such domains are: human-computer interaction, cognitive science, human factors, and text comprehension. Since the activity of Web navigation is in itself new, one has to make analogies with activities that have been extensively studied and understood. Throughout this report we will mainly use interaction paradigms such as reading/writing and talking/listening. Within these paradigms, using the Web can be seen as a dialogue: users inform Web applications about their choices and Web applications "reply" with content. There are also specifics of Web navigation that become apparent during these analogies. For example, involving spatial features (syntax) in processing contents (semantics) is recognized as a distinguishing characteristic of Web navigation (Di Blas, Paolini & Speroni, 2004).

Research attempting at modeling cognitive mechanisms involved in Web navigation gains increasing influence in the Human-Computer Interaction community (Kitajima, Blackmon & Polson, 2000; Pirolli & Fu, 2003). Existing theories such as Information Foraging (Pirolli & Card, 1999) and Construction-Integration (Kintsch, 1998) and models such as SNIF-ACT (Pirolli & Fu, 2003) and CoLiDeS (Kitajima, Blackmon & Polson, 2000) have been taken as input for our own research. Based on this input, we have tried to make a step further and propose amendments to the existing models. As we will try to convince the reader in the next sections, the main idea of our model is that not only semantic but also syntactic (spatial) processes must be accounted for in models of Web navigation tasks (Juvina & van Oostendorp, 2004).

The objective of our research was to build a cognitive model that predicts and explains human performance in Web-assisted tasks. We intended to gather facts and descriptive statistics of Web navigation behavior in order to ground our model in reality. These data were expected to help us answer the question: *what are the most important factors determining success in Web-assisted tasks*? An important part of our data was purposely automatically recorded. The reason for this was two-fold: (1) automation allows efficiency in data gathering; and (2) when user data is automatically recorded, dynamic (real-time) adaptations of the Web application can be designed. However, automatically recorded data (referred also as Web-logging data) need to be properly interpreted; therefore, an important part of our work was concerned with answering the question: how can one make sense of behavioral navigation data?

The next intended step was to build a cognitive model that explains the role of the factors that appeared to be important in the previous phase. This model was expected to give insights into how the information space is represented in users' minds, what features are represented or abstracted out, how relevance of information perceived on the screen is judged, how this judgment of relevance can be computationally modeled, how contextual information is used in judging relevance, and how one can computationally model the involvement of contextual information.

This project has also a more practical objective: improving existing interfaces and tools and providing a better navigation support to users. The full realization of this practical component is, certainly, beyond the scope of a PhD project. However, practical criteria are useful for guiding and testing the theoretical and empirical work. Within this objective we try to answer questions such as: *What type of support is necessary and possible*? *What are the consequences and implications of providing theory-based Web navigation support*? Empirical studies showing the usefulness of such support will be presented, confirming the validity of the proposed model.

Applications of this research are suggested. In particular, using the Web via screen readers by visually impaired persons (VIPs) is treated as a demonstrative case. We have conducted an empirical study in which vision impairment was simulated, in order to investigate *how VIPs can be supported in their Web use.* Participants had to perform Web navigation tasks with the screen of the computer switched off and, instead, with the aid of a screen reader. We also suggest that our work can be applied to other cases where information overload interacts with users' cognitive limitations (mobility, multitasking, etc.) and to other target populations (e.g.,elderly, cognitively impaired).

Summarizing, the research presented here is driven by the following questions:

- What are the most important factors determining success in Web-assisted tasks?
  - How can one measure or estimate these factors in an automatic way?
- What are the explanatory cognitive mechanisms for the identified factors?
  - How can one implement these mechanisms in a (computational) cognitive model?
- What kind of Web navigation support can be conceived based on the knowledge gained from the two previous questions?
  - What impact has this support on users?

The remainder of this thesis is structured as follows:

Chapter 2 presents the field of studying Web navigation from behavioral and cognitive perspectives. It shows previous work and results on this topic and it introduces the necessary concepts for the next sections. It starts with presentation of general facts about Web navigation, continues with individual differences, goes into details about cognitive processes involved in Web navigation and cognitive models of Web navigation, and it ends with practical constraints and opportunities.

Chapter 3 discusses methodological issues. Since the field is relatively new, there are important problems that one is confronted with, such as: how to collect and analyze navigation data, what and to what extent results can be generalized, where to place ourselves between the paradigms of common user versus individual differences, and how descriptive concepts can be computationally modeled.

In Chapter 4, several empirical studies are presented together with their results. The structure of this chapter follows the same logic as the one in Chapter 2. It starts with facts about Web navigation, how they were measured, recorded and analyzed, and how they can be interpreted. It continues with individual differences and how these were investigated based on correlational analyses. Then it presents how we modeled the main cognitive processes involved in Web navigation and what is the value, both theoretically and practically, of our modeling approach. Subsequently, two experimental studies are presented aimed at testing our model and its relevance. The results of these studies suggest that it is empirically justified and practically attractive to provide model-based navigation support to users of Web applications.

Chapter 5 presents possible extensions and applicability of our research. The case of using the Web via screen readers by VIPs is used for demonstrative purposes. A computational implementation of the model in a cognitive architecture is also proposed and demonstrated.

Chapter 6 summarizes the main contributions of this project to the fields of information science, human-computer interaction and cognitive science. The value and limits of our research reported here are discussed, together with directions for further research.

Chapter 2. Web navigation – behavioral and cognitive perspectives

This chapter presents behavioral and cognitive perspectives in studying Web navigation. The following two paragraphs show that the Web navigation behavior is triggered by the specific features of Web-based applications. Section 2.1 describes the most commonly studied domains and tasks, the main paradigm (the Web as a hyperspace), and the main characteristics of the web navigation behavior. Section 2.2 reviews individual differences in Web navigation as presented in the literature. Section 2.3 presents cognitive perspectives (both theoretical and computational) in studying Web navigation. Section 2.4 reviews several practical constraints and opportunities that guided our research.

Navigation is a major part of user experience on the Web (Lazar, 2003). This particular type of behavior is triggered by a specific type of applications, which has become very common nowadays, namely *Web-based applications*. The user interface of these applications – Web interface (WI) as called in Ivory (2001) – has some characteristics that differentiates it from other types of interfaces, such as command language interface (CLI), graphic user interface (GUI), direct manipulation interface (DMI), WIMP interface (windows, icons, menus and pointing devices), virtual reality (VR), etc. Unlike in GUI, DMI and WIMP interfaces, where mainly the functionality of an application is explored, WIs prompt the user to explore the domain knowledge. In fact, Web users face two different interfaces:

- the browser interface, which remains consistent in daily use, and
- the site interface, which changes from site to site.

While the browser interface is rather easy to learn, it is impossible to provide adequate training on how to navigate through the many thousands of websites that the user may visit (Lazar, 2003).

2.1. Instances and description of Web navigation behavior

Web interfaces (WIs) and the facility of navigation through large information spaces brought new problems for application designers and usability specialists; *cognitive overload* and *disorientation* are the main ones (Conklin, 1987; Edwards & Hardman, 1988). Interfaces have been traditionally designed with the function of providing users with information and tools so that they can perform their tasks. In the case of WIs, the function of providing information has developed so much that it has almost become a burden for the user. Therefore, adequate tools to filter the information that is offered to the user and to guide navigation through the information space are necessary. The user must also be assisted in deciding what information is relevant, trustworthy, useful, etc. In order to achieve these functions, WIs must be aware of the user; in other words, they must incorporate a model of the user. Such a model should be based on observed facts about users and about the interaction between users and Web applications.

2.1.1. What is the Web used for? Domains and tasks

The Web was initially used by scientists and mainly within the domain of science. Nowadays it is becoming increasingly popular and is used in almost every domain. Morrison, Pirolli and Card (2001) conducted an extensive *Web and Internet Use* survey aiming at understanding what types of activities users perform on the Web. Participants were asked to answer the following question: "Please try to recall a recent instance in which you found important information on the World Wide Web, information that led to a significant action or decision. Please describe that incident in enough detail so that we can visualize the situation." The survey recorded 3292 participants and the authors selected 2188 usable responses. The following is an example of such responses:

> *I accessed Netscape's financial site to check my credit card balance and how long it would take to pay it off. I'm now MUCH more fiscally aware of my spending habits and am trying to pay off my balance more actively.*

Responses were classified based on: the *Purpose* of people's search on the Web, the *Method* people use to find information, and the *Content* of the information for which they are searching. This taxonomy has been used for understanding people's activity on the Web and for developing ecologically-valid tasks to be used when studying Web behavior. Other authors also collected naturalistic tasks and used them in their studies. For example, Choo, Detlor and Turnbull (2000) studied knowledge workers at their workplace.

These studies have shown that users themselves report a type of task that could be called a *performance-oriented task* (e.g., checking credit card balance, shopping for second-hand car, and finding treatment or medication). In this type of task, the user usually has a goal or a target possible to be specified (at least partially) in advance. Task execution is driven by this goal and clear criteria can be imagined to judge task execution and task success. These tasks are rather complex; they involve locating information, comprehension of, and selection from various options. There are certainly other types of activities that people do on the Web (for example, exploring the Web without something specific in mind), but this research project has focused on performance-oriented tasks.

In reading comprehension research (one of our sources of analogy, as mentioned in Chapter 1), there are reading tasks in which a specific objective is given in advance (e.g.,find discrepancies, or write an argument based on the processed text). These tasks are most similar with the type of Web navigation tasks that we have focused on in this project. These tasks are to be differentiated from other reading comprehension tasks in which there is no pre-specified goal and an understanding of the text passage is constructed during the process of reading.

2.1.2. The information space

Using the Web involves both processing contents and navigating through a structure of hyperlinks that connects the individual information units. Following links between multiple information sources adds a navigational load (Sanchez & Wiley, 2005) to the existing comprehension load. As a network of information units must be represented and manipulated, spatial processing and spatial abilities become critical.

Theodor Nelson, who coined the term "hypertext", defined it as the hyperspace of concepts from a given text (Rada, 1991). This hyperspace is an important component of the task environment when Web-assisted tasks are performed. The fact that users have to mentally represent and manipulate this hyperspace is a characteristic of Web navigation behavior that distinguishes it from other similar behaviors such as reading of plain (non-hyper) text. The term *information space*, as it is used throughout this report, refers to both contents and (hyper) structure. In Web contexts, information spaces have a "patchy" structure. A patch is a group of related information items relatively isolated from other similar groups. Patches can be static (e.g., a website), or created ad-hoc (e.g., the results page of a search engine query) (Pirolli & Card, 1999).

The structure and contents of information spaces are mentally represented and manipulated during Web navigation sessions. Although the existence and the exact nature of these representations are still under debate (Farris, Jones et al., 2002), we are inclined to favor the position of those who assume that such representations are essential in explaining users' Web navigation behavior and its underlying mental activity. This standpoint seems to be well founded in the domains of cognitive psychology and text comprehension (Kintsch, 1998). Subsequent chapters of this report will bring more arguments in favor of this position, including results from our own studies.

2.1.3. What do users do? Web navigation behavior

Web navigation allows the user to approach an information space in a rather natural manner, basically in the same way as orientation in physical space or seeking for food (Pirolli & Card, 1995). Traditionally, Web navigation behavior is depicted as following links on Web pages or moving backward and forward in a succession of Web pages using adequate buttons of the browser application. However, information spaces tend to become very large, complex and abstract, thus unnatural and unintuitive (see Figure 1 for a visualization of a rather small and simple information space). Relying only on our innate ability to orient in physical spaces might be misleading. It has been shown that complex information structures are difficult to handle (Larson & Czerwinski, 1998). Tools such as search engines have been designed to help users in locating specific information in large information spaces. However, such tools are not completely replacing the link following behavior (also called *browsing*). Search engines provide only shortcuts in a link following chain. Typically, users need to browse through the results page returned by a search engine. Thus, browsing and searching are complementing each other. The amounts of searching and browsing that compose a Web navigation session depend (among other things as user preference or skill) on the type of task to be performed. Tasks with a specific goal that is easy to express in keywords are better done with search engines whereas tasks with under-specified goals are better accomplished via browsing. Commonly, tasks require both behaviors (Olston & Chi, 2003). For instance, when shopping for a computer, one might first use a search query to identify the online store and the proper product category, then browsing to select from various options available. It would be virtually impossible to pre-specify in keywords the exact configuration of features that the buyer wants to be included in the final product. Selecting from the available options by browsing through a set of Web pages is the most effective and comfortable option. The research reported here has mainly addressed the browsing component of Web navigation behavior. In our studies, to keep the tasks ecologically valid, users were allowed to use search engines; however, our descriptive and modeling endeavor has focused on users' browsing behavior.

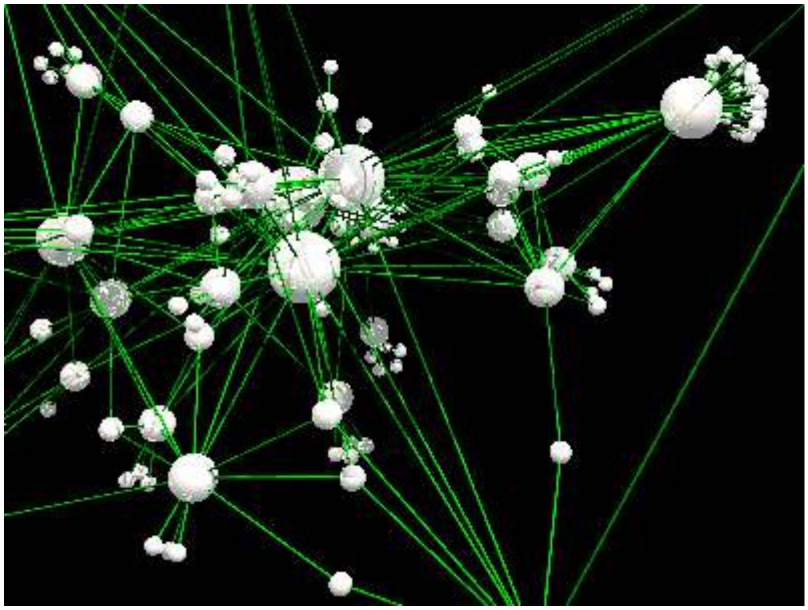


Figure 1. A possible visualization of a rather small and simple information space. Note complexity and abstract character. Imagine the difficulty in visualizing larger and more complex information spaces.

A characteristic of Web navigation behavior frequently reported is re-visitation. According to Cockburn and McKenzie (2000) about 80% of all page requests on the Internet involved pages that a user visited before. More recent studies have corrected down this estimation (Herder, 2005); however the fact remains that users re-visit a relatively large amount of pages. Why do users re-visit pages? An elaborate answer to this question will not be given in this section. However, literature suggests that re-visitation is a means to explore the information space in order to adequately judge the value of a particular information unit. Post-valued recall (PVR) refers to the interest a user may have in recalling information whose value is not recognized until some time after its initial retrieval. PVR occurs after the user has surfed enough to have established a context within which to judge the value of a webpage (Wen, 2003). Thus, contents and structure of information space complement each other: accurate relevance assessment requires traversing the information structure, and development of a representation of the information structure requires knowledge of how information elements are (semantically) connected. Re-visitation behavior seems to be part of a navigation strategy: users explore links and nodes in order to establish a conceptual overview of the information space, and at the same time they mark specific information units to be re-visited if necessary, depending on coherence or comprehension requirements of this developing conceptual overview.

2.2. Individual differences in Web navigation

There is a vast amount of literature showing and analyzing individual differences involved in Web navigation. Thus, Eveland Jr. and Dunwoody (1998) notice 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. Chen and Macredie (2002) show that 'Field-independent' individuals prefer non-linear learning, as opposed to 'Field-dependent' learners who prefer guided, linear learning in a hypertext environment. MacGregor (1999) demonstrated that students who had greater *domain knowledge* evidenced more purposeful navigation and allocated time more variably to different information nodes when they were studying with the aid of hypertext environments. Salmeron, Canas, Kintsch, and Fajardo (2005) demonstrated that low knowledge participants learned more by following a high coherent reading order, whereas high knowledge participants learned more by reading the hypertext in a low coherence order. *Spatial ability* is an important determinant of hypermedia navigation performance, as reported in several studies (Chen & Rada, 1996; Chen, 2000; Gugerty, Treadaway & Rubinstein, 2006; Neerincx, Lindenberg, Rypkema, & Van Besouw, 2000). It has also been shown that individuals with low *spatial abilities* have difficulties in constructing, or do not use, a visual mental model of the space (Stanney & Salvendy, 1995), and they are more directed to the semantic content (Westerman, 1995). Aging is associated with decreases in *working memory capacity* (Sjolinder, 1998) and *computer confidence* (Neerincx, Pemberton & Lindenberg, 1999). Gender appears to involve different navigation patterns: men seem to be better than women at exploring the hyperlinked structures present on Web pages (Roy & Chi 2003). Women report higher levels of spatial anxiety, which is negatively related to the orientation way-finding strategy (Sjolinder, 1998).

2.3. Cognitive processes involved in Web navigation

Besides the behavioral perspectives on Web navigation presented above, there are several attempts to characterize Web navigation in cognitive terms. The most relevant theoretical contributions come from three major fields: language comprehension, information foraging and spatial cognition. Each of these contributions will be shortly introduced in a separate subsection. Mainly the distinguishing characteristics are presented; it is assumed that Web navigation, as a complex behavior, involves basic cognitive processes such as:

- Perception – input information from the environment and interpret it.
- Comprehension – understanding and relating various information elements.
- Reasoning – inferring missing but necessary pieces of information.
- Decision making – selecting between different options.
- Problem solving – searching in a problem space, applying operators on the current state to move to a different state and approaching a goal state.
- Executive (strategic) control – allocating cognitive resources, monitoring progress, switching between tasks, etc.

There are also attempts to specify these cognitive processes in computational terms and the most important models of Web navigation behavior are presented in Section 2.3.4. These models will provide input to the model of Web navigation proposed in Section 4.3.1.1 of this thesis.

2.3.1. Web navigation as a reading comprehension process

According to this view, Web navigation is an iterative process of the same nature as a reading comprehension process. Every new piece of information that is perceived starts a new iteration that eventually updates/reorganizes the existing mental representation (van den Broek, Young, Tzeng & Linderholm, 1999).

The Construction - Integration theory of text comprehension (Kintsch, 1998) postulates a construction phase in which a mental representation is constructed from textual input, reader's goals and prior knowledge, and an integration phase which establishes coherence of the constructed representation via a spreading activation mechanism[1]. Construction is local (context-free) whereas integration is global (context-dependent). Human comprehension is regarded as a top-down and bottom-up

[1] Here is a short explanation of this spreading activation mechanism involved in reading comprehension: as the reader proceeds through a text, she/he constructs an episodic memory representation of the incoming information and uses background knowledge from semantic memory (van den Broek, Rapp & Kendeou, 2005). Since human attentional resources are limited, only a small part of the reader's memory is active at a given moment, that is, only a small amount of knowledge resources can be employed in current processing. There are several sources of activation that determine which concepts are active: the text element that is currently being processed, the preceding recently processed knowledge, the knowledge processed in earlier phases of a reading session, and the reader's background knowledge. Activation spreads from these sources among the concepts of the developing memory representation, causing some concepts to be more active than others.

process (Kintsch, 2005). This theory also makes the distinction between *textbase* representations (mental representations derived directly from the text) and the *situation model* – a mental representation that adds information from the reader's long-term memory to the *textbase*. The meaning of a concept in a particular discourse context is given by its position in the network representing that discourse, enriched with information retrieved from the reader's knowledge net. Also, the meaning is not fixed but must be constructed in every new context. When dealing with potentially ambiguous constructions, readers continue reading hoping that the succeeding text will clarify their problem.

Within the field of Web navigation one can discover analogies with the concepts from the reading comprehension field sketched above. A mental representation of the information space being navigated is constructed, although it is still a matter of debate what the exact nature of this representation is (Farris, Jones et al., 2002). Coherence is less of an issue, but users do relate information elements to each other guided to some extent by the preexisting information structure. Prior knowledge has been shown to be a major factor in successful Web navigation behavior (MacGregor, 1999). Navigation is a combination of goal-directed and *screen-driven* processes, with one type or another being prominent depending on the task at hand. The meaning of a particular information element depends on the assessed semantic value of surrounding information elements (Brumby, 2004). Users reassess the semantic value of particular information elements by re-visiting them after exploring the proximal information environment (post-valued recall) (Wen, 2003).

2.3.2. Web navigation as information foraging. Information Scent

Several approaches to user navigation modeling are inspired by the *information foraging* theory (Pirolli & Card, 1999). Information foraging theory assumes that people have strategies to maximize information gain and minimize the cost (effort) associated with that gain. More specifically, users continuously compare the costs and benefits of alternative actions, for example digging further into a particular information resource versus looking for a different resource. Cognitive models based on this theory assume that selections of users' actions are determined by utility assessments. Users assess meaning of proximal cues such as link labels and make predictions about related information (distal content) (Pirolli, 1995). Information scent is a measure of this subjective assessment of how likely a proximal cue is to lead toward a desired distal content. In terms of Card, Pirolli et al. (2001), "information scent is the (imperfect) perception of the value, cost or access path of information sources obtained from proximal cues, such as WWW links. On a webpage information scent may be delivered by link descriptors, images, contextual clues, such as preceding headings, or by page arrangement". Although it is theoretically valuable, this definition is difficult to apply in more practical situations such as usability evaluation, Web design, user modeling or cognitive modeling. In consequence, an operational definition has been considered necessary. According to this operational definition, information scent is *the assessed semantic relevance of screen objects to users' goals* (Kitajima et al., 2000; Pirolli & Fu, 2003).

2.3.3. Spatial cognition involved in Web navigation

Several authors have mentioned the importance of users' spatial cognition for Web navigation (Czerwinski, van Dantzich et al., 1999; Chen, 2000; Tavanti & Lind, 2001; Tamborello & Byrne, 2005). Since information spaces can be represented spatially (see Figure 1 in Section 2.1), it might come natural that users' spatial cognition is involved. A number of tools incorporated in Web browsers are conceived around the spatial metaphor of Web use. As a matter of fact, even the term "Web navigation" is part of this metaphor and most of the browser tools are called "navigation tools": zoomable and expandable menus, buttons, maps, breadcrumbs, etc. These tools assume a spatial nature of the information space being navigated; they might have *relative position* like in maps, *direction* like in histories, *depth* like in trees or breadcrumbs, etc. The usefulness of such tools is still a matter of debate. There are studies showing both positive and negative effects of spatially inspired tools. For example, McDonald and Stevenson (1999) found that users navigated more efficiently with a spatial map than with a contents list, whereas Goumi, Rouet, and Aubert (2003) found that alphabetically ordered lists are more efficient for finding information than conceptually structured maps.

These results make it worthwhile to ask the question: do users construct *spatial-like* mental representations of an information space or are these representations merely semantic[2]? In an experiment conducted by Farris, Jones and Elgin (2002), participants had to explore a website and afterwards to draw the website's information structure. Analysis of participants' drawings made the authors observe that the website's structure was not represented but the conceptual relations between various information units were. The authors (Farris, Jones et

[2] The term "semantic representation", as opposed to "spatial representation" in this context, refers to mental representations such as propositional representations and semantic networks (Anderson, 1983).

al., 2002) concluded that participants' representations were non-spatial since the website structure was not accurately drawn and, instead, the drawings pictured conceptual relationships. Our interpretation is that users represented the information space structure and not the website structure. The information space structure is more likely to be task relevant than a rather neutral (or task-independent) website structure. The fact that participants' drawings were inspired by semantic relationships does not imply that their mental representations are non-spatial. Participants were still able to draw these relationships in spatial-like configurations. Most probably, users' representations of the information space being traversed are both spatial and semantic and are tailored to the task at hand rather than objectively accurate.

This study also shows that spatial and semantic representations are tightly connected. Users are able not only to draw conceptual structures used in their tasks but also verbalize their tasks in spatial terms (Van Hooijdonk, Maes et al., 2006). It appears that users need to mentally represent not only the semantics but also the syntax – that is, the structural characteristics – of the information space (where the information units are located, how they are connected, and what role every unit plays), and this is reflected in users' language. Web navigation is also an iterative type of task: each step is influenced by previous ones. Users consider the value and cost of more options than those available on screen (Howes, Payne & Richardson, 2002). Assessing relevance of a particular link label to the user's goal depends not only on the user's knowledge about the particular terms used in the label but also on the context of a navigation session, that is, what has been done up to that moment, where the current position is represented in the information space, etc. Thus, semantics and syntax are interconnected.

2.3.4. Cognitive models of Web navigation

There have been quite a few attempts to build models of human performance in human-computer interaction. For example, Kieras and Meyer (1997) argue for the development of a cognitive architecture as a synthesis of theoretical concepts and mechanisms. Practical implications of such architecture can be subsequently explored and tested. ACT-R architecture (Anderson, 1983) has proved to be effective in modeling cognitive tasks involved in human-computer interaction (Byrne, 2003). Cognitive models of Web navigation have originated and been inspired from previous work on menu selection (Norman 1991), menu search and visual search (Byrne, 2001; Hornof & Halverson, 2003), exploratory learning (van Oostendorp & Walbeehm, 1995; Kitajima & Polson, 1997), text comprehension (Kintsch, 1998; Van den Broek, Young et al., 1999), and exploratory choice (Young, 1998).

In the next sections, specific models of Web navigation will be described. ACT-R models (ACT-IF, SNIF-ACT) have been developed for information foraging and Web navigation (Pirolli, 1998; Pirolli & Fu, 2003). Kitajima, Blackmon, and Polson (2000) developed a model in which comprehension of texts and images is assumed to be the core process underlying Web navigation. Another model of Web navigation (Miller & Remington, 2004) that will be described below is not inspired by a specific theory or cognitive architecture but it is rather oriented toward addressing practical needs of Web design. These models are based on the concept of information scent, that is, the assessed semantic relevance of screen objects (such as link labels) to users' goals. They have a computational character, which means they can be run as computer programs trying to generate the same outcomes as an "average" human user would. Comparing real users and simulated users is therefore a way to check the validity of such models.

2.3.4.1. SNIF-ACT: Scent-based Navigation and Information Foraging in the ACT architecture

Pirolli and Fu (2003) developed SNIF-ACT, a computational cognitive model that simulates users performing Web tasks. Their model predicts navigational choices, i.e., where to go next and when to stop (leave the website) based on the concept of information scent. Information scent is calculated as a mutual relevance between the user's goal and link texts based on word occurrences and co-occurrences in the Internet.

2.3.4.2. CoLiDeS: A *Co*mprehension-based *Li*nked model of *De*liberate *S*earch

CoLiDeS (Kitajima, Blackmon et al., 2000) is based on the text comprehension theory of Kintsch (Kintsch, 1998); a similarity in basic principles is assumed between the process of reading a text and navigating through an information space. It explains how users parse and comprehend the content of a webpage and then select what action to perform next.

CoLiDeS compares the user's goal with link texts on Web pages and selects the link that best matches the user's goal (Figure 2). The selected link is clicked on and the process of judging link relevance (information scent) and selecting a link is repeated until the user's goal is attained or the user gives up. The relatedness of screen objects (e.g., link labels) to the user's goal (information scent) is measured based on three factors: semantic similarity, frequency and literal matching. Semantic similarity is calculated based on co-occurrences between words and documents with the aid of a machine learning technique called Latent Semantic Analysis (LSA) (Landauer, Foltz & Laham, 1998). Different LSA semantic spaces are used for different user populations.

CoLiDeS constitutes the theoretical base of a usability evaluation method, called Cognitive Walkthrough for the Web (CWW) (Blackmon, Polson et al., 2002), which is used to identify and repair usability problems related to navigation in websites.

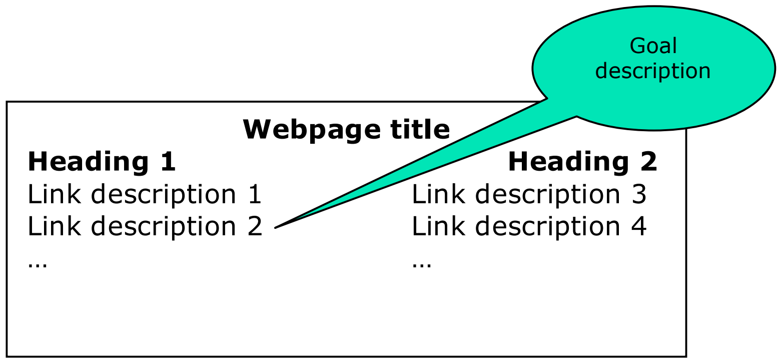


Figure 2. The computational mechanism used by CoLiDeS to model information scent: the LSA semantic similarity is computed between a goal description (a bag of words) and a particular screen object (title, heading or link).

2.3.4.3. MESA: Method for Evaluating Site Architectures

MESA is a computational cognitive model that simulates users navigating an information structure in search of pre-specified targets. Miller and Remington (2004) model the common situation in which link labels are not fully descriptive for their targets or users are not knowledgeable enough to accurately assess the relevance of link descriptions to their goals (information scent). Their model does not give an account for how link relevancies are assessed, but takes them as input. Rather, it focuses on effectiveness of various link selection strategies, given various link relevancies and site structures. MESA gives an account for user’s backtracking behavior and models various navigation styles and strategies to recover from selecting misleading links. By assigning different assessment values, MESA can simulate a variety of users. This is another way to model individual differences than the one used in CoLiDeS.

2.3.5. Value and limitations of existing theories and models of Web navigation

The three models presented above assume that the process of relevance assessment (information scent) is central to Web navigation and they propose explanatory mechanisms and computational instruments to account for this process. It is remarkable that various models originated in different theoretical and practical approaches converge in this basic assumption. It is also interesting to notice that they do complement each other and propose different ways to handle the same issues. Thus, SNIF-ACT uses the whole Web to model users' background knowledge, ignoring individual differences, whereas CoLiDeS proposes specialized semantic spaces to account for differences between groups of users; CoLiDeS and MESA address individual differences in different ways: CoLiDeS by varying semantic spaces and MESA by varying the "noise" associated with relevance judgments[3]; CoLiDeS has a solid theoretical framework while MESA is driven by practical needs.

A major limitation of these models is the relatively narrow conceptualization and operationalization of the information scent concept. They tend to see each judgment of relevance in isolation and ignore the roles of context and history. Thus, they ignore (or abstract out) the current position of the user within a particular page or across pages, the previously viewed information elements and the viewing patterns such as order and frequency of (re)visiting particular information elements. Recent research shows that users' decisions are based not only on the assessed relevance of the currently available screen objects, but also on the relevance of objects that were encountered in earlier steps of the navigation session (Howes, Payne et al., 2002). As mentioned earlier (Section 2.1), backtracking behavior is frequent (Cockburn & McKenzie 2001) and it is involved in the process of judging relevance (Wen, 2003). SNIF-ACT and CoLiDeS do not address backtracking behavior. They model the hypothetical situation of forward linear navigation; in CoLiDeS, backtracking steps are considered erratic actions. When no particular object on the current page sufficiently matches the user's goal, an impasse is said to have occurred. Solutions to impasses are only described and not computationally modeled by Kitajima et al. (2000). Miller and Remington (2004) propose navigation strategies, including backtracking, to deal with ambiguity of link labels or with users' errors in judging link relevance, but the values of link relevancies are not updated as a result of backtracking, as it is the case in the actual use

[3] However, this noise could model two different things: degrees of label quality, or levels of users' domain knowledge, or both.

(Wen, 2003). An attempt to solve this issue is proposed in (Brumby, 2004). It shows that users focus on a limited set of options based on judgments of semantic interdependence between options. However, this proposal is also limited in that it addresses only options at a single webpage level.

In the same line with the previous limitation, these models account more for the semantic dimension and do not give a convincing account for the spatial dimension of Web navigation behavior, while it has been shown – and argued above in Section 2.3.3 – that Web navigation behavior is placed at the intersection between spatial and semantic dimensions (Tamborello & Byrne 2005). Information scent could be increased or decreased based on the position of a particular item or its perceptual salience.

So far, users' background knowledge has been represented as a *static repository* (a certain LSA semantic space (CoLiDeS) or the whole WWW (SNIF-ACT)). There is no convincing account for what happens with the mental model of the user *during* the navigation session. It is conceivable that a user's evaluations of an interface object (e.g., link text) is different at different stages of a navigation session, and the difference is made by the *dynamic knowledge* that is acquired during the interaction.

Traditionally, cognitive models are based on a *common user* assumption. Making cognitive models able to account for individual differences would be the next challenge, given the importance of personalization of content delivery in Web context. Addressing individual differences is the bridge from computational cognitive models to user models of adaptive applications. The model should not only be able to simulate the user's behavior but also work alongside the user and provide individualized support.

2.4. Practical constraints and opportunities

Research on Web navigation cannot be only theoretical. The field has originated in practical needs of human-computer interaction and Web design; practical constraints and opportunities are constantly arising when setting up and conducting studies on Web navigation; and results are always expected to have practical implications.

The main constraint on this type of research is the imperative of ecological validity. While it would be convenient to create simple experimental websites, in which to control every factor, researchers often decide to confront the complexity of real websites, and they try to use realistic Web tasks.

In turn, researchers get a great opportunity to test their theories and models in real life applications; and they make use of continuously improving technological facilities for data gathering and analysis, experimental manipulations, etc.

2.4.1. Web Human Factors and Web Usability

Research on Web navigation is often applied in the fields of Web Human Factors and Web Usability. Web usability extends the area of usability to Web interfaces; it addresses specific problems posed by the network paradigm, such as information overload, disorientation, accessibility, privacy, trust, etc. Besides the technological aspects, there is a particular focus on users and their tasks (Byrne, 1999) – an area called Web Human Factors. Within this area, Web navigation is a central topic (Nielsen, 1989).

For instance, CoLiDeS, the cognitive model of Web navigation presented in Section 2.3 (Kitajima, Blackmon et al., 2000), has been used to identify and repair Web usability problems (Blackmon, Kitajima & Polson, 2003). One of the Web usability evaluation methods called Cognitive Walkthrough for the Web (CWW) (Blackmon, Polson et al. 2002) needs simulations of user behavior to be applied to various Web design concepts, prototypes or applications. CoLiDeS provides such a simulation at almost no cost and relatively fast. Based on CoLiDeS simulations, CWW identifies three types of usability problems: unfamiliar link labels, confusable link labels, and goal-specific competing links/headings. In general, using model simulations beside (or instead of) real users to evaluate Web applications is beneficial especially when involving real users is difficult or too expensive (for example, it is difficult and costly to recruit medical doctors for usability studies) (Ritter, Baxter et al., 2000).

An application area of particular interest for our research on Web navigation is Web accessibility, in particular, using the Web by visually impaired persons (VIPs). VIPs access the Web by the aid of 'screen readers' – tools that read out loud the Web content and options. Since using the Web naturally involves re-visits to certain pages (Cockburn & McKenzie, 2001), blind users would have to repeatedly listen to large amounts of menu options or contents. Therefore, tools are needed to assist users in selecting the relevant information (Di Blas, Paolini et al., 2004). Our research could contribute to conceiving such tools, as we will present in Chapter 5.

2.4.2. Navigation support

A natural outcome of research on Web navigation is in conceiving, testing and improving navigation support. There are various types of navigation support:

- Attention cues. Since the amount of information presented on screen is usually larger than what users can process within their limited attention resources, some Web applications use attention cues such as *highlighting* (or other type of emphasis) to point at the content that is supposedly more important for users to see. For example, the phrase "Buy now!" is often emphasized in passages describing products on commercial websites. Users' attention is triggered by the emphasized item and more cognitive resources are dedicated to it than to non-emphasized items (Tamborello & Byrne, 2005). Moreover, research has shown a beneficial effect of highlighting independent of the relevance of items that are highlighted (Kickmeier & Albert, 2003): highlighting 10 to 25% of items in a text triggers processing of highlighted elements and elements surrounding them, resulting in an overall increase in depth of processing for the entire text.
- Categorization aids. One way to reduce users' information overload is by organizing items based on some meaningful criterion. In particular, menus group similar elements together in a consistent way so that they can always be found in the same place. Users do not need to remember the location of each particular item if the item's category can be easily retrieved. Menus are usually placed at the borders or the screen and they are hierarchically organized. Some menus, called cascading menus, hide levels of their hierarchy and dispatch them as needed when requested by the user. There is an optimum related to the way menu hierarchies are designed, referred to as the depth / breadth tradeoff. It has been shown that users' performance decreases when using either very broad or very deep menu structures (Larson & Czerwinski, 1998). In contrast to cascading menus, indexes are placed centrally on Web pages. They lead to better performance than either horizontal or vertical cascading menus, because the information is directly available without scanning through menu items and layers (Bernard & Hamblin, 2003).
- Orientation aids. Maps are meant to support users in building a mental representation of the information space. It is assumed that users naturally build spatial-semantic representations of information spaces and maps are facilitating this process (Chen & Czerwinski, 1998). When site maps are constantly visible, users make less use of the back button and make navigational movements of greater hierarchical distance (Danielson, 2002).

Efforts are made to find ways of better manipulating and traversing information spaces. Thus, zoomable interfaces allow users to interact with the information space through geometric or semantic zooming (Hornbaek, Bederson et al., 2002).

These navigation support tools are meant to improve navigation performance; however this is not always the case and research is needed to specify what type of support is beneficial and in which circumstances. For example, it has been believed that spatial representations and visualizations of information spaces would increase users' performance by exploiting the powerful human capabilities for spatial cognition (Cockburn & McKenzie, 2002). Based on this belief, a beneficial effect on users' performance would be predicted when increasing the number of dimensions from 2D to 3D (bi-dimensional to three-dimensional). This was indeed the case when a traditional tree hierarchy (2D) was replaced by a sphere-like tree (3D): users made less mouse clicks in vain (Kickmeier & Albert, 2002). However, another study (Cockburn & McKenzie, 2002) found a decrease in effectiveness when changing from 2D to 3D virtual environments.

More research is also needed regarding tradeoffs between different types of outcomes produced by navigation support. Tools might help in one direction and create problems in other directions. For example, an overview was found to increase user satisfaction but also task execution time (Hornbaek, Bederson et al., 2002). Providing a graphic organizer (map) lead to better results in the learning phase of a particular task and worse results in later phases (Nilsson, 2002) or to worse learning results for users with low prior knowledge (Hofman & Van Oostendorp, 1999). It seems that making it too easy for the user has a detrimental effect. Also here results from reading comprehension research are valid in the field of Web navigation: actively involving the reader in representing the text is more beneficial for comprehension and remembering than just providing a graphical representation already elaborated (Montanero, 2004; see also Salmeron et al., 2005). Providing support always produces changes in the way users perform their tasks. Users adapt (e.g., by changing their strategy) to the inclusion of a tool in the task environment. Before a particular tool is recommended to the user, its consequences on user's behavior and task outcomes need to be evaluated.

2.4.3. User models and personalization

A large number of Web applications is designed for a general audience with varying goals. As it is hard to satisfy all categories of users with one design, these applications try to personalize content or link structure. A personalized application needs to develop a model of relevant characteristics of the user. A user model can include (relatively) stable characteristics such as gender, age, education level, and dynamic (changing) characteristics such as goals, preferences and behavior. For example, recommender systems build user models based on users' surfing behavior (Kim, 2003). Traditional techniques for personalization of Web applications involve link hiding, sorting, annotation, direct guidance and hypertext map adaptation (Brusilovsky, 2001).

Of a particular interest for us are user modeling and personalization techniques that are inspired by cognitive models of Web navigation. For example, SNIF-ACT (Pirolli & Fu 2003) has inspired an adaptive tool called ScentTrails (Olston & Chi 2003). ScentTrails highlights links to indicate useful paths to the user based on link semantic relevancy to keywords entered by the user. Later on (Section 4.4) we will propose navigation support that takes into account the user's navigation path. Links will be highlighted if they are relevant to the user's goal and semantically consistent with the user's past selections.

Conclusion

In conclusion to this chapter, Web navigation is a specific behavior involved in the use of Web-based applications. These applications provide access to what we have called 'information space'. Users need to handle both the structure and the contents of this information space. Differences between users regarding their navigation behavior as presented in the literature were reviewed. Attempts to describe the Web navigation behavior in cognitive terms have also been presented. Research on Web navigation is guided by practical constraints and opportunities, such as the need for better support of Web navigation behavior and better access to Web resources.

Chapter 3. Methodological issues

Studying Web navigation behavior is a relatively new endeavor. Usually, in such cases researchers rely on existing paradigms borrowed from related disciplines. As announced in the introduction, we rely on the fields of human factors, human-computer interaction, cognitive science and reading comprehension. However, most of the time these fields are only used as sources of analogies and the borrowed concepts are adapted to meet the specifics of Web navigation. In this section we discuss general methodological issues; details about particular methods are presented together with the studies in which they are applied.

3.1. Measuring and analyzing Web navigation behavior

One important methodological aspect is specifying the proper units of analysis. In our case the largest unit of analysis is a navigation session. Other authors (Herder, 2005) go further than the session unit, analyzing navigation patterns occurring across sessions or days. While this is definitely a worthwhile venture, we decided to focus on sessions and especially on smaller subdivisions of them. A navigation session may consist of several tasks. A task specifies what the user has to do, what is required or expected from the user. Activity is what the user really does in order to accomplish the task. Behavior is the external, observable and measurable part of activity. Each task has a goal, which is an external state to be attained. Users have internal states called intentions corresponding to goals. Activities are driven by intentions toward achieving tasks’ goals. We will analyze:

- tasks with their goals, requirements, and outcomes;
- users with their intentions, knowledge and aptitudes on one hand, and their activity and behavior on the other hand;
- Web environments and their contents, structure and functionality (tools).

Another methodological aspect involved in studying Web navigation is related to the way data is collected. Besides the classical methods of observing, testing and questioning the user, Web navigation studies benefit from automatic recording of user actions and characteristics of the task environment. Automatic recording (also called Web-logging) is justified not only by the need to study user behavior but also by the need to model the user in order to personalize Web applications; it can be applied during the normal use of an application without interfering with users’ natural behavior; data can be gathered continuously for long periods; it can be automatically analyzed; and it releases the user input effort. However, there are also problems associated with Web-logging: samples are often self-selected, it is not always clear how ethical standards of conducting research are respected, and Web-logging data is rather difficult to analyze and interpret.

3.1.1. On-line collecting and mining of Web navigation data

Web navigation behavior can be described with reasonable accuracy based on Web-logging data. However, Web-logging data is only potentially meaningful. Making sense of Web-logging data (also called Web mining) requires careful analysis and interpretation. There are different levels of analysis and different amounts of value one can extract from Web-logging data. For example, it is relatively easy to notice that page re-visits are very common in Web navigation, but this observation has little value in itself. In contrast, it is more difficult and more useful to discover re-visitation *patterns* that answer questions such as: why do users re-visit pages, in which moments of task execution, etc.

Collecting Web-logging data is justified when one has the aim of personalization in mind: analysis of this data can be automated, and adaptive reactions of the application can be based on results of this analysis (Brusilovsky & Maybury, 2002; Jameson, 2003; Juvina, Trausan-Matu et al., 2002).

Interaction events that can be logged during a navigation session are quite numerous: page downloads, view time, use of buttons, etc. Some data about the Web structure being navigated is also available: page title/URL, number of words per page, number of outgoing/incoming links, etc. This constitutes the raw data of Web navigation, which progresses toward information and knowledge via analytic and interpretational processes.

A number of analyses can be performed on the raw data in order to extract some useful *information* out of this data. The results of these analyses are referred to as *navigation metrics*. Extracting information out of navigation data by the aid of various navigation metrics is the way toward acquiring *knowledge* about the user.

In Chapter 4, Section 4.1, metrics that extract relevant information from navigation data will be presented. Here we propose the conceptual framework that has been used in generating and classifying these metrics.

In characterizing users' Web navigation behavior we have used the distinction syntactic-semantic-pragmatic adapted from linguistics. Navigation takes place in an information space and has a particular structure. In pursuing their goals, users follow *paths* through this space.

Paths are linked information elements (e.g., Web pages) that are deliberately selected by the user. One can identify *syntactic* features of navigation paths and syntactic roles of information elements in a path: order of selections, arrangement of pages, pages having specified roles (home, index, hub), etc. Navigation paths can also be *semantically* characterized by, for instance, page contents. Finally, relating navigation paths to users' tasks and goals is a way to *pragmatically* characterize user's navigation behavior.

Syntactic information extracted from navigation data indicates how users move across the information space, what types of links are followed, in which order, in which manner (e.g., linear or nonlinear). *Semantics* indicate what contents the user encountered and which of these contents the user selected and processed. *Pragmatic* information refers to users' goals, tasks, interests, and preferences.
In short, *syntactic* means structural, topologic information, *semantic* refers to the content of visited pages, and *pragmatic* information indicates the reasons and gains for the user to visit certain pages.

This distinction is important because it shows what is required for an analysis to be complete or, conversely, what is missing from a particular approach. For example, a large number of Web navigation metrics are merely syntactic (Botafogo, 1992; Smith, 1996; Ivory, 2001), ignoring the semantic aspects of users' behavior. Semantics of Web-logging data are not to be ignored, as they have been shown to be useful in explaining user behavior: Segal (2000) demonstrated that a content-based machine learning technique was more accurate than a prediction of user actions based only on frequency of selection of various options.

At the other extreme, cognitive models of Web navigation ignore the syntactic dimension affecting plausibility of these models and their correspondence with empirical findings. For example, CoLiDeS only addresses the semantic and pragmatic dimensions of Web navigation behavior. However, a combination of spatial and semantic abilities makes a significant difference in users' performance in Web tasks (Chen, 2000). Spatial abilities are probably involved when users mentally represent the structure of the information space and operate on this representation, that is, when *syntactic* features are involved. Semantic abilities are mainly used to judge the goal-relevance of particular information elements. But syntactic and semantic abilities are complementing each other. No one is self-sufficient: a very accurate representation of the information space is useless if goal-relevance of particular information elements cannot be assessed, and judging goal-relevance of a particular information element necessitates locating and assessing related information elements. Howes, Payne and Richardson (2002) argue that a search strategy is required in addition to an ability to follow label semantics because in real Web navigation label semantics are rarely sufficient to guarantee that users will navigate directly to the location of a goal without exploration of other parts of the search space.

3.2. Cognitive modeling of Web navigation behavior

3.2.1. Modeling as a methodological approach

A model is an intermediary between an object in the actual world and something abstract such as a principle, theory or law (Sterrett, 2002). For instance, in human-computer interaction, it is often difficult to make the link between the abstract theories of cognitive science or information science and a particular situation in which a particular person uses a particular system. To overcome this difficulty, researchers build models of the interaction; these models are abstract enough to be able to test the underlying theories but concrete enough to (partially) map the actual world.

In our case, modeling abstracts out some characteristics of the user and task environment while it focuses on others. The essential decision of which characteristics to represent in the model versus to abstract out was mainly based on exploratory research as we will present in Chapter 4 Sections 4.1 and 4.2. However, practical reasons and project feasibility had also a role in selecting what the model is to focus on. For example, we have abstracted out the issue of how the model is to handle user intentions (how they are formed, monitored and updated) since this is a very difficult problem to address in such a limited time.

A cognitive model of Web navigation must specify how users mentally represent the structure and semantics of the information space, how they operate on these representations, what type of information is used at each particular step in the process (e.g., contextual, incoming and background information), etc. When possible, it is useful to specify such concepts in a computational way, which means to implement the model as a computer program and simulate the users' behavior. In this phase, there are important modeling decisions as well. For example, one way to model contextual information involved in assessments of relevance is to consider the content of all pages viewed by the user up to a certain moment. Another way would be to consider only user selections out of all the inspected content. While the first modeling solution seems comprehensive, the second one is likely to be more appropriate, since Web pages contain a lot of redundant information and most of the time they are not exhaustively inspected. Nevertheless, the second solution is not entirely accurate since it does not include content considered by the user and not selected. In the same line, one could try to model the complete process of Web navigation, including reading text passages, or only the link-following behavior. Thus, most of the time, the modeler needs to take this kind of compromising decisions. Not all researches agree that such compromises are worthwhile. Some of them prefer to work either at a pure theoretical level or at a pure experimental level, without considering the intermediary level of modeling. We consider theory, modeling and experimentation as complementary to one another. Cognitive modeling allows specifications and simulations of theoretical concepts, generating hypotheses, building experimental setups, and comparisons with empirical data to check these hypotheses. Experimentation has its feet in the real world, while theories ensure generalizations. Modeling can predict empirical results that are difficult or expensive to gather via normal experimentation. On the other hand, new empirical results incite novel modeling attempts.

3.2.2. Common user versus individual differences

Quite often cognitive models assume a prototypical user who collapses individual differences. However, individual differences involved in Web navigation have to be taken into consideration when building cognitive models aimed at simulating real users' behavior. Addressing individual differences is also justified by the practical desideratum of building user models of adaptive applications. A cognitive model that is able to simulate the behavior of a particular user could be also used as a user model. As mentioned in Section 2.3, there are notable attempts to model individual differences in Web navigation. The most important ones are addressing individual differences in background knowledge (Kitajima, Blackmon et al., 2000), relevance assessments and memory capacity (Miller & Remington, 2004; Miller & Fucks, 2005).

Making cognitive models able to account for individual differences in spatial cognition would be a next challenge. Given the same task, different users access different contents to achieve it. Using a spatial metaphor, users take different *paths* to the same target. The concept of *path adequacy* that we propose in Chapter 4 might be a starting point in addressing issues of spatial cognition in modeling Web navigation.

3.2.3. Cognitive modeling as source for support

In general, support systems need some sort of a model of the user. Let's take a simple example from the field of Intelligent Tutoring Systems: let's assume a child learns how to solve a particular math problem and we want to build a system to support that task. A computer could probably solve that particular problem in almost no time, instantly. Would that be helpful for the child? On the contrary, we probably want to build a tutoring system based on the way children solve math problems.

In the field of using the Web we have tools as search engines that do a wonderful job. A search engine presenting all possible information sources, sorted by goal-relevance, could take the user directly to the target information. Why would we want more than that? Because using the Web is not always about simple information retrieval tasks. The Web is quite often and increasingly used for more complex and rich tasks than finding a particular piece of information. In environments where knowledge acquisition is important, presenting information in context and facilitating a particular user experience are essential. It can be argued that using a cognitive model as generator of navigation support would preserve the user's experience. A cognitive model that simulates these processes and works alongside the user on the same task could intervene in a flexible and intelligent way, for example only intervene when the user is having serious difficulty, thus preserving users' natural experience.

A question one could ask is related to the effectiveness of using a cognitive model as navigation support. A cognitive model is designed to simulate real users, thus it should for instance make the same errors as real users do. How could such a model effectively support users? Usually a computational cognitive model is initially conceived to give an account for human *competence* in solving a particular task. Then, human-like limitations are considered to make the model accountable for human *performance*. Thus, cognitive models have the capability to solve tasks in an ideal, error-free way, and in addition they can simulate human error-prone behavior. This makes them able to generate support that effectively helps while preserving users' ways of executing tasks.

A supporting agent should have both a model of ideal performance (competence) and a model of (actual) performance. The performance model allows the agent to recognize (diagnose) what the user is doing and why. The competence model allows the agent to figure out what the user should be doing, and intervene when the user is in trouble.

An example from the field of tutoring systems: the student makes errors when adding numbers like 34 and 19. The erroneous answers are always smaller than the correct ones and the difference is always 10. A human teacher is able to immediately figure out where the problem is. But a tutoring system can do that only if it knows how students add numbers (they put them one below the other, then add the digits in the right-most column, if the result is higher than 10, 1 is added to the next column, etc.).

3.2.4. Computational techniques: Latent semantic analysis (LSA)

Since in the next chapters a computational technique called LSA will be used, a general description of this technique is necessary and is provided here.

LSA is a model of language learning based on exposure to texts and a tool for extracting semantic information from texts. It is used for: information retrieval, cognitive modeling, usability, user/student modeling, essay assessment, etc. (Landauer, Foltz & Laham, 1998). In information retrieval LSA overcomes some of the limits of keyword search: documents can be retrieved even if they do not share any words with the search query. In cognitive modeling LSA is used as an underlying mechanism of language learning and to simulate a variety of cognitive phenomena that depend on word and passage meaning. In one of the usability evaluation methods, cognitive walkthrough for the Web (CWW), LSA is used to objectively estimate the degree of semantic similarity (information scent) between representative user goal statements (100-200 words) and heading/link texts on each webpage (Blackmon, Polson et al., 2002).

From a computational perspective, LSA is a machine learning technique that builds a semantic space representing a given user population's understanding of words, short texts (e.g., sentences, links), and whole texts. A semantic space is generated from a large corpus of written materials (including books, magazines, and newspaper articles) read by typical members of that population. Words not included in the corpus are not represented in the semantic space. LSA generates the semantic space by applying singular value decomposition, a mathematical procedure similar to factor analysis, to a huge terms-by-documents co-occurrence matrix.

The first step is to represent the text as a matrix in which each row stands for a unique word and each column stands for a text passage or other context. Each cell contains the frequency with which the word of its row appears in the passage denoted by its column. This matrix is called the initial matrix of co-occurrences. Secondly, the cell entries are subjected to a preliminary transformation, in which each cell frequency is weighted by a function that expresses both the word's importance in the particular passage and the degree to which the word type carries information in the domain of discourse in general. Thirdly, singular value decomposition is applied to reduce the number of dimensions in the initial matrix of co-occurrences (Figure 1). It is this dimensionality-reduction process that induces semantic similarities between words. The meaning of a word is given by the contexts (paragraphs) in which it appears and the contexts in which it does not appear (Landauer &

Dumais, 1997). For example, if a word (e.g., bike) statistically co-occurs with words (e.g., handlebars, pedal, ride) that statistically co-occur with a second word (e.g., bicycle) *and* the first word statistically does *not* co-occur with words (e.g., flower, sleep) that do not co-occur with the second one, then the two words are considered quite similar (Zampa & Lemaire, 2002). Once a semantic space has been built, the meaning of a word, sentence or any text is represented as a vector in a high dimensional space, typically with about 300 dimensions. The degree of semantic relatedness or similarity between any pair of texts is measured by the cosine value between the corresponding two vectors. Cosines are analogous to correlations. Each cosine value lies between +1 (identical) and -1 (opposite). Near-zero values represent two unrelated texts.

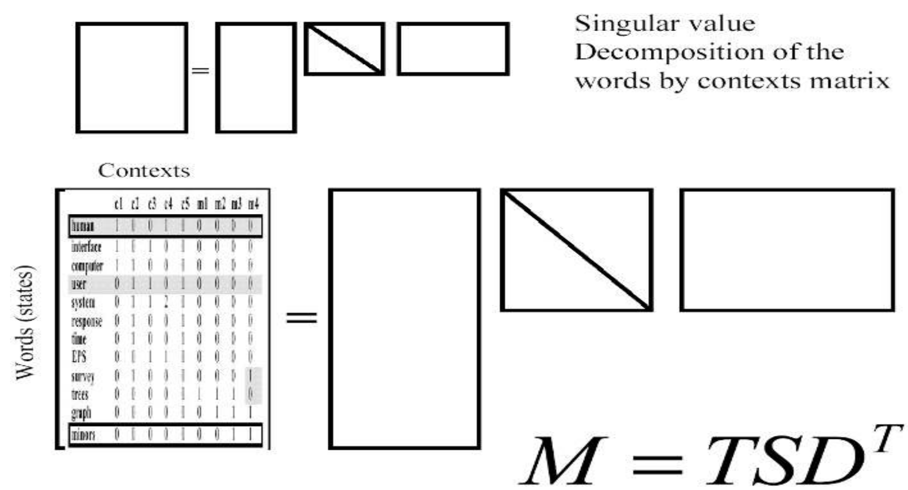


Figure 1. An illustration of the *singular value decomposition* technique (Landauer & Dumais, 1997). The initial matrix is decomposed in three matrices, one representing loading of k dimensions (factors) for words, one representing loading of k dimensions (factors) for passages (contexts), and the third a set of singular values (eigenvalues). The singular values (a diagonal matrix) have decreasing magnitudes, thus a limited set of them is enough to account for most of the variance in the whole set. As a consequence, the lowest singular values are dropped (set to zero). Then the initial matrix is recomposed by multiplying the three composing matrices of which the diagonal one is slightly modified (some of the values are set to zero).

An interesting feature of the method is that the semantic information is derived only from the lexical level. It assumes that word choice is more meaningful than word order. There is no need to represent a domain theory by means of a semantic network or logic formulas. This feature makes LSA valuable in cases where it is very difficult or impossible to construct a fully specified ontology of the task domain (e.g., using the Web). However, the same feature is also a restriction of the method, this is why LSA misses important features of language such as: order, predication, quantification, anaphora, negation, etc. It models reasonably well single words and paragraphs but not so well sentences.

Conclusion

In conclusion to Chapter 3, studying Web navigation faces challenging methodological issues. The way to approach this challenge, which we have adopted here, is to borrow concepts and techniques from related disciplines and adapt them to the specifics of Web navigation. Our approach is rather comprehensive including experimentation, statistical analysis and modeling. A particular emphasis is given to automation in data collection, analysis and modeling, to the benefit of real-world applicability.

Chapter 4. Model Development and Empirical Studies

Our approach in modeling Web navigation was driven by practical criteria: it started from practical requirements and constraints of designing Web applications. We began with studying application domains, identified tasks, criteria and determinant factors (Sections 4.1 and 4.2). Based on the most relevant factors identified by means of task analysis and correlational research, a more analytic process was employed to find concepts and mechanisms that can explain the influence of those factors. Based on these concepts, a cognitive model was developed (Section 4.3). Then, a couple of experimental studies were performed aimed at testing the model and its practical relevance (Sections 4.4 and 4.5).

4.1. Task analysis

Task analysis has the goal of building a task model. With the aid of various methods and tools, information about how users perform (and/or should perform) a task is collected and represented. The task model provides a useful basis to designing usable technology; knowing relevant aspects of the tasks is fundamental to the design process (Paterno, 2002). For example, one way of detecting usability problems in Web interfaces is to compare actual behavior of the user with some ideal, optimal or expected action sequences as specified in a task model (Hilbert & Redmiles, 2000). Paganelli and Paterno (2002) record events in the interaction between user and system by the aid of a logging software and compare them with tasks specified in the task model. Then, the degree of matching between tasks and user actions is used as a usability measure of a website (Paganelli & Paterno, 2002).

Task models are also useful in designing personalized applications (Trausan-Matu, Iosif et al., 2002). They help in constructing an adequate user model and providing appropriate support. For example, identifying errors among user actions is possible in an automatic way only if the application has a model of correct task execution (competence) to be compared with the actual user behavior (performance).

4.1.1. Task domains, instances and criteria

Task analysis methodology was employed in this research in order to:

- Define areas for generalization and applicability (task domains)
- Collect realistic task instances
- Describe and measure Web navigation behavior.

We studied real Web applications, using realistic tasks. By Web applications we refer to websites that are capable of interaction with users in order to provide them with certain services (e. g. Amazon.com). Regarding the last aim, considerable efforts were dedicated to finding adequate measures of user navigation behavior. These measures must be compatible with requirements of Web applications, that is, to be based on data that is easy to collect in real time and unobtrusively. We focused on data that was automatically recorded as a byproduct of a navigation session (Web-logging data).

As mentioned before (Section 2.1.1), we have focused on goal directed and performance oriented Web tasks, since users frequently report them as relevant (Morrison, Pirolli et al., 2001). These tasks can be easily encountered in domains such as: Personal Finance, E-commerce, E-learning, Travel, and Weather forecast. We have selected such domains since they have large applicability and they are relevant for our target population (university students). Furthermore, results can be easily generalized to the general population of Web users. The websites used in this project allow free exploration and use without asking users to register or to disclose personal data.

Realistic task instances had to be collected and tested. Realistic tasks are weakly defined (Pirolli & Card, 1999). Their goals are not always well specified. Nevertheless, we selected tasks with a fairly clear goal to be attained and give them to our subjects in a written form. Realistic tasks are also complex. We selected tasks that were more complex than just locating information on the Web. A certain amount of planning (e.g., "Setup a personal budget"), problem solving (e.g., "How much do you need to save per month in order to buy a car in 4 years") or decision-making (e.g., "What kind of car can you afford") was always involved besides searching for information. As a matter of fact, finding a particular piece of information was considered to be merely a means toward attaining the task goal; information was to be collected along the path and not only at the end. As a consequence, tasks required a fair amount of navigation (in the order of 10-20 pages per task) and a considerable duration to be completed (about 10 min). An example of tasks used in our research is presented in table 1. It places the user in a particular situation that motivates certain actions, given particular constraints (facts).

Table 1. Task description.

Motivation / context: You want to buy a new car but you don't have enough money. The Internet has made getting a loan much easier because it provides you with the resources you need to find the right loan, the approval process is quicker, there is less documentation that needs to be filled out, and you can do it in the privacy of your own home. You want to shop around for an auto loan lender, find an attractive interest rate, and find out how much your monthly payment will be. Also you wonder what will happen if you become unable to make your payments due to various conditions (sickness, etc.). **Given facts:** You cannot afford to pay more than 180 pounds per month for more than 48 months. **Use http://www.alliance-leicester.co.uk/ to:** • Calculate how much you can loan. • Calculate how much your monthly payment will be. • Look for one way to handle situations when you cannot pay.

Performance criteria were used to judge task outcomes. The intention was to find a small number of criteria to cover as many task outcomes as possible. *Effectiveness, efficiency* and *satisfaction* were taken from ISO9241-11 (ISO 1991) and effectiveness and efficiency were grouped under the label *performance*. *Performance* denotes task success (effectiveness) obtained with minimum resources (efficiency). Satisfaction refers to users' affective experience toward task execution and task results.

Besides *performance* and *satisfaction*, another criterion was considered necessary to cover the undesirable aspects of task outcomes. There is a vast literature showing that models of human performance are incomplete if they consider only correct performance and neglect human error or, more general, human fallibility. For example, Reason (1990) states that correct performance and error are like *active* and *passive* sides of a cognitive balance; each *debit* has a corresponding *credit*. For instance, skills development increases performance but also the risk of error, by turning off the conscious control mechanisms (Reason, 1990). In the field of human-computer interaction, Van Oostendorp and Walbeehm (1995) argue for the necessity of (and propose modeling techniques for) considering errors, inefficiency and problem-solving processes in modeling human behavior in interaction with direct manipulation interfaces (Van Oostendorp & Walbeehm, 1995). Since the work presented here takes into account not only errors, but also other undesirable aspects of task execution, such as cognitive workload, disorientation and frustration, a more generic term was chosen, namely *reliability*. In this context, reliability refers to avoiding or minimizing negative outcomes of task execution.

These 3 criteria reasonably cover the whole range of task outcomes. However, their general definition as stated above needed specification to be applicable in concrete studies (table 2). Thus, *Effectiveness* (task success) was measured based on experimenter's ratings from 0 to 4 based on correctness and completeness of participants' answers. *Efficiency* was measured as task completion time. *Satisfaction* was measured by questionnaire items such as "It was an interesting experience to perform these tasks" and "Overall, working to accomplish these tasks was satisfying". *Reliability* was operationalized in this project by the variables *perceived disorientation* and *frustration*. *Perceived disorientation* was measured directly with questionnaire items adapted from (Ahuja & Webster, 2001), such as "It was difficult to find the information I needed on this site" and "It was difficult to find my position after navigating for a while". According to Ahuja and Webster (2001), it is easier, more accurate and quicker to measure users' perceived disorientation directly than to infer it from their actions. *Frustration* was measured with items such as "I felt frustrated when I encountered difficulties in completing the tasks", and "I felt angry when I couldn't find what I needed to complete the tasks".

Table 2. Criteria for task outcomes

Criteria	Performance	Satisfaction	Reliability
Operationalization	Effectiveness Efficiency	Satisfaction with task execution and results	Disorientation Frustration

4.1.2. Metrics of navigation behavior[4]

Graph theory and graph-like visualizations are usually employed to characterize Web tasks and user's behavior (Card, Pirolli et al., 2001). One way to represent the task environment (information space) is by making graph nodes stand for Web pages and graph arcs stand for links between pages. User navigation can then be represented as a sub-graph where nodes and arcs are, respectively, pages visited and links followed by the user. This simple modeling approach captures essential characteristics of Web use. It can be used to model different levels of abstraction on a variable scale. For example, a node can represent only the page title or it can contain all information on a particular Web page;

[4] These metrics were developed in collaboration with Eelco Herder from University of Twente (Herder, 2006).

links can be represented in different colors based on use frequency, etc.; a graph can represent a single user or it can accommodate multiple users; for one hour or for one day, etc.; all of these depending on the modeler's particular goal. However, increasing complexity usually makes these graphs difficult to interpret (they look like "plates with spaghetti").

Mathematical techniques can be employed to make sense of navigation graphs. This section will present a number of navigation metrics that we have calculated and used in further analysis to describe user Web navigation behavior and to ground our modeling decisions. The rationale for computing these metrics and the framework used in classifying them have been presented in Section 3.1.1.

The raw data for computing navigation metrics consisted of:

- Interaction events. For each *navigation session*, the following data was collected: Page visited and link followed; Time of visit; Page load time and length of visit; Navigation action (e.g., link, address bar, refresh, back button).
- Site structure data. For each *page* the following data was collected: URL and host; Size in bytes; Number of words and images; Number of outgoing links; Title and author; Source code. For each *link* the following data was collected: Source and target page; Text associated with the link; Type of link (internal, external).

Based on these raw Web-logging data, two types of *syntactic* metrics (first-order and second-order metrics) and one *semantic* metric (path adequacy) were calculated as presented below. These metrics were used as indicators of user behavior for the analyses presented in the next sections. In this section only the analyses needed to define the metrics and to express them in behavioral and cognitive terms are briefly described. Details of subjects and procedure for data collection are presented in Section 4.2.

4.1.2.1. First-order metrics

The navigation metrics described below are labeled *first-order metrics* because they are derived directly from the raw data, without taking into account any relationship between them. For example, Average Connected Distance (ACD) was calculated independent of Back Button Use (BBU), and we did not take into consideration the fact that low values on ACD were associated with high values on BBU and vice-versa ($r=-0.49$). This latter information was used in calculating second-order metrics.

Path length (path length[5]) is the number of pages that the user has visited during a navigation session, including page re-visits.

The relative amount of re-visits (re-visits) is calculated as the probability that any page visited is a repeat of a previous visit. The following formula was used: one minus the ratio between the number of distinct pages visited and the total number of pages visited (Tauscher & Greenberg 1997).

Back button usage (back button) indicates the percentage of back button clicks among all recorded navigation actions. It includes backtracking multiple pages at once using the back button.

Page return rate (return rate) indicates the average number of times that a page is re-visited. For a particular navigation graph, return rate is calculated by averaging the number of visits to all pages that have been visited at least twice.

The relative amount of home page visits (homepage) is the number of times the home page (index.html) is visited. "Relative" refers to a correction of home page visits based on path length (the number of home page visits is divided by path length).

View time of a page was calculated as the duration between the moment immediately after the page was loaded and the moment of a new page request. For a navigation graph (several pages visited by a user) several metrics related to view time were calculated:

- The average (mean) view time a user spends on a page (avgview). A small number of pages that are carefully inspected affect the value of this metric.
- Standard deviation of view time per page (devview). It indicates variability in allocating view time across pages on a navigation graph. A user with a large devview can be a very slow viewer on some pages and very fast on others.
- Median view time spent on a page (medview). Since users generally spend only little time on the large majority of pages and relatively more time on a limited number of pages, the median view time is a better indicator of central tendency.
- Difference between mean view time and median view time (meanmed). It indicates that a relatively small number of pages are viewed for a relatively long time. A high value on this metric was hypothesized to indicate user's tendency to carefully (or just slowly) read contents.

---

[5] Short names given in brackets will be used later to refer to these metrics.

- The average view time on large pages (viewlarg). Pages with size (number of words) higher than one standard deviation above the average size were considered large pages.
- The average view time on small pages (viewsmal). Pages with less than one standard deviation below the average size were considered small pages.
- The average view time on index pages (viewindx). Pages with more than one standard deviation above the average number of outgoing links are considered index pages.
- The average view time on reference pages (viewref). Pages with less than one standard deviation below the average number of outgoing links are considered reference pages. Reference pages are assumed to mainly deliver content.

The number of links followed per page, also called **fan degree** (fandeg), represents the average number of connections (links) per state (page) (Rauterberg, 1996). In the case of user's navigation graph, fan degree represents the ratio between the number of links followed and the number of distinct pages visited.

The number of cycles (cycles) is the number of linear independent cycles of a graph (Rauterberg, 1996). The **number of cycles of a navigation graph** is calculated as the difference between the number of transitions between pages 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.

Net density (density) is the actual net density compared to the maximal possible net density (Rauterberg, 1996). The **path density** 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.

The average connected distance (ACD) is the average length of the shortest path between any two nodes. The longer the ACD the more steps (e.g., link clicks) are necessary to reach pages from one another. In other words, a high ACD indicates that users do not return to a page very soon, but only after having browsed for a while. They also tend to return using a link rather than using the back button.

Each of these first-order metrics is indicative for Web navigation behavior when properly interpreted. However, they provide limited information by themselves and invite for further investigation. For example, a high return rate might be interpreted as an extensive use of navigation landmarks, but other interpretations are possible as well. To determine which interpretation is correct one has to look at the type of pages that are frequently re-visited. If they are mainly index pages the first interpretation holds; otherwise an alternative interpretation should be found.

4.1.2.2. Second-order metrics (Navigation styles and Reading time)

Two different data analysis approaches were employed in deriving second-order navigation metrics: unsupervised and supervised learning (Huang, Kecman, & Kopriva, 2006). In the unsupervised learning approach, only patterns of covariance in the first-order metrics are considered, regardless any outside criterion. The second-order metrics resulted in this way were called navigation styles. They are completely specified (numerically) by first-order metrics. However, interpreting their meaning and labeling them was based on their correlations with user characteristics and task outcomes.

In the supervised learning approach, the task outcomes defined in Section 4.1 were used as outside criteria in the attempt to combine the first-order metrics. A second-order metric was derived in this way and was called *reading time*. It is a combination of several view time metrics weighted in such a way as to ensure a significant correlation with *task performance*.

Navigation Styles

A principal component analysis with *equamax* rotation was run on the first-order metrics presented above. Details about participants and data collection procedure are presented in Section 4.2. A 4-components solution that accounted for 85.95% of the initial variance has been selected (all the other components were discarded because their eigenvalues were lower than 1). Each component accounted for 27.3, 23.8, 22.8, and 12.0 % of variance, respectively. Component loadings in first-order metrics and the correlations between factors and user characteristics and task outcomes were used to interpret the content of each factor in terms of navigation styles, as follows.

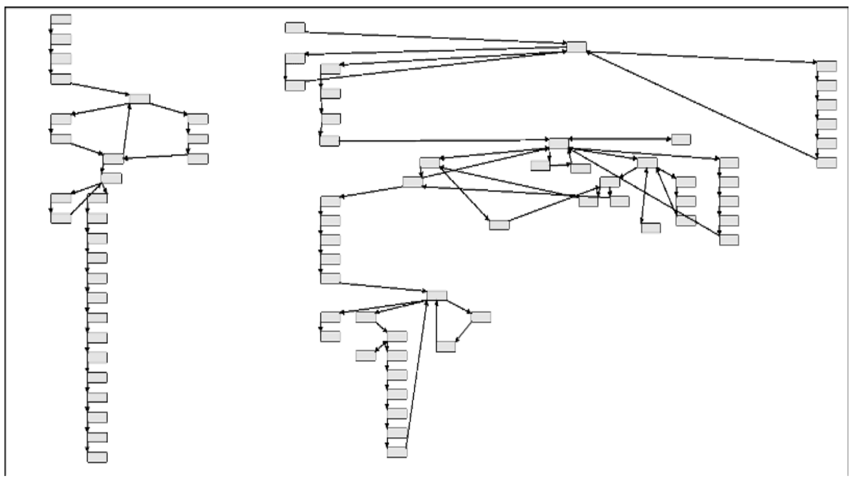


Figure 1. Flimsy (left) versus sturdy (right) navigation. From the figure it can be observed that flimsy navigation is characterized by short navigation paths and a low number of cycles in the navigation graph. The page re-visits that did take place in the flimsy navigation path were made using the back button

Component 1. Flimsy Navigation (fig. 1). This appears to be a parsimonious navigation style. The navigation path was not very elaborate, most of the navigation taking place around the homepage. Time was spent more with processing content than with figuring out the hyperstructure that showed where the relevant information was. High scores on this component were associated with small number of pages visited (r=-0.80), high path density (r=0.80), high median view time (r=0.77), low average connected distance (r=-0.70), low number of cycles (r=-0.53), high rate of home page visiting (r=0.48), and high frequency of back button use (r=0.39). Regarding the correlations between Flimsy Navigation and other variables, a high score on the flimsy navigation style was associated with low Internet expertise (r=-0.5), low active mood (r=0.48), low working memory capacity (r=-

0.38), external locus of control (r=-0.37), and high perceived disorientation (r=0.46).

Component 2. Content Focus. This component grouped together all the view-time metrics. High values on this component indicate high view time on a rather small set of pages. Users with high scores on this style looked for those pages that ought to be read and carefully read them. This style was not associated with any user characteristics or task outcomes.

Component 3. Laborious Navigation (fig. 2). This style involved intensive use of navigational infrastructure provided by the site. Users seemed to employ a trial and error strategy. They followed links just to see if they were useful or not. They figured out quite fast when paths were not leading towards their goal and returned. Re-visits were quite numerous but they were not redundant: once a page was re-visited a different link was followed, it was just another trial. High scores on this component were associated with high number of links followed per page (fan degree) (r=0.95), high number of re-visits (r=0.94), high number of cycles (r=0.79), high returning rate (r=0.73), high use of back button (r=0.71), high density (r=0.43), high number of pages visited (path length) (r=0.40), low average connected distance (short returns) (r=-0.39). Regarding correlations with other variables, this navigation style was associated with high episodic memory (r=0.49), low spatial ability (r=-0.40), and low interest in entertainment (r=-0.38). This style indicates the type of re-visitation that does not relate to disorientation. The user needed to look around for a while until she/he had a good representation of the site structure, because she/he had a weak spatial ability. Yet, her/his memory prevented her/him from making redundant re-visits. This component shows how people compensate for the lack of spatial ability by effort and memory, and do not necessarily decrease performance (no correlation with task performance was found, although spatial ability and task performance were positively correlated). It also shows why re-visitation is not always associated with disorientation.

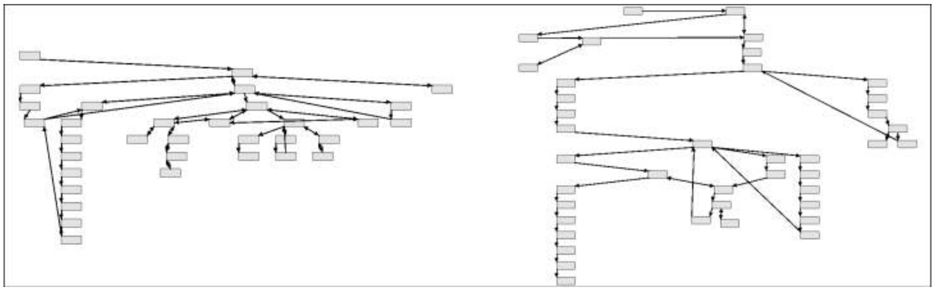


Figure 2. Laborious (left) versus non-laborious (right) navigation. From the figure it can be observed that the laborious navigation style is characterized by a high amount of re-visits, with some pages clearly functioning as navigational landmarks.

The term "laborious" should not suggest a correlation with effectiveness (task success). This style is effective in compensating (to some extent) for lack of spatial ability and avoiding a major decrease in performance and increase of perceived disorientation. But the style itself is not necessarily effective; it is not employed by highly effective users.

Component 4. Divergent Navigation. High scores on this component were associated with low homepage use and high average connected distance (long returns). This navigation style was explorative. Users were not that eager to re-visit pages but rather to explore new directions. This navigation style was only associated with a high propensity to trust (r=0.43).

These four components were labeled navigation styles and not navigation strategies; this decision was based on the way these components were derived – unsupervised learning, more specific, principal component analysis. There was no rational analysis involved on the experimenter's side, or assumptions of rationality or deliberation on the user's side, as in other work of this type (Catledge & Pitkow, 2001; Miller & Remington, 2001). Another reason to label them styles and not strategies was that these components were relatively neutral regarding task performance. We do not intend to suggest any correspondence between these navigation styles and the cognitive styles mentioned in the psychological literature.

An additional explanation is required for the difference between flimsy and laborious navigation styles. Their names and their representations (fig. 1 and 2) might suggest that they are opposite to each other or related in one way or another. However these two styles were derived as orthogonal (not correlated) components. They address different facets of disorientation: lack of skill (flimsy) and compensation (laborious). Important differences between the two styles are:

- A flimsy navigator uses the back button to stay around the home page whereas a laborious navigator uses the back button (and, in general, re-visitation) just as means to explore different options.
- A non-laborious navigator navigates linearly, do not re-visit, whereas a flimsy navigator re-visits quite a lot (especially the homepage).
- Flimsy is correlated with disorientation whereas laborious is not.

Thus, users employ laborious navigation when approaching a new and complex hyperspace to compensate for lack of ability.

Reading Time. Another second order-metric was constructed by a supervised learning approach, that is, by trying to get meaningful combinations of first-order metrics that would be significantly correlated with task outcomes. One such trial that proved to be successful is the following:

Equation 1:

$$readtime = 230.1 + 4.2 * viewlarg + 1.56 * viewref - 3.73 * viewsmal - 4.77 * viewindx + 2.04 * devview .$$

It was obtained by running a stepwise regression analysis with *task performance* as dependent variable and all the first-order metrics as independent variables. The significant predictors selected by the stepwise procedure (*viewlarg*, *viewref*, *viewsmal*, *viewindx*, and *devview*) and their corresponding b-weights together with the intercept (230.1) were used to build the equation 1 (see above). A unique score was calculated as a linear combination of the significant first-order metrics. It was labeled *reading time* since it positively weighted view time on large and reference pages presumed to be content pages (B=4.2 and B=1.56, respectively) and negatively weighed view time on small and index pages (B=-3.73 and B=-4.77, respectively). High *reading time* was significantly associated with low *flimsy navigation* (r=-0.5), low *divergent navigation* (r=-0.36), high *Internet expertise* (r=0.35), high *performance* (r=0,43), and low *disorientation* (r=-0.39). *Reading time* (high score) can be interpreted as a measure of reading efficiency during navigation: view time is higher on large and reference pages (since they require reading) and lower on small and index pages (since they only require scanning).

4.1.2.3. Semantic metrics

As argued in Section 3.1.1, semantic metrics are needed beside the syntactic ones, in order to more accurately characterize users' navigation behavior.

Labels. A first simple way to semantically characterize users' navigation behavior is to add semantic contents to a navigation graph. Color-coding can also be used to categorize this semantic information (fig. 3)

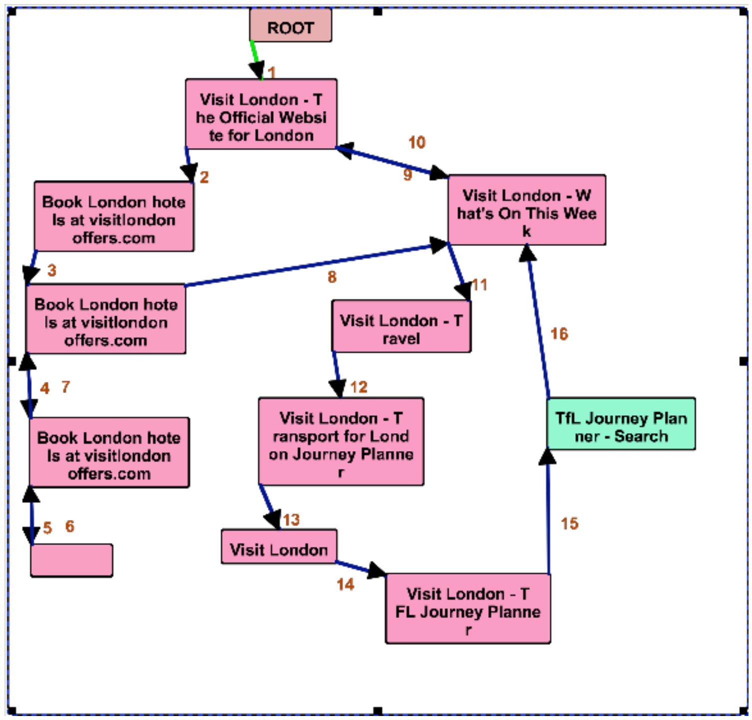


Figure 3. Navigation graph with semantic content (page titles) added to each node. Representing a page by its title is a convenient solution, though not always appropriate, as there can be different pages having the same title. Color-coding (or shading) can also be used to semantically categorize nodes (in this case the node containing the most goal relevant information has a different color (shade) than the rest of the nodes)

Objects that bear semantic information within a node are: the page title (as shown in fig. 3), headings and link descriptions, text paragraphs, or the page as a whole. Moreover, it is often the case that task relevant information is to be collected from multiple nodes. This is why it is important to define semantic objects that lie across multiple pages (e.g., navigation paths).

Semantic relevance. A semantic metric, called *Path adequacy*, was calculated based on navigation data and the task descriptions that subjects were provided with at the beginning of the navigation session (fig. 4).

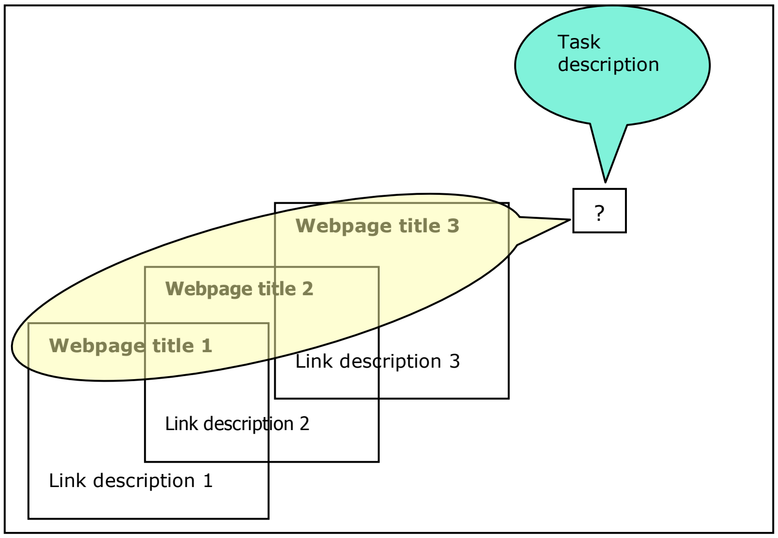


Figure 4. Path adequacy is calculated as a semantic similarity (LSA) between a task description and a navigation path composed of page titles.

A navigation path was considered to be a concatenation of semantic objects that the user has encountered in her/his way toward the current location. As semantic objects, one can consider link anchors, page titles, page contents, URLs, clickable icons, banners and images, etc. Here we present an example of navigation path composed of page titles:

> *If the user visited the pages titled 'Should I finance or pay cash for a vehicle? Calculators', 'How much will my vehicle payments be? Calculators', 'Glossary', and 'What vehicle can I afford? Calculators', then his/her navigation path was represented as a string of all words in these titles: <should, I, finance, or, pay, cash, for, a, vehicle, calculators, how, much, will, my, vehicle, payments, be, calculators, glossary, what, vehicle, can, I, afford, calculators>.*

Navigation paths were compared with task descriptions. The following is an example of task description:

> *Suppose you want to buy a car in 2 years. You have already saved $ 500. How much do you need to save on a monthly basis in order to make a down payment of $ 8000 for the car? Assume*

> *that the savings and tax rates are as listed. What is the most expensive car you can afford if you will be able to pay 40 monthly payments of at most $ 150 after the down payment?*

Path adequacy was determined as a coefficient of semantic similarity between a navigation path composed of page titles and a task description (fig. 4). Latent Semantic Analysis (LSA) was used to measure the semantic similarity between navigation paths and task descriptions. LSA was selected because is an automated measure of semantic similarity (see description of LSA in Section 3.2.4). Other measures, such as expert judgments, are not suited to be implemented in realistic Web applications. *Path adequacy* calculated at the end of a particular task was significantly correlated with *return rate* (r=-0.48), *spatial ability* (r=0.36), and *task performance* (r=0.47). These correlations showed that path adequacy was a relevant metric of user navigation behavior and encouraged us to consider path semantics as a first approximation in our modeling approach.

4.2. Individual differences in Web navigation – empirical findings[6]

In this section, user characteristics that have an influence on Web navigation behavior are presented. The relevance of several user characteristics in predicting task outcomes (performance, satisfaction, and reliability) was tested with the aid of multiple regression analysis. Navigation metrics were used as indicators of user characteristics and task outcomes. Results suggest that spatial-semantic cognitive mechanisms are crucial in adequately performing Web navigation tasks. The fact that user characteristics and task outcomes can be estimated with reasonable accuracy based on navigation metrics suggests the possibility of building adaptive navigation support in Web applications.

4.2.1. Research Objectives and Questions

The initial step was to identify factors that were able to predict the task outcomes performance, satisfaction and reliability. Some of these factors were person-related (user characteristics) and others were interface- and context- related. Subsequently, navigation data was used to estimate person-related factors (user characteristics) and predict task outcomes.

[6] This section includes passages from the paper „Individual Differences and Behavioral Metrics Involved in Modeling Web Navigation“ by Ion Juvina and Herre van Oostendorp, appeared 2006 in Universal Access in Information Society, Springer.

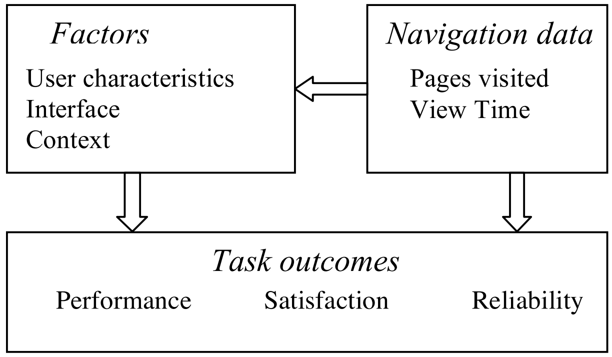


Figure 5. Overview of the issues considered in this study: hypothesized factors predict task outcomes (research questions 1,2, and 3), navigation data predict user characteristics (research question 4) and task outcomes (research question 5), and factors and navigation data predit task outcome (research question 6)

An overview of the issues that were considered in this study is presented in Figure 5. Research questions about these issues were formulated as follows:

1) Are the hypothesized factors indeed significant as predictors of task outcomes? Which ones are the best predictors?
2) What is the relative importance of each factor in predicting task outcomes?
3) How well can each of the task outcomes be predicted?
4) Is it possible to predict user characteristics based on navigation behavior? For example, how accurately *spatial ability* can be predicted based on navigation metrics?
5) Is it possible to predict task outcomes based on navigation behavior? For example, how accurately user's *perceived disorientation* can be predicted based on navigation metrics?
6) How accurately can task outcomes be predicted based on both user characteristics and navigation metrics?

4.2.2. Methodology

An exploratory task analysis allowed us to understand which were the most relevant success factors. For example:

- Some subjects were capable of deploying a fast, elaborate and effective Web navigation behavior. Consequently, a factor concerning Internet expertise was considered.
- Although tasks were conceived in such a way to require as little previous knowledge as possible, it was noticed that a certain familiarity of users with the application domain (personal finance in this study) was an advantage.

- Spatial ability was included in the hypothetical model based on the high frequency of spatial terms used in subjects' verbalizations, even when they were dealing with completely non-spatial issues. Examples of verbalizations with spatial connotation include: "where am I", "let's go in another place", "I'm stuck in these analyzers", "I saw it somewhere".

Based on this type of observations, variables and indicators were established to measure potentially relevant aspects. Criteria for task outcomes were specified during task analysis. They are presented in Section 4.1. Potential predictors included user characteristics, interface and context factors, and navigation metrics. User characteristics that were hypothesized (based on task analysis and previous research) to have an influence on task outcomes will be described in more detail in this section. Navigation metrics (syntactic and semantic) as described in Section 4.1.2 were also used as independent variables. Here the first-order syntactic metrics are summarized in table 3.

Given the scope of this study, only a few of the interface and context factors that could have an influence on task outcomes could be investigated. Others were randomized or kept constant. By choosing three different websites to be used as research material (www.financenter.com, www.thisismoney.co.uk, and www.amazon.com), factors pertaining to site structure or interface design were randomized. The task domain – Web assisted personal finance – was kept constant. Sites' usability was explicitly measured as an interface factor.

With regard to different usage contexts that could have an influence on task outcomes, we kept as constant as possible the room, the type of computer and all the other contextual factors that could influence users' navigation behavior, except the factor *Time constraints* that was experimentally manipulated by trying to induce the feeling of time pressure in a half of the participants. The other half received no specific instruction.

Table 3. First-order metrics

Metric	Short label	Description
Path length	*pathlength*	The number of pages visited during the task.
Relative amount of re-visits	*re-visits*	The probability that any URL visited is a repeat of a previous visit.
Return rate	*return*	The average number of times that a page was re-visited.
Back button usage	*backbutton*	The percentage of back button clicks among the navigation actions.
Relative amount of visits to homepage	*homepage*	Amount of visits to the homepage; "relative" refers to a correction based on path length.
Average view time	*meanview*	Average duration spent on viewing pages
Median view time	*medview*	The median of the view times spent on every page.
Difference *meanview* and *medview*	*difview*	The extent to which *meanview* is influenced by a minority of pages that are viewed carefully.
Deviation view time	*devview*	How much the view time varies between pages.
Time spent on large pages	*viewlarg*	Time spent on pages with number of words larger than average plus standard deviation.
Time spent on small pages	*viewsmall*	Time spent on pages with number of words smaller than average minus standard deviation.
Time spent on index pages	*viewindx*	Time spent on pages with a large number of outgoing links.
Time spent on reference pages	*viewref*	Time spent on pages with a small number of outgoing links.
Structural complexity (fan degree)	*fandeg*	The ratio between the number of links followed and the number of distinct pages visited.
Number of cycles	*cycles*	The difference between the number of pages and the number of links.
Path Density	*density*	To what extent users make use of all possible links of the site structure.
Average connected distance	*ACD*	The average length of the path between any two connected pages. A high ACD indicates that pages are not re-visited immediately but only after a number of other visits.

The following are descriptions of variables used in this study:

- An *Internet expertise* measure was constructed based on users' self reported frequency of Internet use and their self-assessed level of knowledge and skills in Web navigation.
- *Domain expertise* (Web-assisted personal finance) was measured with items such as: "Have you ever used a personal finance website (Yahoo Finance, MSN Money, etc.)?".

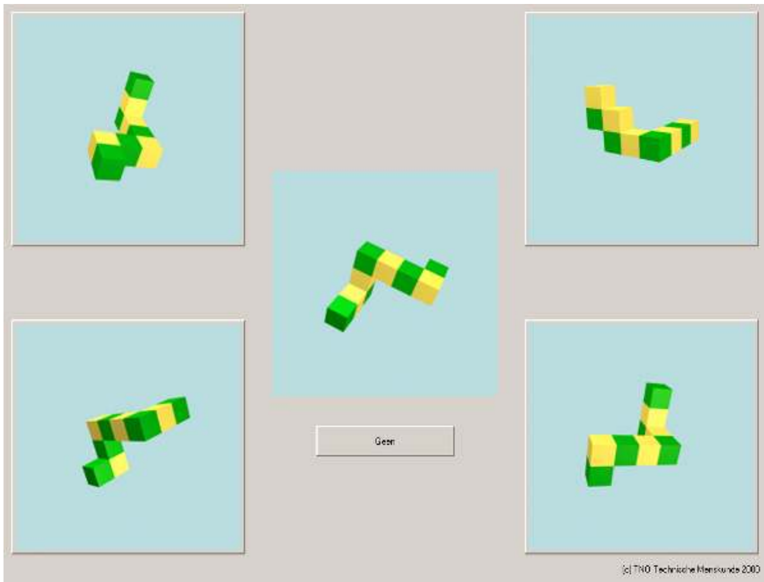


Figure 6. Snapshot from the spatial ability test. Participants had to indicate which of the four figures in the corners is identical with the figure in the center.

- The variables *spatial ability*, *episodic memory*, and *working memory* were measured with computerized cognitive tests provided by TNO – Human Factors Institute (Neerincx, Pemberton & Lindenberg, 1999). The 'Spatial ability test' used the classical mental rotation task, and the spatial ability score was the number of correct solutions obtained by rotating three-dimensional objects (correct matches between objects and their rotated equivalents) (fig. 6). The 'Episodic memory test' presented 3 lists of 60 images each; the participants had to loudly name the images in the first 2 lists; between lists 2 and 3 there was a distraction task (we used the 'spatial ability test' as a distraction task, to efficiently use the testing time); list 3 contained images that were presented before in lists 1 and 2 together with new images; the participants had to recognize the images that were presented in list 1. The 'Working memory test' used a reading span task (Linderholm, 2002): subjects were presented with series of phrases, the size of series increasing progressively from 2 to 7 phrases; the participants were asked to loudly read the phrases and try to understand their content; after each series, the participants were asked to recall the last word of each phrase in that particular series; for one random phrase in the series participants were asked to fill in 2 missing words, to ensure that they really treated the whole content and not only the last words. The working memory score was calculated based on correctness of recalls. 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 individual's belief regarding the causes of his or her experiences, and those factors to which an individual attributes his or her successes and failures. 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). Locus of control was measured with a 20-item scale (Pettijohn, 2002). The *sequential-holistic* cognitive style was measured with items such as: "I like to break down large problems into smaller steps" and "I like to look at the big picture" (Peak Performance, 2002).
- 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 current disposition. Subsequently, users' ratings were factor analyzed and grouped in three basic *moods*. Thus, *active mood* was composed by the following affective states: Determined, Calm, Alert/vigilant, Sluggish/lethargic/lazy (negative sign), and Blue/Depressed (negative sign); *Enthusiastic mood* was mainly composed of the Enthusiastic, Excited, and Strong states; and *irritable mood* contained mainly the states Irritable, Sluggish (lethargic, lazy), Nervous, Sleepy, and Relaxed (negative sign).
- Participants' *propensity toward trust* (Egger, 2003) was measured with items such as: "People always can be trusted" and "People always take care only of themselves".
- A factor called 'Motivation' was included based on observations during the experiment and inspections of students' answers to questionnaires items. A dichotomous variable that differentiates between participants from Utrecht University and Twente University was initially recorded just to check for sampling errors. Afterwards, it was noticed that the students in the two universities reported consistently different types of interests, that

is, students from Utrecht University declared higher levels of interests in entertainment and personal development, whereas students from Twente University declared higher levels of interests in personal and professional businesses. This variable was hypothesized to pertain to students' motivation and goal orientation. The differences between the two groups of students (Utrecht vs. Twente) seemed similar to the difference between mastery and performance goal orientation. *Mastery* oriented students perceive new tasks as an opportunity to learn or to acquire new skills, whereas *performance* oriented students perceive tasks as opportunities to demonstrate already existing competence and skills (Miron, 2003). This hypothesis must be checked in further research, but, for this study, the new dichotomous variable was used with the general label "Motivation".

- *Self-efficacy* was measured with a questionnaire, adapted from (Compeau, 1995), containing items such as "I could perform better using these websites if I had a lot of time to complete the job for which the sites were provided".
- An *interests* factor was operationalized based on principal component analysis. Participants were asked to indicate for what purposes they use the Internet. Answers were factor analyzed, and the 2 components that resulted were called 'Interest entertainment and personal development' (for brevity, *interest entertainment*) and 'Interest personal and professional business' (for brevity, *interest business*), respectively.
- *Perceived usability* of the three sites used in this study was measured with a selection of items from two known usability questionnaires (Sumi; Wammi), consisting of items such as: "It was easy to use this website" and "I could effectively complete my tasks using this website".
- The factor *time constraints* was experimentally manipulated. Half of the subjects (15) were instructed that only 30 minutes are available to complete the navigation tasks, while the other half did not receive any time indication. In fact, all subjects were given a maximum of 40 minutes to execute the navigation tasks. No clock or other time indication was available.

Participants

The study was run with 30 participants in a single session, lasting approximately 2½ hours. 15 participants (7 females and 8 males) were registered as students in the Information Management Department of Twente University, and the other 15 participants (8 females and 7 males) were students in the Information Science Department of Utrecht University. Participants were randomly selected out of students' catalogues of both universities. Participation was not compulsory.

Students who declined the invitation were replaced by making new random selections in the pool of candidates. Half of the participants were randomly assigned to the 'Time constraints' condition in which the participants were instructed to finish the navigation tasks in 30 minutes.

Procedure

The first part of the sessions was dedicated to questionnaires and cognitive tests aimed at measuring user characteristics. The second part consisted in execution of Web navigation tasks. This part lasted maximum 40 minutes for all participants (including those in the 'time constraints' condition). No clock was available, participants were asked to put away their wristwatches, and the computer clock was disabled. During the navigation task, navigation behavior and task performance were recorded. Subjects were informed that their navigation behavior was recorded. Task performance was recorded by the participants on a dedicated form and coded afterwards by the experimenter. The third part of the sessions consisted of administration of usability and satisfaction questionnaires. Each participant received a compensation of Euro 20 at the end of the session.

4.2.3. Results

Results are presented in the same order as suggested in the overview of the issues considered in this study (fig. 5). Multiple linear regression analysis was used to investigate the significance of Hypothetical factors in predicting task outcomes, as well as the possibility of using navigation metrics as estimates of user characteristics and predictors of task outcomes. Including predictors in regression models was based on the *stepwise method* (Tabachnick & Fidell, 2001), thus the results must be seen as the best one can get with the minimum number of predictors. The input of regression analysis is composed of predictors and criteria described above. After each analysis, the outcomes of the stepwise procedure were summarized in tables presenting:

- the criterion (dependent variable);
- the proportion of variance in criterion explained by the significant predictors (R square);
- the predictors retained at the end of the stepwise method as significant;
- the relative importance (beta coefficient) of each significant predictor.

Predictors found to be not significant (excluded from the model) are not listed with every analysis but they can be easily recovered from the list of predictors presented above.

4.2.3.1. Predicting Task Outcomes Based on Hypothesized Factors

All task outcomes could be predicted based on a limited number of predictors with various effect sizes. The effect size for regression was calculated with the following formula $ES^2 = R^2/(1-R^2)$. An effect size of 0.02 is considered a small effect, 0.15 a medium one, and 0.35 a large one (Cohen, 1992). In our case (table 4), the smallest multiple R squared (0.22) corresponds to a medium-large effect size.

Table 4. Predictions of task outcomes based on Hypothetical factors. Perceived disorientation and frustration are aspects of the reliability criterion.

Task outcome	R square	Predictors	Beta
Performance	0,39	Spatial ability	0,496
		Finance expertise	0,385
Satisfaction	0,67	Motivation	0,612
		Usability	0,506
		Interest business	-0,319
Perceived Disorientation	0,42	Usability	-0,505
		Working Memory	-0,344
Frustration	0,22	Time constraints	0,471

Task performance was best predicted by *spatial ability* and *domain (finance) expertise*. In other words, the user ability to represent the information space structure and their domain knowledge were the most important determinants of task success. *Satisfaction* was best predicted by *motivation*, *usability* and *interest in business*. Users who were motivated, perceiving the websites as usable and not interested in business were more likely to be satisfied with task completion. *Business interest* was negatively correlated with *satisfaction* (r=-0.38). A possible interpretation is that subjects with interests in personal and professional business have higher expectations and they are more vulnerable to be dissatisfied when task execution and results do not meet their expectations. Concerning *Reliability*: *Disorientation* was best predicted by *usability* and *working memory*. Low working memory capacity and low perceived usability were associated with increased probability of users' perceived disorientation. *Frustration* was predicted by *time* *constraints*. Users in the 'time constraints' condition reported a higher level of frustration than users in the control condition.

4.2.3.2. Predicting User Characteristics Based on Navigation Metrics

As presented in Section 4.1, several types of navigation metrics were calculated: *first-order*, *second-order* (navigation styles and reading time) and a semantic metric called *path adequacy*.

Table 5 shows that a considerable number of user characteristics could be predicted with reasonable accuracy based on navigation metrics. For example, 25% of the variance in *Internet expertise* was predicted based on the *flimsy navigation style*. Spatial ability can be estimated based on the number of re-visits.

Table 5. Predictions of user characteristics based on navigation metrics

User characteristic	R square	Predictors	Beta
Internet expertise	0,25	Flimsy navigation	-0,50
Spatial ability	0,195	Re-visits	-0,442
Episodic memory	0,245	Laborious navigation	0,495
Working memory	0,28	Flimsy navigation	-0,83
		Median view time	0,58
Locus of control	0,398	Flimsy navigation	-0,704
		View reference	0,874
		Deviation view	-0,553
Active mood	0,258	Path density	-0,508
Trust propensity	0,233	Average connected distance	0,483
Interest entertainment	0,147	Re-visits	-0,383

When first- and second-order metrics occur together as predictors of a particular user characteristic, they are statistically independent.

4.2.3.3. Predicting Task Outcomes Based on Navigation Metrics

Table 6. Predictions of task outcomes based on navigation metrics

Task outcome	R square	Predictors	Beta
Performance	0,299	Path adequacy	0,385
		Reading time	0,298
Perceived Disorientation	0,213	Flimsy navigation	0,462

Two of the task outcomes, *performance* and *disorientation*, were significantly predicted based on second-order (syntactic) and semantic metrics (table 6). These results showed that second-order and semantic navigation metrics add value to first-order metrics. There were no significant predictions for *satisfaction* and *frustration*.

4.2.3.4. Predicting Task Outcomes Based on Hypothetical Factors and Navigation Metrics

When both Hypothetical factors and navigation metrics (see Figure 5) were entered in the regression analysis, the selection of the most significant predictors was consistent with the previous analyses (table 7). Small differences can be noticed but they are not surprising. For example, *number of cycles* appeared among the significant predictors of disorientation instead of *flimsy navigation*. But *cycles* was highly correlated with *flimsy navigation,* actually it was one of its most important components. Two first-order metrics appeared as significant predictors of *satisfaction,* which can be interpreted as follows: the amount of variance in *satisfaction* left unexplained by *motivation*, *usability* and *interest for business* could be explained by *view time per large pages* and *view time per small pages*. Note that the *beta* coefficients had opposite signs: users were more satisfied when they spent relatively long time on large pages and relatively short time on small pages than vice versa. In other words, spending relatively long time on large pages (presumed to be content pages) determined that part of *satisfaction* in task completion that was not determined by *motivation*, *interest for business* and *sites usability*.

Table 7. Predictions of task outcomes based on Hypothetical factors and navigation metrics

Task outcome	R square	Predictors	Beta
Performance	0,515	Spatial ability	0,399
		Finance expertise	0,340
		Path adequacy	0,216
		Reading time	0,318
Satisfaction	0,793	Motivation	0,754
		Usability	0,619
		Interest business	-0,277
		View large	0,491
		View small	-0,302
Perceived Disorientation	0,57	Usability	-0,496
		Cycles	-0,388
		Working Memory	-0,305

Another effect of considering factors and metrics together as predictors of task outcomes was a considerable increase in predictive power: R square was higher than 0.50 for all criteria (the average increase of R square is 0.133) except *frustration* (see table 4).

4.2.4. Conclusion and discussion

4.2.4.1. Summary of results and conclusions

This study has shown that user characteristics such as *domain expertise*, *spatial ability*, *working memory*, *motivation*, and *interest* are important determinants of task outcomes. Interface and context factors such as *sites' usability* and *time constraints* have also an influence on some of the task outcomes (see research questions 1, 2, and 3 in Section 4.2.1).

However, user characteristics as determinants of task outcomes can only be measured in experimental settings. The study has also shown that some of the user characteristics such as *Internet expertise*, *spatial ability*, *working memory*, *episodic memory*, *trust propensity*, and *interests* can be estimated with a reasonable level of accuracy based on Web-logging data that can unobtrusively be collected in a real-world navigation session (table 5) (research question 4).

The predictions of task outcomes based on user characteristics, interface and context factors appeared to be more accurate than those based on navigation metrics. This difference suggests that there is still enough work to be done in searching for accurate and relevant indicators of navigation behavior (research question 5). However, both categories of predictors are important, one from a more theoretical perspective and the other from an applied one (research question 6).

A considerable number of factors proved to be less relevant than expected or reported in literature. For example, the demographic factors gender and age were not correlated with any of the task outcome or navigation metrics. This might be due to the structure of our sample. Neerincx, Lindenberg, Rypkema, and Van Besouw, (2000) found important effects of age on Web navigation performance. The affective and conative factors were not as important as the cognitive factors.

4.2.4.2. Discussion and implications

Studying a large number of factors in relation to a comprehensive range of outcomes of Web navigation tasks in a particular domain (Web-assisted Personal Finance) was useful in several respects. A limited number of significant predictors were identified, and their relative contribution to the accuracy of predictions was estimated. Since factors were studied together and the stepwise method of regression analysis was employed, it was possible to rule out factors that were only marginally significant or confounded with one another. This is an important contribution of this study in comparison with other work of this type. Most of the studies addressing individual differences in Web navigation (including those referenced here) are restricted to a limited number of user characteristics, and for this reason they can easily overlook other (more important) characteristics. For example, the influence of working memory on hypertext navigation as reported by Tucker and Warr (1996) might have not appeared as significant if spatial ability was included as a predictor in their model (Tucker & Warr, 1996). Our results show that *spatial ability* is more important for Web navigation performance than *working memory capacity*.

The distinction between *spatial ability* and *working memory capacity* is not clear-cut from a theoretical point of view. As a matter of fact, one component of working memory – visuo-spatial sketchpad – is believed to be involved in storing and manipulation complex spatial patterns (Baddeley, 2000). This component is most probably involved in *mental rotation* – the probe we have used to measure individual differences in spatial ability. To assess individual differences in working memory we have used the *reading span* test, which addresses another component of working memory – the articulatory loop. Thus we have taken into consideration both spatial and non-spatial aspects of working memory and it turned out that the spatial ones are relatively more important for Web navigation than the non-spatial ones.

The proposed approach to calculating different types of metrics based on navigation data proved to be profitable. Different types of knowledge about user can be inferred based on the kind of information that is extracted from this data: *syntactic* (structural) information indicated mainly users' navigation styles, for example, if they rather re-visit pages than viewing new pages, if they return using the back button or just by following links, etc. (first- and second-order metrics); and *semantic* information indicated if users were effective in pursuing their goals (path adequacy) independent of their navigation styles.

From a theoretical perspective, it appears that spatial-semantic cognitive mechanisms are crucial in adequately performing Web navigation tasks. This study has identified some individual differences that are consistently associated with specific task outcomes. Next sections are concerned with modeling cognitive mechanisms that are responsible for these individual differences and for their influence on task outcomes. Presently, the existing cognitive models of Web navigation (Kitajima, Blackmon et al., 2000; Pirolli & Fu, 2003) ignore almost completely the spatial dimension, and treat solely the semantic dimension of Web navigation (information scent).

The results of this study have important practical implications. A Web application can be designed in such a way that it takes into consideration (or compensates for) those factors that proved to be significant in predicting task outcomes. For example, since *spatial ability* is one of the determinants of *task performance*, some interface features (e.g., maps) should be designed to compensate for low spatial ability. The indicators of navigation behavior that are automatically calculated during a navigation session and are able to predict relevant user characteristics and also task outcomes can be used to model the user in real time and personalize the application. For example, *Internet expertise* and *spatial ability* are virtually impossible to measure in real-time use of a Web application, but they can be estimated based on user's navigation behaviour. The application can be programmed to provide additional navigation aid when users are diagnosed "at risk" and to hide useless hints when users are assessed as "doing well".

4.2.5. Additional analyses based on data from other studies

The fact that spatial ability was the most important determinant of Web navigation performance was an unexpected and non-intuitive result. Why would performance on a semantically void mental rotation task predict performance on a semantically intensive Web navigation task? Before trying to answer this question we wanted to make sure the correlation between spatial ability and Web task performance is a robust result. The following analyses and results show that this is indeed the case.

In 2 other subsequent studies reported in Section 4.4 and 5.1 respectively, this correlation was replicated (table 8).

Table 8. Correlations between spatial ability and Web task performance in three studies

Study	Correlation coefficient (r)	Significance (p)	Number of participants
Individual differences	0.494	0.005	30
Voice suggestions	0.682	0.005	15 (control condition)
VIP	0.564	0.023	16 (control condition)

Possible confounders for this correlation that have been checked for and proven to have no influence are: working memory capacity, episodic memory, Internet expertise, reading comprehension, reading speed, and cognitive style. The correlation between spatial ability and another type of computer task performance has been checked and found non-significant. Thus, the correlation between spatial ability and Web task performance is a robust result.

A look at the correlations between spatial ability and navigation metrics might help understanding the relationship between spatial ability and Web task performance. Spatial ability is negatively correlated with metrics involving re-visitation (table 9).

Table 9. Spatial ability is negatively correlated with metrics involving re-visitation

Variables	Correlation coefficient (r)	Significance (p)	Number of participants
Spatial ability – re-visits	-0.442	0.014	30
Spatial ability – back button	-0.426	0.019	30
Spatial ability – fan degree	-0.427	0.019	30

It seems that spatial ability helps users in figuring out the information space structure, so less re-visitation is needed. Supposedly common cognitive processes are used to represent and operate on an information space and to mentally rotate objects in a three-dimensional space (see fig. 7 for a possible analogy).

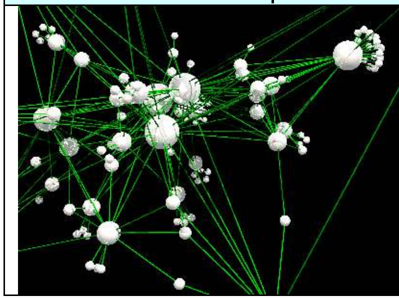
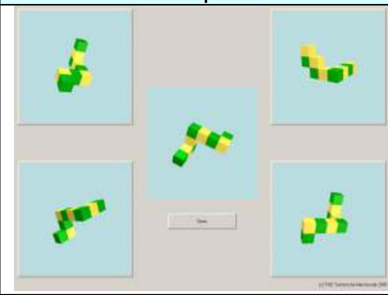
Information space	3D space
	

Figure 7. Mentally represented information space (left) and physical space (right). Supposedly common cognitive processes are involved in operating on these representations, explaining the correlation between spatial ability and Web task performance.

4.3. Modeling cognitive processes involved in Web navigation

A subsequent objective was to find explanatory mechanisms for the relevant factors identified by task analysis and correlational research, as presented in previous sections. Explanatory mechanisms are useful from a theoretical perspective but also from a practical one – more valid, more general and more fine-grained design recommendations can be provided.

Complementariness between spatial and semantic aspects involved in Web navigation (Chen, 2000; Tamborello & Byrne, 2005) could explain the relationships between navigation data, user characteristics (e.g., spatial ability) and task outcomes found in our research. While navigating, users probably represent the information space and select appropriate steps based on assessed relevance of screen objects to their goals ("information scent"). Spatial ability is involved in representing the information space and operating on it while users' domain knowledge is involved in judging relevance.

4.3.1. Cognitive model

A cognitive model integrates analyses of cognitive processes that account for the observed empirical effects. This ensures coherence with existing theories and research and helps in validation of particular findings.

Our proposed model of Web navigation attempts to implement the complementarities between syntactic (spatial) and semantic processes that have been found empirically. More specifically, we try to model the role of contextual information in selecting specific navigation actions. "Context" could refer to information on the current screen that surrounds a particular object (Brumby, 2004) or to information encountered in earlier steps of the navigation session. We focused on the latter since it is well documented in literature (Chen, 2000; Howes, Payne et al., 2002; Wen, 2003) while it has not been addressed yet from a cognitive modeling and computational perspective.

The account our model gives for spatial cognition is indirect; it relies on a number of assumptions. We assumed that the user's navigation path is a good approximation of the user's mental representation of the information space being traversed. We also assumed that the exact nature of this representation, in other words how a navigation graph is mentally represented, is less relevant from a computational perspective. Syntactic aspects of the path (what have been called syntactic metrics in previous sections) are implicitly considered. For example, a path with a high *Average Connected Distance* (ACD) is more informative and less redundant than a path with a high *Density* (D)[7]. What matters for the user's current decision is how much relevant information the path conveys.

4.3.1.1. CoLiDeS+

We have made a few amendments to an existing cognitive model of Web navigation, namely CoLiDeS (Kitajima, Blackmon et al., 2000; see Section 2.3.4.2). The altered model has been labeled CoLiDeS+, to indicate that it shares the main assumptions with the original and is intended to eventually be an augmented model.

CoLiDeS+ brings in the concept of "path adequacy" as a complement to the concept of "information scent". As shown is Section 4.1.2.3, path adequacy is the semantic similarity between a navigation path and a user's goal. A navigation path is a succession of links that have been selected prior to a particular moment in a navigation session. Users are assumed to base their selections based not only on goal-relevance of incoming information but also on whether a candidate selection is consistent with past selections or not. Therefore, in CoLiDeS+ selecting a link on a specific webpage is a function of goal description, link description and path description[8].

---

[7] High ACD means delayed re-visitation, that is, a high number of different visits are made between a visit and re-visit to a page, thus the resulting path contains non-redundant information. High D means revists are made whenever possible, thus the resulting path contains information about a limited number of pages.

[8] Here the navigation path is composed of link labels instead of page titles as in the previous section and path adequacy is calculated at each step of a navigation session.

CoLiDeS models mainly the ideal situation of forward linear navigation; backtracking steps are considered erratic actions. When no particular object on the current page sufficiently matches the user's goal, an impasse is said to have occurred. Solutions to impasses are only described and not computationally modeled by Kitajima, Blackmon and Polson (2000). However, backtracking and impasses seem to be natural in Web navigation and rather frequent (Cockburn & McKenzie, 2001; Wen, 2003). Therefore they need to be modeled within the same framework as forward linear navigation. Miller and Remington (2004) propose navigation strategies to deal with ambiguity of link labels or with users' errors in judging link relevance (Miller & Remington, 2004).

CoLiDeS+ also tries to incorporate navigation strategies by maintaining a developing representation of the information space being navigated (navigation path) and checking at each step for impasses based on path adequacy. An impasse occurs when path adequacy does not increase after selecting a link and it is a reason to switch the path. At this point, CoLiDeS+ reacts with a strategy that we called "next best" and it is to some extent similar with the opportunistic strategy of Miller and Remington (2004). "Next best" means that not only the link with the highest similarity to the user's goal is considered but also links with lower similarities provided that they contribute to an increase in path adequacy. And eventually the options of backtracking one or more pages or going to index pages are available.

A short description of the algorithm used by CoLiDeS+, presented below, shows how the concept of path adequacy is considered in addition to link relevance (see also Figure 8):

- A task description is taken as input and assumed to be equivalent to the user's goal.
- A webpage is attended to, parsed in several areas, and a particular area is focused on (e.g., a menu).
- Menu entries are comprehended (based on how semantically similar to the user's goal they are) and one entry (the one that is most relevant to the user's goal) is selected and acted on (e.g., clicked on).
- A new page is attended to and if the target information cannot be found, the cycle is reinitialized.
- The selected element is retained in a memory structure that maintains user's navigation paths.
- Starting with the second cycle, a navigation path is available and the metric called *path adequacy* is computed. Selections of screen objects to be acted on are made if they contribute to an increase in path adequacy.

- Otherwise an impasse is declared and dealt with by considering "next best" options, changing the focused-on area and backtracking.
- The algorithm stops when the user decides that the current page contains the target information.

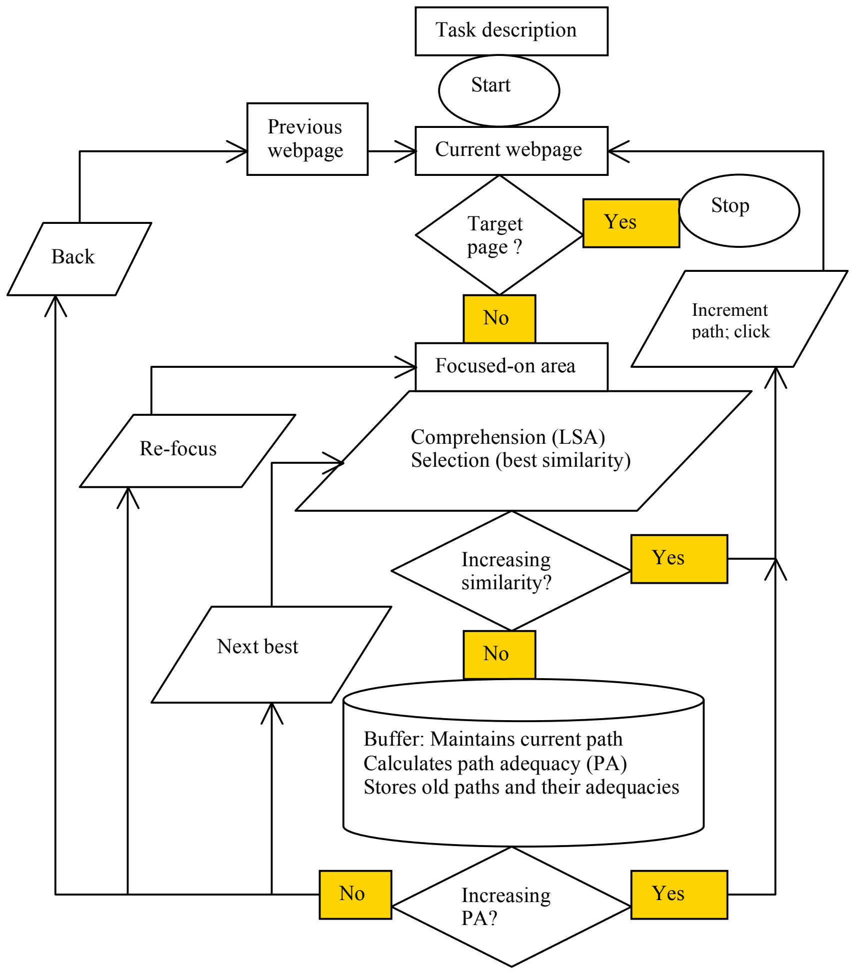


Figure 8. A diagram of the algorithm that implements the CoLiDeS+ model

The following is an example of how the algorithm of CoLiDeS+ works:

- Let's assume the user has the following goal: "I want to find a hotel as soon as possible and as cheap as possible".
- The current page has two options: "sleep" and "eat". Their goal-relevancies (the LSA semantic similarity coefficients[9]) are .32 and

[9] The "general reading" semantic space available at http://lsa.colorado.edu was used.

.27, respectively. Therefore, the link labeled "sleep" is selected and clicked on.

- The next page has two options: "sign up for a waiting list" and "book accommodation now". Their goal-relevancies are .73 and .44, respectively.
- The goal-relevance of "sign up for a waiting list" is increasing in value as compared with the one of "sleep", so this link is selected and clicked on.
- The next page has two options: "registered customer" and "new customer". Their goal-relevancies are .10 and .25 respectively. Since the highest one is not increasing in value as compared with the previous selection, path adequacy is calculated.
- The path consisting of past selections "sleep sign up for a waiting list" has an adequacy of .72. When the link label "new customer" is added, the adequacy of the new path "sleep sign up for a waiting list new customer" decreases to the value of .61. Since the path adequacy does not increase, a path switch is necessary.
- After one step back and a 'next best' selection, the link "book accommodation now" is clicked on.
- The next page has now the options "economical rooms" and "exclusive apartments". Their goal-relevancies are .36 and .24, respectively. The higher of these two relevancies (.36) is lower in value than the previous selection "book accommodation now" (.44). However, its corresponding path adequacy ("sleep book accommodation now economical rooms" = .55) is increasing as compared with the adequacy of the old path ("sleep book accommodation now" = .51). Therefore, the link "economical rooms" is selected and clicked on.

Based on its conditions of increasing relevancies and adequacies as well as its mechanisms of recovering from impasses, the algorithm was able to find the right path. Without these conditions, based only on the rule of selecting the most relevant link on each page (as the original CoLiDeS proposes), a wrong path would have been selected – "sleep sign up for a waiting list new customer".

Based on this type of simulation, it is possible to determine at each step in the simulation process what is the model's successful path up to that moment and what are the model's unsuccessful trials (detours from the successful path). In the above example, at the end of the simulation, the model's successful path is "sleep book accommodation now economical rooms" and the model's unsuccessful trial is "sign up for a waiting list".

This algorithm is not entirely automated as an executable computer program. Some important aspects are not computationally modeled. As other cognitive models, CoLiDeS+ makes use of assumptions for those processes that are abstracted out. One important assumption is that the user is familiar with Web use and knows how and when to use various interface widgets (both browser and websites interfaces). The processes of parsing a webpage and focusing on a screen area are not modeled. The experimenter just provided the model with the right focus-on area. An implementation of CoLiDeS+ as a computational cognitive model is presented in Chapter 5.

Cognitive grounds of CoLiDeS+

CoLiDeS+ is a process model of Web navigation, which describes how information presented on the screen is processed in order to perform various types of Web tasks. It is originated in established theories and models of text comprehension.

The Construction - Integration theory of text comprehension (Kintsch, 1998) postulates a construction phase in which a mental representation is constructed from textual input, reader's goals and prior knowledge, and an integration phase which establishes coherence of the constructed representation. Construction is local (context-free) whereas integration is global (context-dependent). Human comprehension is regarded as a top-down and bottom-up process (Kintsch, 2005). CoLiDeS (the original source of CoLiDeS+) only implements a local top-down feature: assessing an incoming text element in the view of user's goal. By adding contextual information (global bottom-up feature) we made the model more consistent with its theoretical assumptions. Contextual information is essential in comprehension of particular text elements especially for text elements with equivocal or metaphorical meaning. Lack of supportive sentence context may lead to fast reading but poor comprehension (Budiu & Anderson, 2004). In Web navigation, contextual information allows users to build a representation of the information space that is being navigated and this representation, in turn, supports locating and integrating relevant pieces of information.

Link labels have various degrees of ambiguity (due to either bad design or user's comprehension limits). Users are generally able to disambiguate an ambiguous term by integrating it in context (Kintsch, 1998). Budiu and Anderson (2004) demonstrated from a computational modeling perspective that when a word seems inappropriate, a rich sentence context can help people grasp the intended meaning of that word.

As the reader proceeds through a text, he/she constructs an episodic memory representation of the incoming information and uses background knowledge from semantic memory (van den Broek, Rapp et al., 2005). Since human attentional resources are limited, only a small part of the reader's memory is active at a given moment, that is, only a small amount of knowledge resources can be employed in current processing. What determines which concepts are activated? According to the authors of *the Landscape model of reading* (Van den Broek, Young et al., 1999) there are several sources of activation: the text element that is currently being processed, the preceding recently processed knowledge, the knowledge processed in earlier phases of a reading session, and the reader's background knowledge. Since the Web navigation tasks that we modeled are goal-directed, we consider the user's goal as the primary source of activation. Thus, in CoLiDeS+, information can be actively involved in current processing because: (1) it is the current goal, (2) it is attended to (incoming text elements), (3) it has some residual activation carried over from previous cycles of processing (previously encountered text elements), (4) it is required for understanding of current information, and (5) it is strongly associated with an already active knowledge element.

Existing models of cognition (e.g., Landscape) assume that concepts can be activated to different degrees. CoLiDeS+ gives priority ranks to various sources of activation and this results in some sources of activation having a bigger influence than others. For instance, user's goal is more important as a source of activation than information previously encountered. The attended text element is first assessed in the view of user's goal. Only when the result of this assessment is not straight forward previously attended text elements come into play to disambiguate the relationship between the current text element and the user's goal. This ranking mechanism can also give a simple account for decay effects. Thus, background knowledge and previously encountered information is likely to have less influence than the currently attended information on the current processing. Our particular ranking of various sources of activation is tailored to the type of tasks we are simulating. In a Web based information search task, already encountered information has been assessed as partly useful or useless. We only use selected text elements, so they were judged as having some relevance, but the most relevant elements are not there yet, and are expected to be found in the incoming information. This is why incoming information has received a higher rank as an activation source.

As mentioned in the text comprehension literature, readers often import concepts that are not mentioned in the current sentence. Employing background knowledge from user's long term memory in the current processing happens either because there is an explicit need for that knowledge or the prior knowledge is strongly associated with the current knowledge (Van den Broek, Rapp & Kendeou, 2005). We have shown in Section 4.2 that domain expertise (prior knowledge and experience) plays an important role in successfully performing Web navigation tasks. CoLiDeS+ (as well as CoLiDeS) uses a LSA semantic space to model user's knowledge representation. A particular semantic space (corresponding to a particular user population) is used to represent concepts and relationships between them (semantic similarities). Based on this representation, a text element can be "comprehended", that is, it is possible to specify the concepts of the semantic space that are most associated with it.

An LSA semantic space is not a perfect model of user's background knowledge but it is useful in building computational models (Kintsch, 1998). LSA allows an objective estimation of the strength of association between knowledge elements; this estimation is essential in calculating the amount of activation that is spread between various knowledge elements. CoLiDeS+ relies crucially on a knowledge representation that allows comprehension of incoming information. An attended text element (e.g., a link label) is represented in the semantic space and it can be computed if the current link label is semantically connected to the current goal. If the connection is not strong enough, the processor employs contextual information, that is, text elements that have been previously attended to and selected (navigation path). The semantic space allows determining if the new text element is connected to the previous information that has been selected. If this is the case the current text element receives an extra activation from the path elements and contributes itself to the *cohort activation* (Van den Broek, Rapp & Kendeou, 2005) of the whole new path (including the new element). The new path as a whole can now be stronger related to the goal (*path adequacy* is higher) than the old path. If this is the case, the new element is integrated in the path and (if the goal is not fully attained) the user proceeds further to attend to new information. If the added element makes the path less connected to the goal, it is rejected, and other options are considered. The semantic space provides a reference frame to represent all of these connections and to estimate their strength.

4.4. Study "Voice suggestions"[10]

This study was conducted in order to experiment with CoLiDeS+, check its validity and practical relevance. It was hypothesized that CoLiDeS+ would be able to simulate real user's navigation behavior and the navigation support generated based on simulations of successful paths

[10] This study has been conducted in cooperation with Poyan Karbor and Brian Pauw, master's students at the Institute of Information and Computing Sciences, Utrecht University.

would have a positive influence on user's navigation behavior and task performance. This positive influence was expected to be bigger for users with a deficit of spatial ability, since CoLiDeS+ took over the job of representing the information structure and remembering past selections.

4.4.1. Method

Participants and design.
Participants (students, with sufficient Web experience, non-domain and non-internet experts) were randomly assigned to two conditions: a *control* condition in which 15 participants had to perform as many of the six tasks as possible in 45 minutes, and a *support* condition in which 14 participants did the same while receiving the generated navigation support (suggestions). These participants were instructed in advance that suggestions were generated by a robot, they were meant to help with task execution, and they were not mandatory: participants could follow them or not at their discretion. Suggestions were provided in the auditory modality. This way of delivering navigation support was selected since it can be implemented in combination with screen readers in order to improve the Web access of visually impaired users.

Tasks.
Six realistic Web tasks were constructed based on the collection of cases gathered by Morrison, Pirolli and Card (2000) and using experience of Kitajima, Blackmon and Polson (2000) concerning the size and elaboration of task descriptions. Each task had an associated website (see table 1 for an example of a task).

The six tasks were simulated with CoLiDeS+ prior to the actual navigation sessions. The results of simulations were successful paths, that is, successions of links leading to the target pages, and 'dead-ends', that is, pages that are not linked with the target pages, making it necessary for the user to backtrack. Based on these results of the simulations, two types of suggestions were generated: *link* suggestions – when a link contained in a successful path was visible on the screen, the user received the suggestion *Click on* <link label>; and *path switch* suggestions – when a 'dead-end' page was downloaded, the user got the suggestion *Go back*.

Task performance.
Solutions to tasks were reported on paper and evaluated afterwards for correctness. The answers were scored for correctness and completeness on a 4-point scale for each task (0 – not attempted task, 1 – erroneous answer, 2 – partly correct answer, 3 – correct answer, 4 – correct and

complete answer). The total correctness score for the six tasks ranges from 0 to 24. Calculating a general correctness score across the 6 tasks was justified by reasonably good consistency (Cronbach's alpha = .63) and by the need to build a well-formed variable (continuous and normally distributed). The average duration of tasks per participant was calculated by dividing the total navigation time (45 minutes) to the number of tasks attempted. An overall estimate of *task performance* was calculated by dividing the total correctness score to the average duration of tasks. The natural logarithm of this ratio was taken to correct for a skewed distribution.

Participants' *spatial ability* was tested with a mental rotation task, the same task described in Section 4.2. Navigation actions of participants were automatically recorded with Web-logging software (Scone). Navigation sessions were recorded on video using the software tools "The Observer" and "Camtasia Studio".

4.4.2. Results

The first outcome was that CoLiDeS+ was able to generate successful paths and 'dead ends', although the way it navigated the websites was not as similar to real users as suggested by Kitajima, Blackmon and Polson (2000). It made extensive use of 'next best' trials, refocusing, and backtracking. The number of steps to solutions was higher than the actual users took. However, even when the model took different decisions than the actual users, the correct paths and dead-ends were correctly identified, due to the mechanisms of solving impasses. This result justified using the model's outcomes (successful path and dead-ends) in generating navigation support.

4.4.2.1. How much "plus" does CoLiDeS+ bring in?

By making use of path adequacy, CoLiDeS+ simulated users' behavior slightly better than CoLiDeS. To show this, we have analyzed 10 user sessions recorded on video randomly selected from the control condition[11]. Only data from the control condition was used since the support condition was already manipulated based on CoLiDeS+ simulations. We looked at what selections users made and compared with what selections CoLiDeS and CoLiDeS+ would have made based on the same set of options.

---

[11] This analysis has been conducted in cooperation with Poyan Karbor, master's student at the Institute of Information and Computing Sciences, Utrecht University.

First, for each user, for each task, for each visited webpage, we have recorded the (last) area of the page the user has focused on, options available in this area, and which option was selected by the user. This latter information was recorded as the *user selection*. Second, the task description and the same options were entered in the LSA algorithm to compute semantic similarities between the task description and each option. The option with the highest similarity was chosen and recorded as the *CoLiDeS selection*. Third, *CoLiDeS+ selection* was calculated on the same data by adding the path to each option and selecting the option with the highest path adequacy (goal-similarity of a path) (see the CoLiDeS+ algorithm presented in Section 4.3). Results are summarized in table 10. It shows that CoLiDeS+ simulates users slightly better than CoLiDeS (Chi square = 3.52; p=.06; marginally significant).

Table 10. CoLiDeS+ makes the same selection as the user in 54.9 % of cases, simulating users slightly better than CoLiDeS.

Total valid selections	275	%
CoLiDeS selection matches user selection	129	46.9
CoLiDeS+ selection matches user selection	151	54.9

4.4.2.2. Effect of navigation support on task performance

Providing navigation support made a significant difference in users' navigation behavior and task performance. The number of navigation steps was lower in the support condition than in the control condition (t=3.86; p=.001). It took an average of 30 steps to execute a task in the control condition and only 19 steps in the support condition. The average duration of tasks was shorter in the support condition than in the control condition (t=2.16; p=.04). It took an average of 10.26 minutes to complete a task in the support condition and 12.49 minutes in the control condition. Task performance was significantly higher in the support condition (mean=1.16) than in the control condition (mean=.68) (t=2.16; p=.04). The score on task performance (natural logarithm of correctness divided by time) ranges from –1.12 to 2.07 and the distribution is sufficiently normal.

4.4.2.3. Effect of navigation support on task performance for users with low spatial ability

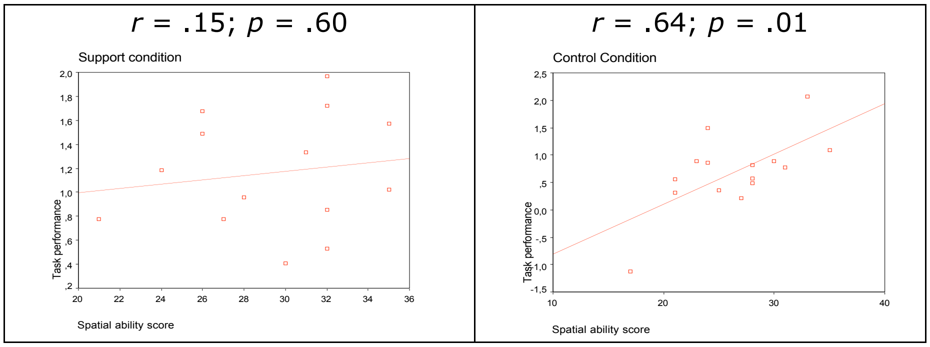


Figure 9. The correlation between spatial ability and task performance becomes non-significant in the support condition.

As expected, the correlation between spatial ability and task performance was significant for the control condition (r=.64; p=.01) and not significant for the support condition (r=.15; p=.60) (Figure 9). The difference between the two r coefficients does not reach statistical significance (z=1.45; p=.07) because of the small number of cases in the two groups (14 and 15, respectively).

Participants were divided in two groups with high and low spatial abilities (the median of test scores was taken as a cutting point). The difference in task performance induced by navigation support was checked separately for low and high spatial ability participants. Results based on the means are depicted in Figure 10. One can see that the difference induced by navigation support is bigger for participants with low spatial ability (t=2.27; p=.04) than for participants with high spatial ability (t=.73; p=.48). Again, because of the small number of participants in the two groups, the interaction between condition and levels of spatial ability did not reach statistical significance (F=1.45; p=.24).

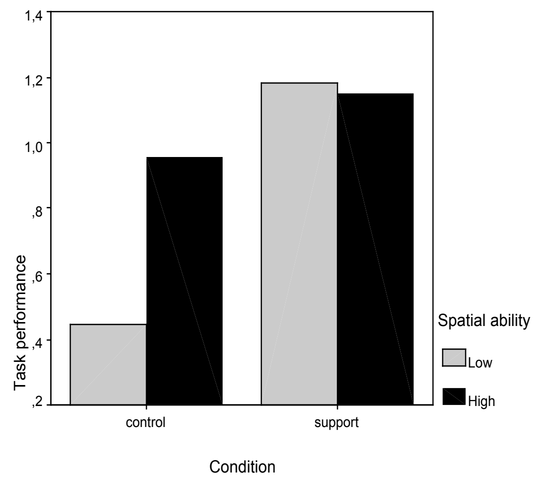


Figure 10. The effect of providing navigation support on task performance for users with low and high spatial ability

4.4.3. Conclusion

CoLiDeS+ simulated users' task execution better than CoLiDeS based on information scent and path adequacy – a measure of contextual information involved in user navigation. When simulated solutions were offered as navigation support, users performed better the given tasks. It seems that the offered navigation support prevented users from spending time and cognitive resources with those navigating actions that are not directly effective but are usually employed in order to accurately represent the information structure. Users engage in apparently useless navigation actions in order to get acquainted with the context of a particular piece of information, which is eventually useful in judging the value of that particular information. By making use of path adequacy CoLiDeS+ gives an account for this type of behavior.

The correlation between spatial ability and task performance indicates that users' ability to mentally represent and manipulate information spaces is crucial for Web navigation tasks. However, when provided with navigation support, users with low spatial abilities had a higher performance increase than users with high spatial abilities. It follows that users with low spatial abilities are probably less able to represent the information space and this is why they benefit considerably when the cognitive model is doing this job for them. We consider this result to be an additional proof that CoLiDeS+ gives an account not only for the process of assessing relevance of link labels to users' goals but also for the ability of users to represent and manipulate an information space.

To illustrate these ideas, let us review the example from 4.3.1.1. The model suggests only its successful trials, in this case the path "sleep book accommodation now economical rooms". It does not recommend (it suggests "go back") its own unsuccessful trial "sign up for a waiting list". Thus, the model "took over" the task of exploring the information space and ruled out the irrelevant information. For humans, exploring an information space requires spatial abilities and less skilled users are at risk of "getting lost". If these users take the offered suggestions their risks of disorientation and decreased task performance are lowered.

In conclusion, path adequacy is a possible means in modeling user's ability to take into account contextual information when assessing relevance of link labels. By making use of this means, a cognitive model might be able to give an account for individual differences in users' spatial ability (measured with a mental rotation task).

4.5. Study "Graphical suggestions"

During the previous study ("Voice suggestions") it was noticed that providing suggestions in the auditory modality was perceived as rather disturbing for users' natural navigation behavior. In a post task questionnaire, 8% of the participants perceived the suggestions as annoying and 31% of the participants felt being manipulated by the given suggestions. Therefore, we decided to offer suggestions in the visual modality and also to investigate subjective consequences of providing navigation support. In addition, we were interested in examining further the difference made by the provided navigation support in users' navigation behavior.

Navigation support was generated in the same way as in the first experiment. Only link suggestions were offered this time. Figure 11[12] presents how the link "Good to the last drop" is suggested by having two (red) arrows pointing at it. Path switch suggestions ('Go back') were dropped based on observation from the first experiment. It was difficult for the experimenter to determine the usefulness of such suggestions, since users with a higher tendency to pursue dead-ends were more exposed to 'Go back' suggestions, which were perceived as annoying, this perception could have affected task performance, etc.

[12] Only the focus-on area is presented in this figure, the whole webpage can be found at http://home.howstuffworks.com/coffee.htm

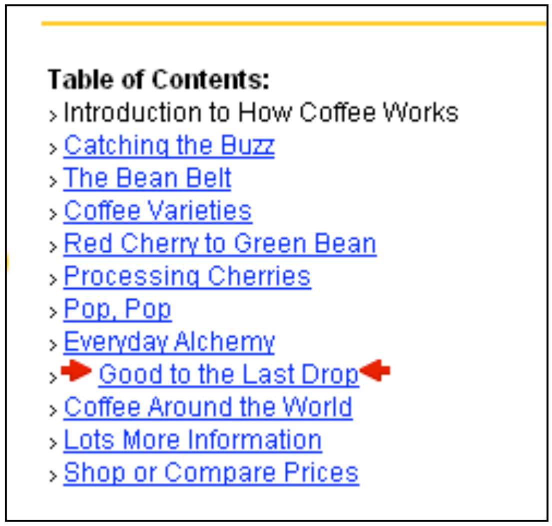


Figure 11. Example of link suggestions:
the two arrows point at a particular link text (website: www.howstuffworks.com).

4.5.1. Method

Participants, tasks and materials had the same characteristics as in the first experiment. This time 32 participants were recruited and randomly assigned to two conditions (*navigation support* versus *control*).

As dependent variables we considered users' perception measures, navigation metrics, behavior measures and task performance.

Users' perceptions (opinions, attitudes) were measured with a post-navigation questionnaire. The questionnaire contained items referring to perceived usability of the used websites (e.g., 'I could effectively complete my tasks using the provided websites'), and users' perceived disorientation (Ahuja & Webster 2001) (e.g., 'Quite often I unexpectedly returned to a page I have visited before'). 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 had to fill in 4 additional items about how they perceived the provided suggestions. These items are: '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 set of metrics of users' navigation behavior described in Section 4.1 was used.

User sessions were video recorded and coded afterwards by two experts. The following measures were coded: reading time, reactions to suggestions and task performance. Reading time included time users spent on reading contents such as text paragraphs and images; inspecting links and headers and user actions such as scrolling and typing were not scored as reading time. For each suggested link, the two experts noted if the user noticed and followed the suggestion. Task performance was coded based on predefined criteria specifying what a correct and complete answers is for each task. Part of videos were double coded in order to check for the interrater reliability; the correlation beween scores given by the two experts was: $r = .984$, $p = .000$, $N = 40$.

4.5.2. Results

Suggestions were generally well received. Based on the questionnaire, it appeared that most of the participants did not perceive the suggestions as intrusive, annoying or manipulative; a relatively high number of participants (11 out of 16) trusted the suggestions; but there is no clear evidence that suggestions were perceived as useful.

The differences between conditions were non-significant with regard to perceived disorientation and perceived usability. 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 .10 (two tailed). An interaction between support and gender was suspected to be responsible for this and the variable *gender* has been included in the model. As a result, there is a significant interaction between the variable *gender* and the variable *support* in relation to perceived disorientation ($F=5.12$, $p=.03$). Thus, men and women subjectively benefit to different extents from being provided with suggestions. When this interaction between gender and support is taken into consideration, the effect of support becomes significant ($F=9.43$, $p<.01$). Therefore, it is now clear that there is a significant effect of providing suggestions on perceived disorientation, but only for men (Figure 12).

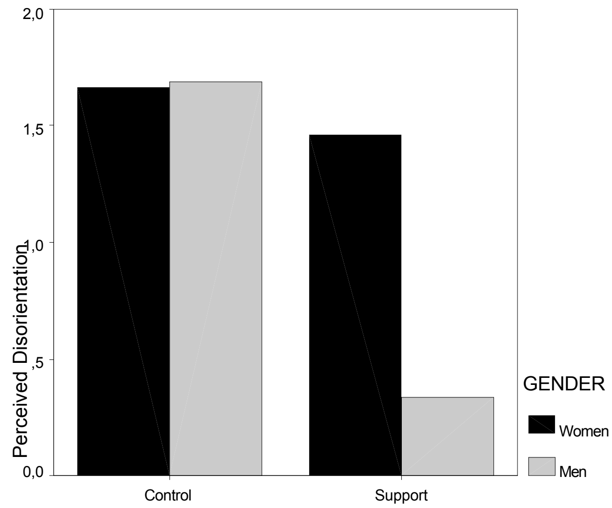


Figure 12. Perceived disorientation decreases in the support condition to a greater extent for men than for women.

We also checked if link suggestions changed the structure of users' navigation behavior. In the support condition, participants used the back button less (t=-2.24, p=.03) and the average connected distance in the navigation path was higher (t=2.26, p=.03), than in the control group. Thus, link suggestions caused the subjects to navigate in a more linear manner and reduced the amount of backtracking.

We expected the *coded reading time* to be correlated with the *logged reading time* – a second-order metric calculated based on first-order view times (see Section 4.1). This correlation was non-significant. However, there was a significant correlation (r = 0.376; p = 0.041; N=30) between *coded reading time* and a first-order metric – *meanmed*. As described in Section 4.1, meanmed indicates that a relatively small number of pages are viewed for a relatively long time. This first-order metric was better at indicating reading time than a second-order metric that we hypothesized to indicate reading time. This second-order metric was derived in such a way that it is correlated with task performance (see Section 4.1). Assumed was that more careful reading would lead to better task performance. This assumption was proven to be not entirely correct: the correlation between the coded reading time and task performance does not reach statistical significance (r = .23, p = .22, n = 30). Perhaps the relationship between reading time and task performance is more complex. In line with our findings so far, task performance should depend on both syntax and semantics, whereas reading time refers to merely the semantic aspects. To check for this new hypothesis we built a stepwise regression model in which we included besides the *coded reading time* all syntactic measures we had (first-order metrics and navigation style). Results of this regression analysis are sumarised in table 11:

Table 11. Results of stepwise regression analysis showing that syntactic (ACD) and semantic (reading time) measures are complementary in determining task performance.

Dependent variable	Predictors	Beta	t	Sig.	R	R square	Adjusted R square
Task performance	ACD	.597	3.954	.000	.633	.400	.356
	Reading time	.326	2.157	.040			

These regression results show that there are two factors affecting task performance: the average connected distance (ACD) of the navigation graph (a measure of how effectively an information space is traversed) and the amount of reading (a measure of how carefully the content is inspected). These two factors are independent of each other and the first one is relatively more important (Beta = .597) than the second (Beta = .326).

When comparing these results with the prediction of task performance based on *spatial ability* and *domain expertise* from the study reported in Section 4.2., one can notice once more the complementariness between *structural* aspects (spatial ability, ACD) and *content* related aspects (domain expertise, reading time) that characterizes Web use.

With regard to users' reactions to the provided suggestions, we expected that suggestions were often taken, and taking suggestions would increase task performance.

On average, a user was offered 168 link suggestions over the five tasks. Out of these offered suggestions, only an average of 21% were considered (the rest were not noticed or ignored). Out of those suggestions that were considered, only 40,39% were taken. In absolute terms, an average of 14 suggestions per user were taken.

When trying to check if taking suggestions increases task performance, first a non-significant correlation was found. However, by inspecting the scatter plot (fig. 13) one can notice that users' tendency to take the offered suggestions interferes with task performance, that is, high performers tend to take fewer suggestions than average and low performers.

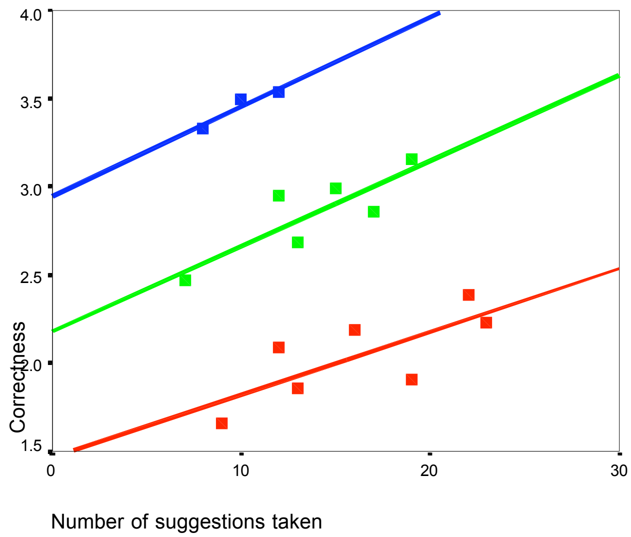


Figure 13. Scatter plot for the correlation between Number of suggestions taken and Task performance (Correctness)

Three different levels of task performance can be distinguished based on users' tendency to take suggestions and to benefit from taking them. High performers (upper line) tend to take a small number of suggestions, average performers (middle line) tend to take a higher number of suggestions, and low performers (lower line) tend to take the highest number of suggestions. However, within each performance level, taking more suggestions is associated with increased performance. The partial correlation between the Number of suggestions taken and Task performance, controlling for the variable composed of the three performance levels is: $r = .79$; $p = .001$; $N=16$.

The difference in task performance induced by the offered navigation support followed the same trend as in the first experiment (performance is higher in the support condition), although this time it didn't reach statistical significance ($F = 2.07$, $p = .16$). However, when the *coded reading time* is controlled for (included in the ANOVA model as a covariate) the difference in performance between conditions becomes significant ($F = 5.325$, $p = .029$). Thus, for the same level of reading time supported users perform better than controls.

A last analysis was performed to check for a possible interaction between gender and support with regard to task performance as it was the case for perceived disorientation. This interaction was not found ($F = 0.14$, $p = 0.71$), mainly because men slightly outperform women (the difference is not significant) in both conditions (see fig. 14, compare with fig. 12).

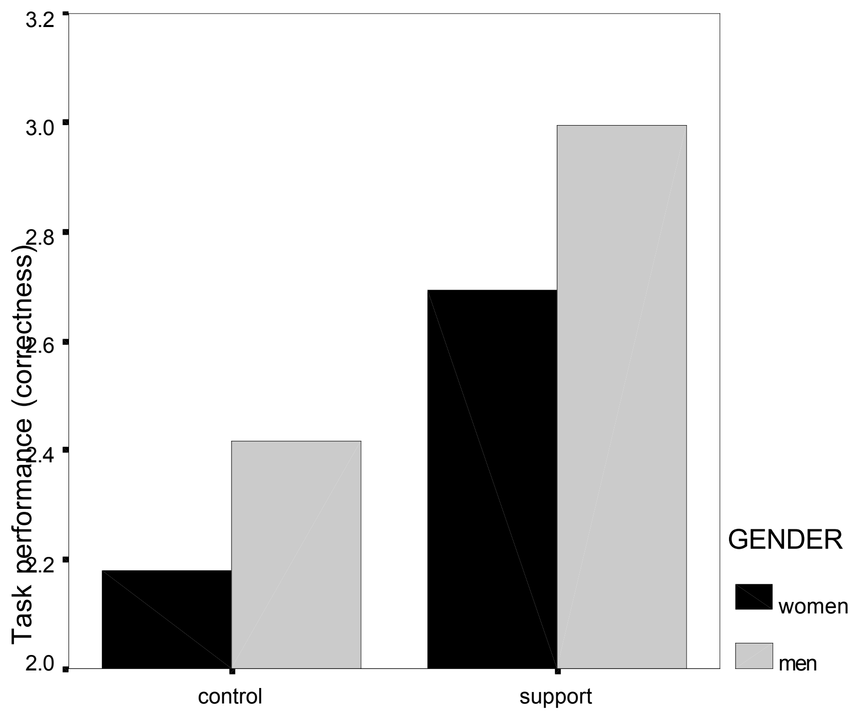


Figure 14. Differences in task performance between men and women and control versus support; the difference between men and women is constant across conditions, thus there is no interaction between gender and support.

Thus, based on an objective measure of task performance, one can say that both men and women benefited from suggestions. Other authors reported significant differences between men and women in favor of men with regard to Web task performance (Roy and Chi, 2003)

4.5.3. Conclusions and discussion

This study confirmed the dual (syntax-semantics) nature of Web task performance and showed the subjective and objective value of providing model-based navigation support in a graphical form.

Throughout the entire chapter, key steps of the model development process have been documented:

- Task analysis has identified the relevant criteria and measures, necessary to ensure ecological validity from the very beginning of the research process.

- Correlational research was aimed at selecting the most important determinants of Web navigation behavior. These determinants were shown to be: domain expertise and spatial ability.
- A cognitive model (CoLiDeS+) was built based of existing models of Web navigation. This model was intended to specify and explain the role of the two main factors identified by correlational research. Specifically, the role of domain expertise was modeled as judgments of relevance based on knowledge represented in semantic spaces; and the role of spatial ability was modeled as the involvement of contextual information in relevance assessments.
- The model has been empirically tested for how well it simulates the actual user behavior and whether it is useful in generating Web navigation support. Although it does not simulate the user behavior particularly well, CoLiDeS+ was shown to perform better that its previous version (CoLiDeS).
- CoLiDeS+ was used to generate navigation support and this support had a positive impact on user behavior and task outcomes.

While these results seem promising, there are a number of limitations that, on one hand weaken the scientific power, and on the other hand, prevent a direct implementation of these ideas in the practice of Web engineering and design:

- We have assumed that spatial ability was related to representing the information space and operating on it. In other words we have related spatial ability to navigation across pages and not with the process of visual search within a page. This assumption was based on the correlation found between spatial ability and syntactic metrics such as the re-visits, back button use, and fan degree (see Section 4.2). Our modeling decisions were based on this assumption. However, there is no clear evidence that this is indeed the case. In normal Web use, there are quite a few visuo-spatial features that are not related to navigation across pages (e.g., page layout, color, size, highlights) and that could determine the correlation between spatial ability and task performance. A complete proof would be to eliminate these features and show that the effect of spatial ability on task performance remains. Luckily, a situation like this (i.e., lack of visuo-spatial features) exists in real Web use of visually impaired persons via screen readers. Next chapter will present a study in which the screen was turned off and users had to make use of a screen reader in order to navigate and complete a series of Web tasks.
- The results of CoLiDeS+ simulations are toward the expected direction but not particularly compelling. This is mainly due to

the weakness of the machine learning technique used to compute semantic similarities (LSA). The situation is aggravated by the fact that we have used real Web applications and (as much as possible) realistic tasks. This has resulted in highly heterogeneous and complex experimental material. For example, we have used goal descriptions and link labels composed of multiple words. In this case, small errors of LSA at the word level are propagated to the passage level, and eventually impact on the accuracy of our simulations.

- CoLiDeS+ is to a large extent handcrafted, it is a bit better specified than a pure boxes-and-arrows model, but it is still far from being a computational model. However, there are important scientific and practical reasons for such a model to be implemented as a computational model. The scientific value of computational models has been discussed in Chapter 3 and is extensively argued elsewhere (Goldman, Golden & Van den Broek, 2006). From a practical point of view, a computational model can be used as a user model and can be integrated with other tools in an adaptive Web application.

It is beyond the scope of this project to give complete solutions to all these limitations. However, the next chapter suggests possible ways to overcome these limitations.

Chapter 5. Extension and Applicability of Model Development

It is in the nature of a PhD research to raise more issues than resolve. This chapter is intended as a continuation and extension of Chapter 4. While doing the research reported there, we have become aware of a series of limitations and opportunities and have generated ideas about future research directions and practical applicability of this research. This chapter has two sections:

First section of this chapter reports a study focused on a particular user population – visually impaired persons (VIPs). Motivation for this study was two-folded:

- From a scientific point of view, we needed to isolate an essential feature of Web navigation – handling a hyper-structure of linked documents – from other aspects of screen-based interaction. Using the Web by VIPs via screen readers is a perfect instance of such isolation: visual features such as page layout, font color and size are absent.
- From a practical point of view, this user population is really in need of Web navigation support, much more so than sighted users. We wanted to investigate a way to deliver the kind of support CoLiDeS+ generates and whether this support makes a difference in users' performance and satisfaction.

This section also elaborates on the idea of providing model-based navigation support for VIPs. It shows that our empirical and modeling work is only a modest beginning and more work is necessary in the areas of Web engineering, artificial intelligence and user interface design in order to build effective navigation support.

The second section addresses another limitation of the empirical and modeling work presented in Chapter 4 – the large amount of hand-coding involved in CoLiDeS+ simulations – and presents a way to go toward a fully computational cognitive model of Web navigation by using the ACT-R cognitive architecture and modeling environment.

5.1. Study "Visually Impaired Persons"[13]

This study was intended to investigate the Web use of visually impaired persons (VIPs), to what extent the kind of support generated by CoLiDeS+ can be delivered in a way that does not increase the information load of the interface and is unobtrusive, and whether or not

[13] This study has been conducted in cooperation with Arnaud Lek, Content & Knowledge Engineering master's student at the Institute of Information and Computing Sciences, Utrecht University.

this support has a positive subjective and objective impact on users. We hypothesized that:

- Using the Web via a screen reader depends essentially on the same cognitive mechanisms as in the case of using a graphical screen. In particular, the correlation between spatial ability and task performance will be replicated.
- Suggestions can be effectively delivered by increasing the reading priority of the relevant items.
- Suggested users will perform better and be more satisfied that controls.

5.1.1. Problem definition and background

Visually impaired persons accessing the Internet via *screen readers*[14] have difficulties locating goal-relevant information (Jones, Farris, Elgin, Anders & Johnson, 2002). Using the Internet naturally involves re-visits to certain pages and one of the most frequent user actions is pressing *the back-button* (Cockburn, Greenberg, Jones, McKenzie & Moyle, 2003). Re-visitation is not only a means to correct for superficially processed or forgotten information, it is also a way to involve information contexts in judging the relevance of a particular piece of information. Due to re-visitation, VIPs have to redundantly listen to large amounts of content and options. This slows down the process and adds information load that must be handled with users' limited cognitive resources. Users' cognitive overload increases and the quality of their Internet experience decreases. This might explain why only 21% of the VIPs have access to the Internet, whereas 57% of the sighted persons have access to the Internet (Gerber & Kirchner, 2001). But, even when basic access to the Internet is ensured, questions arise regarding the quality of users' experiences with the Internet. How much do they really benefit from using the Internet? Are they fully capable of taking advantage of the whole functionality? (Di Blas, Paolini & Speroni, 2004)

Current screen readers read out in speech or Braille characters the content and options on Web pages in a non-selective way. They do not provide enough support in discriminating between various information types. Sighted persons have size, colors, position, shape, motion, etc. as aids to discriminate between various types of information. Screen readers have only phonetic cues (e.g., male vs. female voice) and verbal cues (e.g., the word "link" is added after each link). Moreover, there is no support in discriminating goal-relevant information from "noise". Information is presented serially and not in parallel as in visual interfaces. Visual and contextual cues are either absent since they

[14] A screen reader is a software program which converts screen information to Braille characters and/or speech

cannot be translated in a textual form or difficult to retrieve because re-visitation is non-selective. Selective reading as an alternative to exhaustive reading would be a natural solution. This requires adding "intelligence" to screen readers. An intelligent agent could be working together with the screen reader prompting the user with goal-relevant selections.

Significant work has been recently devoted to building tools for assisting VIPs in their Internet use. One approach is concerned with making auditory interfaces as similar as possible to visual interfaces. Thus, tools and techniques proven useful for sighted users can be converted so as to be useful for VIPs as well (Frauenberger, Stockman, Putz & Holdrich, 2005). For instance, an analog of visual scanning can be performed on audio data by speeding up the reading process (Hurst, Lauer, Burfent & Gotz, 2005). Auditory icons (Petrie, Morley, McNally, O'Neill & Majoe, 1997) and audio feedback indicating the user's location can be provided by a combination of speech and non-speech sounds (Strain & McAllister, 2005). Another approach focuses on treating the content of Web pages (Yu, McAllister, Kuber, Murphy & Strain, 2005) and modifying them so as to make them accessible for VIPs, for instance, by summarization (Zajicek, Powell & Reeves, 1998).

When such tools are based on theory and research, they have a higher chance to be successful and they can be generalized beyond their initial application domain (Neerincx, Lindenberg, & Pemberton, 2001). In turn, attempts to build and implement such tools can lead to validation or refinement of the theories that inspired them in the first place.

5.1.2. Method

A study of the same type as the two studies presented in Sections 4.4 and 4.5, respectively, was conducted. The main difference was the context of use: participants had to perform Web navigation tasks with the screen of the computer switched off and, instead, with the aid of a screen reader.

5.1.2.1. Website and Tasks

A modified version of www.howstuffworks.com has been used. This website provides explanations of various phenomena (e.g., how toilets work, how electronic pet fences work). A selection of approximately 200 topics spread on approximately 1000 pages has been used. All pages have been modified as to contain only text-based information about the selected topics; ads and graphics have been removed.

Two tasks were developed in the same manner as described in Section 4.4. One task was about sleeping and dreaming and the other one about dieting. The screen reader "Home Page Reader" developed by IBM has been used in executing the tasks. This is one of the most common tools of this kind on the market.

5.1.2.2. Experimental manipulation

A between subjects design of the same kind as used in Sections 4.4 and 4.5 was used: a manipulation of the website was performed only in the "support" condition and it was hypothesized that this condition would differ from the control condition with regard to user behavior and task outcomes.

Unlike in the experiments presented in Sections 4.4 and 4.5 where suggestions were built by running the model in advance and identifying correct paths and dead-ends, this time the navigation support was less deterministic. The items to be emphasized were selected based only on current information and past selections. The reasons for this were to avoid a large amount of hand-coding and to allow partial automation in support generation.

As to the way of emphasizing the suggested item, there were not so many options. One option was to add a word or a small expression to the suggested item in the same manner as in the experiment reported in Section 4.4. This option has been proven to be effective but not so well received subjectively by users. In addition, in the case of VIPs' Web use, this option adds information load, since all the information comes via the auditory modality. Instead we have decided to suggest an item by changing its reading priority. Since the screen reader always reads from upper-left to lower-right parts of the screen, increasing the reading priority of an item means moving it in the reading sequence upper and to the left. For example, if the user is looking for "electronics" in the following menu "books/music/electronics" it will take at least 3 times longer to find it than if the user were looking for "books". So "books" is "highlighted" by default because of the sequential character of reading. If "electronics" is to be highlighted, it should be placed on the first position in the menu. The assumption behind this was that users would hear the relevant item sooner and would make their selection before the whole screen is read, avoiding information overload and, in cognitive terms, having to deal with simpler representations of the information space. To maintain a minimal consistency of the website's structure, the main menu on each page (i.e., the highest level in the website's hierarchy) was left unchanged.

5.1.2.3. Measures, Participants and Procedure

Since this time the experimenter has not controlled the quality of suggestions in advance, a specific measure has been conceived to verify that emphasizing the goal-relevant items was successful, that is, the reading priority of the relevant items was indeed higher in the support condition than in the control condition. We have recorded and determined the position of goal relevant options in the reading sequence. For example, if a goal relevant item is placed second on a webpage it will get position 2.

Task performance was measured in two ways:
- directly, by counting the number of target pages participants have arrived at;
- indirectly, based on correctness of answers to a set of multiple-choice questions.

Prior knowledge on the topics of the tasks was also measured by administrating the task performance questionnaire before task execution.

Perceived usability, Satisfaction and Perceived Disorientation have been measured with the same questionnaires as in the previous studies; spatial ability test and the same set of navigation metrics were used (Chapter 4). A few questions were added to measure the perceived usability of the screen reader. For example: "It was easy to learn how to use this screen reader" and "The screen reader had all the functions and capabilities I expected it to have".

A number of 29 undergraduate and graduate students at Utrecht University participated in this experiment. An equal number of participants (13) has been assigned at random to the two conditions. Three subjects used for the pilot tests have been added to the control condition, so in the end there were 16 subjects in the control condition and 13 in the support condition. For the three pilots added to the control condition no data has been recorded on navigation metrics.

Participants were not visually impaired and had no prior experience in using a screen reader. Vision impairment was simulated by turning the computer screen off. Instruction and training on how to use the screen reader were provided. Sessions took at most two hours and consisted of taking the spatial ability test, performing the two navigation tasks and filling out the questionnaires. The two navigation tasks were presented to the participants via the screen reader. The order of the two tasks was counterbalanced. Participants received financial compensation for their participation.

5.1.3. Results

Results are presented in the same order as objectives: first some characteristics of using the Web via a screen reader, second the effectiveness of the experimental manipulation and third the impact of this manipulation on users.

5.1.3.1. Particularities and commonalities of using the Web via a screen reader versus via a graphical screen

Unlike in the previous experiments (Chapter 4), there is a high cost associated with visiting many pages via a screen reader: the longer the path (number of pages visited), the higher the chance of perceived disorientation ($r = .408$, $p = .039$, $n = 26$), the lower perceived usability ($r = .657$, $p = .000$, $n = 26$), and the lower perceived usability of the screen reader ($r = .637$, $p = .000$, $n = 26$).

There are also results suggesting that using the Web via screen readers involves the same underlying cognitive processes as using the Web via a graphical screen:

- Spatial ability is correlated with task performance ($r = .564$, $p = .023$, $n = 16$, control condition).
- Users with high spatial ability employ less re-visitation ($F_{1,23} = 3.02$, $p = .095$, $n = 26$, marginally significant).

Other results suggesting a high commonality between the two contexts of use are:

- The number of re-visits is negatively correlated with perceived usability ($r = -.545$, $p = .004$, $n = 26$) and perceived usability of the screen reader ($r = -.534$, $p = .005$, $n = 26$).
- Task performance is correlated with perceived usability ($r = .421$, $p = .023$, $n = 29$) and perceived usability of the screen reader ($r = .393$, $p = .035$, $n = 29$).

5.1.3.2. Effectiveness of the "suggestion" mechanism

As a result of our manipulation, the position of an item has been decreased in the reading sequence from an average of 4.88 in the control condition to an average 3.84 in the support condition. That is, in the support condition links have successfully been pushed up. In 9 out of 25 cases the relevant item has been pushed up to the first position in the reading sequence.

5.1.3.3. Impact of suggestions on users

An analysis of covariance with task performance (correctness) as a dependent variable, condition as a factor and prior knowledge as a covariate resulted in no significant differences between the two conditions ($F_{1,26} = .446$, $p = .510$, $n = 29$). There were also no significant differences with regard to the other measure of task performance (number of target pages reached), satisfaction, disorientation, and navigation metrics.

This lack of impact of our manipulation on users can be explained by one or both of these factors:

- The magnitude of our manipulation (moving the relevant item with approximately one position up in the reading sequence) was not big enough to produce a significant effect.
- Changing the order of items on Web pages produced a break of the initial coherence as established by the authors. This "side effect" of our manipulation presumably interacted with (and cancelled out) the expected positive effect.

While not excluding the first explanation, we tend to favor the second one, based on two related findings:

- The correlation between spatial ability and task performance behaved in the same way as in the "Voice suggestions" study (Section 4.4): the correlation is significant in the control condition and non-significant in the support condition. This means that our manipulation had indeed an impact on users.
- Users in the support condition visited a larger set of pages than users in the control condition, while re-visitation was constant across conditions. This could probably be caused by users' efforts to repair the breaks of coherence caused by changes in the initial order of items on Web pages.

The following is an example of manipulation that might have caused coherence breaks. The original page has the following options:

- What is caffeine?
- Caffeine in the diet
- Caffeine and adenosine
- Caffeine and dopamine

The manipulated page has an increased reading priority of the option "Caffeine and adenosine", causing the items to be presented in the following order:

- Caffeine and adenosine
- What is caffeine?
- Caffeine in the diet
- Caffeine and dopamine

Perhaps it is confusing for the reader to first present very specialised information on the effects of caffeine on the brain ("caffeine and adenosine") and only after that to present general information about what caffeine is.

5.1.4. Conclusion and discussion

This study was intended to get insight in VIPs' Web use and to further test the assumptions behind our cognitive model and its practical relevance. It can be concluded that the same cognitive mechanisms are employed as in the case of sighted users – representing the information space, operating on it, and making selections based on judgments of goal-relevance (information scent). However, using the Web via a screen reader is more taxing from a cognitive point of view, and users are more vulnerable to dissatisfaction and disorientation if they have to visit a large number of pages to reach their goal.

Our way to emphasize the goal-relevant items by increasing their reading priority did not have the expected impact on users' performance and satisfaction. It might be that other ways of emphasizing could have been more effective, for example, changing the audio properties of the reading voice. The lesson learned from this result is that there is no simple way to translate results of empirical research into practical applications.

Visual impairment was simulated by turning off the computer screen and asking sighted users to perform Web tasks via a screen reader. This might raise the question to what extent results of this study can be generalized to real VIPs. The decision to use such a study design was taken based on (besides practical limitations) discussions with experts in the area of Web accessibility. According to these experts (Velleman & Snetselaar, 2000) most of the VIPs are faced with the same situation: they need to learn using a screen reader because they (partially) lost their vision. However, our sample has a lower age than the VIPs population. It is to be expected that real VIPs have even higher difficulties in using the Web than our participants, due to their increased age. It would be useful to replicate this study using VIPs as participants.

5.2. Toward a fully computational cognitive model of Web navigation

There are theoretical, methodological and practical reasons for developing not only conceptual but also computational cognitive models. According to Goldman, Golden, and Van den Broek (2006), computational models promote the development and evolution of conceptual theories showing where the theories accord with behavioral data and where they do not; they can be used to understand and test alternative explanatory constructs; and they promote communication among researchers within and across research areas. In addition, particularly in our field, cognitive models can be used as substitutes for users (Ritter, Baxter et al. 2000) and as basis for building user models in adaptive Web applications.

It was not an objective of this PhD project to develop a computational cognitive model of Web navigation. However, if this research were to be continued and extended toward developing support tools for Web interfaces, converting our conceptual model in a computational model would be highly opportune. This section provides only a demonstration of such a computational model.

For this demonstration, the ACT-R (Adaptive Control of Thought - Rational) cognitive architecture (Anderson, Bothell et al., 2004) and its associated modeling environment - ACT-R6 (Bothell, 2005) have been used. ACT-R has validated means to model the key concepts of Web navigation behavior presented earlier. Thus, ACT-R perceptual modules can handle sequential information input as in the case of using the Web by VIPs. Information is coded based on its content ("what") and location ("where"). The declarative memory module is appropriate to model the dynamic representation of the information space being navigated, since information elements (chunks) can be stored, modified and retrieved according to cognitively plausible sub-symbolic mechanisms such as spreading activation, decay, frequency and recency. Symbolic rules from procedural memory fire depending on the state of the contextual information stored in buffers. These rules issue appropriate actions that have internal and/or external consequences (fig. 1).

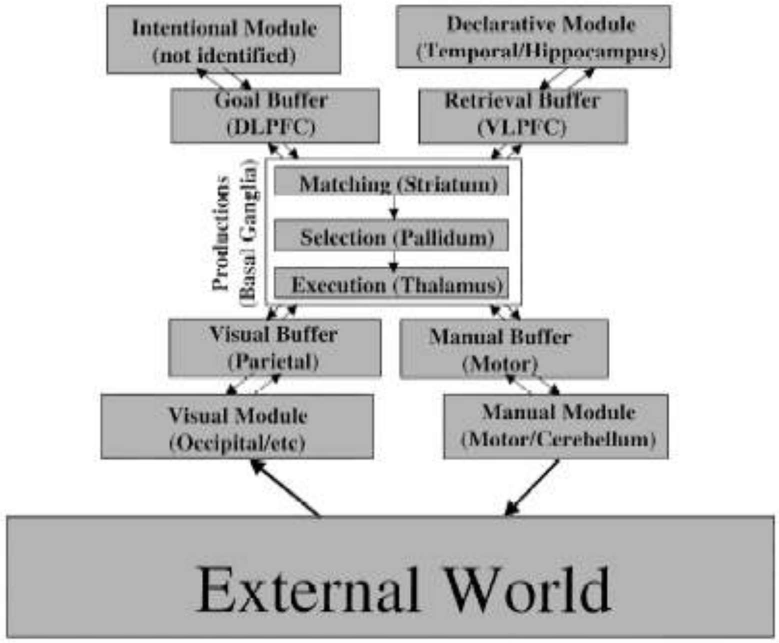


Figure 1. An overview of the ACT-R architecture (Anderson, Bothell et al., 2004)

The remainder of this section will present an ACT-R model of Web navigation and discuss it in comparison with previous models such as CoLiDeS, SNIF-ACT, MESA and CoLiDeS+.

5.2.1. An ACT-R model of Web navigation

In order to achieve its demonstrative purpose, the model presented here interacts with a small scale website and uses an external tool – Generalized Latent Semantic Analysis (GLSA) – to represent background knowledge and calculate semantic similarities. The website is a simplified version of How Stuff Works (http://www.howstuffworks.com/). As its title suggests, this website provides explanations of various phenomena such as hurricanes, hypnosis, etc. Information on Web pages has been reduced to link labels and each label is composed of a single word. Starting with the home-page, Web pages are presented to the model one after another depending on the model's selections. For instance, Figure 2 shows the homepage as it is presented to the model. The model selects one of the presented link labels causing the corresponding next webpage to be displayed.

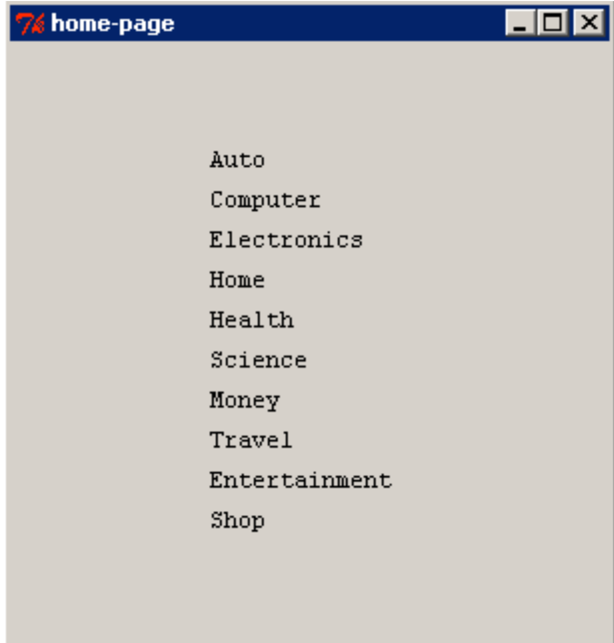


Figure 2. A example webpage presented to the model

For this demonstration, the model has the task to find out *how coffee works*, more simply stated, to find a webpage with information about *caffeine*. Therefore, the following goal has been set for our ACT-R model:

ACT-R syntax	English
GOAL ISA COMPREHEND-HYPER-TEXT INTENTION "coffee" SELECTION NIL LAST-SELECTION NIL THRESHOLD NIL PAGE-SWITCH NIL STATE START	The goal chunk has the type comprehend hypertext, and the following slots: intention, selection, last selection, threshold, page switch, and state. Each slot has a value, for instance, the intention slot has the value "coffee".

The Vision module[15] of ACT-R is used to perceive (locate and encode) information on Web pages. The following rule finds the location of an unattended text element on the current Web page:

ACT-R syntax	English
(P attend-location =goal> ISA comprehend-hyper- text state start ==> +visual-location> ISA visual-location screen-y lowest :attended nil =goal> state attend)	If the goal is to comprehend hypertext, and the state of the goal is “start”, try to find a location starting at the top of the screen that has not been attended before, and change the state of the goal in “attend”.

The following rule moves attention toward the current location:

ACT-R syntax	English
(P attend-option =goal> ISA comprehend-hyper- text state attend =visual-location> ISA visual-location ?visual> state free ==> +visual> ISA move-attention screen-pos =visual-location	If the goal is to comprehend hypertext, and the state of the goal is “attend”, and there is a location in the visual-location buffer, and the vision module is free, move attention toward that location and encode whatever content is there.

After having seen all the options on a particular webpage, the model has to select one that is goal relevant, that is, semantically similar with the word “Coffee”. Semantic similarities are not computed by the ACT-R model, but imported from a tool called Generalized Latent Semantic

---

[15] The Audition module could have been used to simulate the web use of VIPs. However the vision module is more appropriate for our demonstrative purpose and it works exactly the same with regard to the core feature of VIP use – sequential input of information.

Analysis (GLSA). GLSA (http://glsa.parc.com) is similar in principle with LSA (see Section 3.2.4 for a description of LSA) but it uses the whole Web as a text corpus and other optimizations (e.g., stemming) meant to improve its performance.

Our model selects an option with a goal-relevance higher than a specified threshold; in this case the threshold is a free parameter and has the value 0.10. Goal relevancies for all the options on the home-page are presented in the Annex.

The following rule is used to select the goal relevant item:

ACT-R syntax	English
(P judge-goal-relevance =goal> ISA comprehend-hyper-text intention =intention selection nil state assess =retrieval> ISA similarity-fact term1 =intention term2 =term2 ==> =goal> selection =term2 state click)	If the goal is to comprehend hypertext, and a selection has not been made and the state of the goal is "assess", and a similarity fact has been retrieved then select the value of the second slot of the similarity fact and update the selection slot of the goal.

The option "shop" is selected because it was retrieved as having a similarity with coffee of .142 (higher than the threshold), and the shop page is displayed (fig. 3).

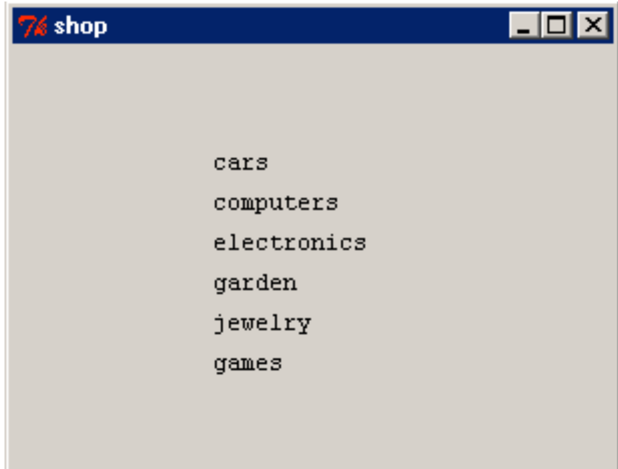


Figure 3. Example of a page loaded after a selection of the model. The option “shop” was selected on the previous page and the corresponding “shop-page” has been downloaded.

Starting with the second page, beside goal relevancies, information encountered on previous pages can be used when making current selections. Different ways to use past information are employed by the model and they will be discussed in separate sub-sections. The Annex presents a synopsis of information needed to understand the behavior of our model. It shows the most important Web pages, options on these Web pages (link labels), goal relevancies and several coherence coefficients. Goal relevancies are shown on the first line under each option. For example, the option “shop” on the home-page has a goal relevancy of 0.142. On the second line under some of the options, one can see a coefficient called “back-coherence1” – a semantic similarity between the current option and the previous selection of the model. For example, the option “garden” on the shop-page has a back-coherence1 of 0.33 (i.e., the semantic similarity between the current option “garden” and the model’s past selection “shop” is 0.33). On the third line, a coefficient called “back-coherence2” can be seen. This is calculated as the semantic similarity between the current option and the model’s past selection made 2 steps prior to the current step. For example, the option “supplies” on the garden-page has a back-coherence2 of 0.22 (i.e., the semantic similarity between the current option “supplies” and the model’s selection at two steps back “shop” is 0.22). The coefficients back-coherence3 and back-coherence4 are defined in an analog way.

5.2.1.1. Forward linear navigation

For applying a selection strategy based only on goal relevancy, the model retrieves a goal relevancy higher than 0.10 using the following rule:

ACT-R syntax	English
(P retrieve-similarity =goal> ISA comprehend-hyper-text intention =intention state attend ?retrieval> state free ==> +retrieval> isa similarity-fact term1 =intention > value 0.1 :recently-retrieved nil =goal> threshold 0.1 state assess)	If the goal is to comprehend hypertext, and the intention is temporarily stored in the variable "=intention", and the state of the goal is "attend", and there is no other retrieval process currently performed, request a retrieval of a similarity fact having the intention ("coffee") as the first term and a value higher than 0.1, that hasn't been retrieved recently, and remember the threshold, and change the goal state in "assess".

In case this retrieval attempt is successful, the model makes a selection and proceeds further on a new page. Based on this strategy, the model selects here "shop" on the home-page and "garden" on the shop-page. On the garden page, no option is found with a goal relevancy higher than the threshold.

5.2.1.2. Backtracking and lowering the threshold

In case a goal relevancy cannot be found, the threshold is lowered (as in MESA) to 0.05 and the request for retrieval is repeated. If there are no goal relevant options on the current page, options encountered on previous pages are considered (as in CoLiDeS+). ACT-R's sub-symbolic computations (base-level activation in this case) ensure that previous options are considered in the inverse order of their perception (the more recent ones are considered before the older ones). In other words, the model backtracks one step at a time, as most of the users do by using the most common function of the back-button.

As a result of this strategy, the model deploys a very elaborate navigation behavior. From the garden-page in our example, it selects "supplies" then "pool", it backtracks from the pool-page and selects "accessories" on the supplies-page, it backtracks again from the accessories-page and supplies-page to select "tools" on the garden-page, etc., etc. However, most of the selections it makes are not very helpful in understanding how coffee works. It finds some information about "hangover" and "alcohol" that is somehow related to "coffee", but other more relevant options are not selected. In particular, the page containing the option "caffeine" is not selected, although the option "caffeine" itself has a high goal relevancy.

5.2.1.3. Coherence with past selections

Based on the previous strategy the model took too large detours and was trapped in too many ambiguities and ramifications of the natural language. For example the word "shop" might have some semantic association with the word "coffee" but it has also thousands of associations with other words. When it jumps to "garden" (which is also somehow related to "coffee"), the initial association between "coffee" and "shop" is lost. By enforcing a coherence constraint one can ensure that the model stays within a limited set of meanings.

In case of this strategy, the model checks for both goal relevance and coherence with past selections. The model selects a goal relevant item only if its back-coherence1 is higher than 0.30. For this example, checking for back-coherence1 was enough to make the model avoid highly inefficient divagations such as "decor", "supplies" and "tools". Checking for higher order coherences would have made the model far too conservative. While this strategy was good enough at improving model's efficiency, it was not as good at improving the model's effectiveness: pages relevant to how coffee works are still not selected. This time not even the "hangover" and "alcohol" options are selected.

5.2.1.4. Conservative and explorative strategies intertwined

The coherence strategy is useful to avoid divagations but also too conservative. It leads to blockages – situations where there is no goal relevant and coherent item to be selected and backtracking and lowering the threshold have already been applied. In these cases the model is allowed to employ the explorative strategy just until it overcomes the blockage. A rule that fires only in blockage situations omits the coherence check and selects based only on goal relevance. After this, the model goes back to the conservative (coherence-based) strategy. This combined strategy makes the model both efficient and effective (as effective as the explorative strategy can ensure). However, the model is still not effective enough: pages needed to understand, in our example, how coffee works are still not encountered.

5.2.1.5. Post-valued recall

So far, our model has only used current information displayed on the screen and information remembered from its own past selections. Information previously encountered but not selected has been ignored. However, users reassess previously discarded items based on newly acquired information, a phenomenon known as post-valued recall (Wen, 2003).

Whenever the model makes a selection of a link label to be clicked on, it also requests a retrieval of a previously non-selected item that has a relatively high semantic similarity (0.20) with the currently selected item. If such an item is found, it is selected in spite of its low goal relevance and/or back coherence. After such unusual selection the model continues its regular behavior – enforcing goal relevance and back coherence.

Based on this strategy the model found a high semantic similarity between "alcohol" and "drugs" (.37), which led ultimately to finding the "caffeine" option on the addictive-page.

5.2.2. Conclusion and discussion

This model implements key features of Web navigation behavior as reported in literature and found in our empirical research. Some of these features are shared with previous models – selections based on goal relevance (information scent) (SNIF-ACT and CoLiDeS); backtracking, threshold and opportunistic strategies (MESA); back coherence (CoLiDeS+) – others are implemented here for the first time – intertwining between conservative and explorative strategies, and post-valued recall.

The ACT-R cognitive architecture allowed representation and simulation of the combination between syntactic and semantic aspects involved in Web navigation that has been shown to determine task performance. This syntax-semantics combination is inherent in the architecture. For example, a combination between "what" and "where" allowed building a complete and functional memory representation of the information space being navigated. Based on this representation it was possible to combine semantic similarity judgments with syntactic backtracking strategies; an item is retrieved at a particular moment not only because of its content but also because of its position in the sequence of items that have been inspected: a recency effect modeled by the base-level activation of chunks simulated the order of items within a path – the most recent item has the closest position to the current item; and spreading activation ensured retrieval of the chunk required by a particular state of the environment – a memory chunk is more likely to be retrieved if it is associated with the goal chunk.

There is still further work needed to ensure full cognitive plausibility and fit of the model with human data. We claim that our modeling decisions are founded in literature and in our own empirical research, but we also admit that more empirical and modeling efforts are needed to fine tune some of the parameters of the model. For example, the existence of a selection threshold and the satisficing behavior are generally accepted (Krug, 2000; Miller & Remington, 2001); however we don't know the exact value of this threshold and to what extent it should be lowered during backtracking. The values we have used (0.1 and 0.05) are based on our experience with LSA-like tools and modeling and experimentation with a variety of Web navigation tasks and websites. The same kind of "educated guesses" have been used for the back coherence and post-valued recall parameters.

In conclusion, in spite of the aforementioned limitations, the model presented here demonstrated a way to proceed toward building a fully computational cognitive model of Web navigation. This model needs to be extended in order to gain more automation and autonomy, that is, it should be less dependent on user input and able to interact with a standard interface and with other models/agents. Significant work is also needed for making this kind of model scalable to large information spaces.

In conclusion to the entire chapter, spatial ability is, indeed, related to handling a hyperspace and this relation is independent of visuo-spatial features on Web pages; providing model-based navigation support for VIPs should not break the intrinsic coherence of the content on Web pages; and development of a computational model of Web navigation is possible and opportune.

Chapter 6. Conclusion and discussion

This chapter starts with a summary of the main contributions of this thesis, then shortly concludes on the extent to which the objectives of this research stated in Chapter 1 have been attained, and ends with discussing some restrictions, limitations and opportunities encountered throughout this project.

6.1. Summary of contributions

Throughout this project, attention was paid simultaneously to theory, method and real-world applicability. Web navigation was grounded in the theories of Cognitive Science, in particular Text Comprehension, and Information Science, in particular Human-Computer Interaction. Methodological paradigms of experimentation, statistics and modeling have been applied. The whole research was placed as much as possible in realistic settings and practical needs of Web engineering and design has always guided research decisions.

This section presents a summary of the main contributions of our research. The first subsection presents contributions to Information Science and Human-Computer Interaction, and the second subsection presents contributions to Cognitive Science and Hypertext Comprehension. Arguably, this distinction is somewhat artificial since each contribution is part of all these fields; however, presenting our findings in this way emphasizes both the theoretical and the applied value of our research.

6.1.1. Contributions to the fields of Information Science and Human-Computer Interaction

There are three main contributions of our research to these fields: facts regarding the use of Web applications, methods to gather and interpret information about user behavior, and ways to conceive and deliver effective Web navigation support. Each of these contributions will be briefly summarized here. The interested reader is pointed back to sections of this thesis where more detail can be found, and we also cite here our own publications in which these aspects are treated.

6.1.1.1. Facts regarding the use of Web applications

We have documented how real Web applications are used. A large number of users have been invited to our Usability Lab, where they have performed Web tasks of various sorts from several domains (Juvina & Van Oostendorp 2003). Their behavior was recorded by various means: Web-logging, screen capture, video, audio, paper-and-pencil, etc. A large amount of data has been analyzed and interpreted in order to discover the regularities that can inform scientists and practitioners. The most important findings in this category are:

- Using the Web can be seen as a dialogue: users inform Web applications about their choices and Web applications "reply" with content. Thus, interaction paradigms such as reading/writing and talking/listening can be applied in understanding Web navigation behavior. In addition, involving spatial features (syntax) in processing contents (semantics) is a distinguishing characteristic of Web navigation.
- In order to adequately characterize Web navigation behavior, a complete set of criteria is needed, including objective (performance), subjective (satisfaction) and undesirable aspects of task execution (disorientation).
- Users re-visit pages not only because they forget what they have seen before, but mostly to get acquainted with the structure of the information space, which in turn helps them in preventing disorientation (see Section 4.1.2.2 where navigation styles are discussed).
- Using the Web via a screen reader is more taxing from a cognitive point of view, and users are more vulnerable to dissatisfaction and disorientation if they have to visit a large number of pages to reach their goal (see Section 5.1).

Observational and empirical research has allowed us to discover what are the most important user-related factors that determine performance in Web assisted tasks. Evidence from various studies (see Section 4.2) converged toward a combination of two factors: a structure related factor (spatial ability and average connected distance) and a content related factor (domain expertise and reading time).

Studying a large number of factors in relation to a comprehensive range of outcomes of Web navigation tasks was useful in several respects. A limited number of significant predictors were identified, and their relative contribution to the accuracy of predictions was estimated. Since factors were studied together and the stepwise method of regression analysis was employed, it was possible to rule out factors that were only marginally significant or confounded with one another. This is an important contribution of this research in comparison with other work of this type. Most of the studies addressing individual differences in Web navigation (including those referenced here) are restricted to a limited number of user characteristics, and for this reason they can easily overlook other (more important) characteristics. For example, the influence of working memory on hypertext navigation as reported by Tucker and Warr (1996) might have not appeared as significant if spatial ability was included as a predictor in their model (Tucker & Warr

1996). Our results show that *spatial ability* is more important for Web navigation performance than *working memory capacity*. These findings have been well received by the research community (Juvina & Van Oostendorp, 2005) and confirmed by more recent studies (Gugerty, Treadaway et al., 2006).

6.1.1.2. Methods to gather and interpret information about user behavior

Factors such as spatial ability can be measured only with specialized tests, which cannot be implemented in realistic Web applications. For this reason, we have proposed using Web-logging data to calculate metrics of Web navigation behavior (Section 4.1).

By summarizing raw Web-logging data such as use of navigation actions, page visits and re-visits, links followed and duration of visits, *first-order metrics* have been computed, such as:

- Path length
- Amount of re-visits
- Back button usage
- View time per page
- Fan degree
- Number of cycles
- Net density
- Average connected distance

Second-order metrics were computed as linear combinations of the first-order metrics by the aid of principal component analysis. They were completely specified (numerically) by first-order metrics. However, interpreting their meaning and labeling them was based on their correlations with user characteristics and task outcomes. The interpreted second-order metrics were labeled *navigation styles*. Two of these navigation styles are described below:

- **Flimsy Navigation** is a parsimonious navigation style. The navigation path was not very elaborate, most of the navigation taking place around the homepage. Time was spent more with processing content than with figuring out the hyperstructure that showed where the relevant information was. A high score on the flimsy navigation style was associated with low Internet expertise (r=-0.5), low working memory capacity (r=-0.38), and high perceived disorientation (r=0.46).
- **Laborious Navigation** involves intensive use of navigational infrastructure provided by the site. Users seemed to employ a trial and error strategy. They followed links just to see if they were useful or not. They figured out quite fast when paths were not leading towards their goal and returned. Re-visits were quite numerous but they were not redundant: once a page was re-

visited a different link was followed, it was just another trial. This navigation style was associated with high episodic memory (r=0.49) and low spatial ability (r=-0.40). This style indicates the type of re-visitation that does not relate to disorientation. The user needed to look around for a while until she/he had a good representation of the site structure, because she/he had a weak spatial ability. Yet, her/his memory prevented her/him from making redundant re-visits. This component shows how people compensate for the lack of spatial ability by effort and memory, and do not necessarily decrease performance (no correlation with task performance was found). It also shows why re-visitation is not always associated with disorientation. The term "laborious" should not suggest a correlation with effectiveness (task success). This style is effective in compensating (to some extent) for lack of spatial ability and avoiding a major decrease in performance and increase of perceived disorientation. But the style itself is not necessarily effective, it is not employed by highly effective users.

A *semantic metric*, called *Path adequacy*, was calculated based on navigation data and the task description that subjects were provided with at the beginning of a task. A navigation path was considered to be a concatenation of semantic objects that the user has encountered in her/his way. As semantic objects we have used page titles and link labels. Link labels were better than page titles in characterizing user's navigation path and computing semantic metrics, because they convey specific information. A navigation path was used in simulations of Web navigation as an indicator of contextual information involved in selecting specific navigation actions. *Path adequacy* was determined as a coefficient of semantic similarity between a navigation path and a task description. Semantic similarity was calculated with Latent Semantic Analysis (LSA). *Path adequacy* calculated at the end of a particular task was significantly correlated with *spatial ability* (r=0.36), and *task performance* (r=0.47). *Path adequacy* calculated at each step of a navigation session was used in simulations of Web navigation as a coherence criterion involved in selecting specific navigation actions.

Thus, we have shown that different types of knowledge about users can be inferred based on the kind of information that is extracted from Web-logging data: *syntactic* (structural) information indicated mainly users' navigation styles, for example, if they rather re-visit pages than viewing new pages, if they return using the back button or just by following links, etc. (first- and second-order metrics); and *semantic* information indicated if users were effective in pursuing their goals (path adequacy) independent of their navigation styles. We have argued for using navigation metrics in building adaptive Web applications, such as recommender systems (Juvina & Van Oostendorp, 2004).

6.1.1.3. Ways to conceive and deliver effective Web navigation support

We have suggested that a cognitive model of Web navigation can be used as generator of Web navigation support, particularly when the model is specified in computational terms and can be run as a computer program (Juvina & Van Oostendorp, 2005). In a series of empirical studies, several ways to deliver model-based navigation support have been tested.

Suggestions of goal-relevant links via voice (Section 4.4) have been shown to increase task performance. In addition, users with low spatial abilities had a higher performance increase than users with high spatial abilities. It seems that the offered navigation support prevented users from spending time and cognitive resources with those navigating actions that are not directly effective but are usually employed in order to accurately represent the information structure. Users engage in apparently useless navigation actions in order to get acquainted with the context of a particular piece of information, which is eventually useful in judging the value of that particular information. It follows that users with low spatial abilities are probably less able to represent the information space and this is why they benefit considerably when the cognitive model is doing this job for them. However, suggestions via voice were not well received from a subjective point of view, users found them annoying and manipulative.

Graphical suggestions in the form of small red arrows pointing at the relevant link (Juvina & Herder, 2005) were not only effective but also well received from a subjective point of view (Section 4.5). Men receiving support showed a decreased level of perceived disorientation as compared with men in the control condition, whereas such a difference was not found in women. Navigation support in the form of graphical link suggestions changed the structure of users' navigation behavior. In the support condition, participants used the back button less and the average connected distance in the navigation path was higher than in the control group. Thus, link suggestions caused the subjects to navigate in a more linear manner and reduced the amount of backtracking. High performers tend to take fewer suggestions than average and low performers. However, within each performance level, taking more suggestions is associated with increased task performance.

In the case of using the Web via screen readers, we tried to implement suggestions by changing the order of items on Web pages, in the sense that relevant items were placed upper so that they are read sooner in a sequence. This manipulation was not successful, most probably because changing the order of items on Web pages breaks the coherence established by the content authors (Section 5.1).

6.1.2. Contributions to the fields of Cognitive Science and Hypertext Comprehension

There are two main contributions of our research to these fields: amendments to existing models of Web navigation and applying well-established theories in new settings.

6.1.2.1. Amendments to existing models of Web navigation

CoLiDeS+, our proposed augmented version of CoLiDeS, has been shown to account for important aspects of user navigation behavior such as: considering contextual information when judging goal-relevance and employing navigation strategies (Section 4.3). This was done by including the user's navigation path in the model and allowing the model to backtrack and reconsider past selections (e.g., next-best strategy). The model has been empirically tested for how well it simulates the actual user behavior and whether it is useful in generating Web navigation support. Although it does not simulate the user behavior particularly well, CoLiDeS+ was shown to perform better that its previous version (CoLiDeS). CoLiDeS+ was used to generate navigation support and this support had a positive impact on user behavior and task outcomes. A number of limitations of CoLiDeS+ have been identified, such as: low accuracy in simulating the users' behavior, caused mainly by its reliance on LSA to compute semantic similarities, and a high amount of hand-coding required for running simulations.

An ACT-R model of Web navigation has been presented in order to demonstrate the possibility of overcoming some of the CoLiDeS+ limitations. This model offered computational solutions to implement key features of Web navigation behavior as reported in the literature and found in our empirical research. Some of these features were shared with previous models – selections based on goal relevance (information scent) (SNIF-ACT and CoLiDeS); backtracking, threshold and opportunistic strategies (MESA); back coherence (CoLiDeS+) – others were implemented here for the first time – intertwining between conservative and explorative strategies, and post-valued recall.

6.1.2.2. Well-established theories applied in new settings

Well-established theories of text comprehension (Kintsch, 1998), memory and cognition (Anderson, 1983), and working memory

(Baddeley, 1986) have been used in our experimentation and model development processes. While using these theories to explain Web navigation behavior, specific aspects have been noticed. For example, coherence of the memory representation plays an essential role in reading comprehension (Van den Broek, Young et al., 1999) whereas in Web navigation goal relevancy is essential and coherence is secondary (Section 4.3).

In addition, in Web navigation a spatial representation of the information space is much more important than in the case of reading plain text. The fact that spatial ability was the most important determinant of Web navigation performance was an unexpected and non-intuitive result. Why would performance on a semantically void mental rotation task predict performance on a semantically intensive Web navigation task? We have shown that the correlation between spatial ability and Web task performance is a robust result:

- It was found in three successive studies.
- Possible confounders for this correlation – working memory capacity, episodic memory, Internet expertise, reading comprehension, reading speed, and cognitive style – have been checked for and proven to have no influence.
- The correlation between spatial ability and another type of computer task performance has been found non-significant.

Correlations between spatial ability and navigation metrics have helped us understand the behavioral mediators between spatial ability and Web task performance. Spatial ability is negatively correlated with metrics involving re-visitation (re-visits, back button, and fan degree). It seems that spatial ability helps users in figuring out the information space structure so less re-visitation is needed. Supposedly common cognitive processes are used to represent and operate on an information space and to mentally rotate objects in a three-dimensional space. Even in the case of VIPs' reading or Web use, spatial aspects are essential (Section 5.1). Presumably both sighted and VIPs represent information in terms of *what* (content) and *where* (location). For sighted users, information of the type *where* could literally mean visual location. For VIPs *where* information could refer to: temporal position of a particular information element in a sequence, category membership, etc.

Our findings showing the importance of spatial cognition in Web navigation (Chapters 4 and 5) can be added to the body of evidence of established theories of text comprehension, memory and cognition.

6.2. Conclusions

This section provides conclusive answers to the research questions stated in Chapter 1.

6.2.1. What are the most important factors determining success in Web-assisted tasks? How can one measure or estimate these factors in an automatic way?

Success in Web assisted tasks depends essentially on a combination of spatial cognition and domain knowledge. Spatial cognition is involved in representing the structure of the information space while domain knowledge is necessary for understanding and selecting relevant content. These factors are expressed in the user's behavior. We have shown that users' behavior can be automatically logged and various navigation metrics can be computed based on these logs. Metrics referring to the structure of user navigation were called syntactic metrics, whereas metrics referring to the visited content were called semantic metrics.

6.2.2. What are the explanatory cognitive mechanisms for the identified factors? How can one implement these mechanisms in a (computational) cognitive model?

Users build and update a mental representation of the information space being navigated. This representation has a spatial character (in the sense that positions and distances are relevant, but not in the sense that it is visual or three-dimensional), and it is relatively independent of the contents (semantics) being represented. Users make assessments of relevance and decisions to select particular contents based on:

- prior knowledge they have about those contents, and
- knowledge they gain from the local context of those particular contents (i.e., what contents they link to).

We have proposed a cognitive model (CoLiDeS+) in which assessments of relevance are made based on both

- prior knowledge, modeled by an LSA semantic space, and
- contextual information, modeled by users' past selections.

We have also shown in an ACT-R demonstrative model how a developing memory representation can be used to mentally traverse (backward or forward) an information space.

6.2.3. What kind of Web navigation support can be conceived based on the knowledge gained from the previous questions? What impact has this support on users?

Navigation support should aim at preserving the information space structure and helping users traverse it in an efficient way. We do not favor extracting the relevant information from its original context as in the case of search engines. Instead, we have shown that emphasizing the relevant information in its original context helps users to discern between relevant and irrelevant information, and compensates for their deficient spatial abilities (when this is the case).

Providing contextual aids for the link-following behavior is not a new idea (Hardman, Bulterman, & Rossum, 1993). We have investigated it in the context of common use of the Internet under the assumption that the structure and contents of the information space are not known in advance but defined dynamically based on the user-system interaction. For instance, keeping the information in its original context is also used in the ScentTrails agent (Olston & Chi, 2003). ScentTrails starts by performing a search based on keywords entered by the user. Pages returned by the search engine are not immediately displayed to the user; instead, paths to these pages are suggested by means of highlighting. Thus, the user gets the chance to see the context and to collect additional information along the path. This approach assumes that the supporting agent knows the information space (e.g., a website's content and structure), so that a path can be identified from the search results to the user current position and relevancies of items to search results can be computed. In contrast, our approach assumes that the supporting agent is exploring the information space in the same manner as the user. Suggestions are made based on the goal relevance of links on the current page and their consistence with the user's past selections. It only needs to know the user's goal and to track the path followed by the user. These features allow the supporting agent to behave in an adaptive way. Our solution is less accurate than solutions based on an *a priori* knowledge of the system; however, it has less implementation constraints, a lower computational complexity and it is more realistic: users do switch websites in searching for their targets (Nielsen, 2006) and most information spaces are changing environments.

6.3. Discussion and future developments

It was one of our most important desiderata to do ecologically valid research; we have used real websites and realistic tasks as much as possible. While this impetus is not to be regretted, it must be admitted that this was the cause for results that were not always very compelling, at least in terms of statistical significance and magnitude of effects. Limitations of our studies in terms of sample size and accuracy of measurement prevented us from applying powerful statistical techniques such as Structural Equation Modeling. In addition, it is a matter of discussion how accurate and complete a cognitive model of Web navigation can be. We have opted for a trade-off between accuracy and practical relevance. CoLiDeS+ was conceived with the aim of building model-based navigation support. For this reason, some of the methodological criteria of cognitive modeling were relaxed. The simulation of user behavior was not complete. For example, the model did not have a mechanism to identify target content pages. Such a mechanism would have been extremely difficult to build and it was not really necessary for our intended use of the model. The model was meant to work alongside the user and suggest links that are relevant to a given user goal. The user was assumed to take those suggestions or not and stop when the target page was reached. Performance of cognitive models in the field of Web navigation depends on progress made in other fields such as machine learning and natural language processing. For cognitive modeling, a task domain as weakly structured and knowledge-intensive as the one proposed here is a great challenge. A cognitive model of Web navigation needs to handle natural language, large knowledge networks, and a great deal of sub-symbolic computations. These aspects are not part of the traditional work in cognitive modeling research (Gluck & Pew, 2005) but are becoming increasingly prominent in the cognitive modeling community (Pirolli, 2005). A foreseeable problem will be to handle the computational complexity required by up-scaling cognitive models to be included in adaptive Web applications as generators of navigation support.

The research presented here has not dealt with domain-specific applications such as intranets and specialized knowledge repositories. The target population was that of the common Web users performing everyday Web tasks, such as online shopping, requiring no or little domain knowledge. For this reason, the Semantic Web technology (Bocconi, Nack, & Hardman, 2005) was not applicable. Instead of particular ontologies corresponding to particular application domains, we have used semantic spaces corresponding to particular user populations. A semantic space contains all the concepts that a particular user population is assumed to have encountered, and it can be used to compute associations between concepts based on co-occurrence and dimensionality reduction. The advantage of semantic spaces over ontologies is a lower amount of hand-coding needed (they are just bags of words); the drawback is lower accuracy.

CoLiDeS+ (as well as other cognitive models) relies on the users' goals to be known. It was not our concern in this research how user goals are entered in an adaptive Web application or how such an application could infer them. In a previous paper (Juvina & Van Oostendorp, 2004) we suggest a way to infer pragmatic information (including users' goals) based on users' navigation history. This would be an interesting direction for future research. Another point to be addressed in future research is the dynamic characteristic of the user's memory representation (including the goal representation); so far in our models the goal remains unchanged during a navigation session, and this is a questionable assumption.

References

Ahuja, J. S., & Webster, J. (2001). Perceived disorientation: an examination of a new measure to assess Web design effectiveness. *Interacting with Computers*, 14(1), 15-29.

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, Massachusetts, Harvard University Press.

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review,* 111(4), 1036-1060.

Baddeley, A. D. (1986). Working Memory: the interface between memory and cognition. *Journal of Cognitive Neuroscience*, (4), 281-288.

Baddeley, A. D. (2000). The Episodic Buffer: a New Component of Working Memory? *Trends in Cognitive Sciences*, 4(11), 417-423.

Bernard, M., & Hamblin, C. (2003). Cascading versus Indexed Menu Design. *Usability News*, 5(1), retrieved July 16th, 2004 from psychology.wichita.edu/surl/usabilitynews.

Blackmon, M. H., Polson, P.G., Kitajima, M., Lewis, C. (2002). Cognitive Walkthrough for the Web. *Proceedings CHI 2002*, Minneapolis, Minnesota, USA.: ACM Press.

Bocconi, S., Nack, F., & Hardman, L. (2005). Supporting the Generation of Argument Structure within Video Sequences. *The sixteenth ACM Conference on Hypertext and Hypermedia*, 75-84.

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

Bothell, D. (2005). ACT-R 6 Official Release. *Twelfth Annual ACT-R Workshop*, Trieste, Italy: Universita degli Studi di Trieste.

Brumby, D. P. (2004). A model of single-page Web search: the effect of interdependence on link assessment. *6th International Conference on Cognitive Modeling*, Pittsburgh, PA.

Brumby, D. P. & Howes, A. (2003). Adaptive decision making in menu search: The role of interdependence and past experience on link selection. *Proceedings of HCI 2003*: Designing for Society, Bath, UK.

Brumby, D. P. & Howes, A. (2004). Good enough but I'll just check: web-page search as attentional refocusing. *6th Internal Conference on Cognitive Modeling*, Pittsburgh, PA.

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

Brusilovsky, P. & Maybury, M. T. (2002). From Adaptive Hypermedia to the Adaptive Web. *Communications of the ACM*, 45(5), 31-33.

Budiu, R. & Anderson, J. R. (2004). Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science,* 28, 1-44.

Byrne, M. D. (2001). ACT-R/PM and menu selection: applying a cognitive architecture to HCI. *Int. J. Human-Computer Studies,* 55, 41-84.

Byrne, M. D. (2003). Cognitive Architecture. In J. Jacko. & A. Sears. (Eds.). *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications.* (pp. 97-117) Mahwah, NJ, Lawrence Erlbaum.

Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled Web we wove: A taskonomy of WWW use. *Human Factors in Computing Systems: Proceedings of CHI 99*, Reading, MA, Addison Wesley.

Card, S. K., Pirolli, P. Van Der Wege, M., Morrison, J.B., Reeder, R.W., Schraedley, P.K., & Boshart, J. (2001). Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability. *SIGCHI'01*, Seattle, WA, USA., ACM.

Catledge, L. D. & Pitkow, J.E. (2001). Characterising Browsing Strategies in the World- Wide Web. *Computer Networks and ISDN Systems,* 27(6), 1065-1073.

Chen, C. (2000). Individual Differences in a Spatial-Semantic Virtual Environment. *Journal of The American Society for Information Science,* 51(6), 529-542.

Chen, C. & Czerwinski, M. (1998). From Latent Semantics to Spatial Hypertext - An Integrated Approach. *HyperText 98*, Pittsburgh, PA, USA, ACM.

Chen, S. Y., & Macredie, R.D. (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.

Choo, C. W., Detlor, B. & Turnbull, D. (2000). Information seeking on the Web: An integrated model of browsing and searching", First Monday, 5(2), 2000. Available at http://firstmonday.org/issues/issue5_2/choo/index.html

Cockburn, A. & McKenzie, B. (2001). What Do Web Users Do? An Empirical Analysis of Web Use. *Int. J. Human-Computer Studies,* 54(6), 903-922.

Cockburn, A. & McKenzie, B. (2002). Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments. *CHI 2002*, Minneapolis, Minnesota, USA, ACM.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin,* 112(1), 155-159.

Compeau, D. R., & Higgins, C.A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quart,* 19(2), 189.

Conklin, J. (1987). Hypertext: An introduction and survey. *IEEE Computer Magazine,* 20(9), 17-41.

Czerwinski, M., van Dantzich, M., Robertson, G.G., & Hoffman, H. (1999). The contribution of thumbnail image, mouse-over text and spatial location memory to Web page retrieval in 3D. *Interact'99,163-170,* Edinburgh, Scotland: IOS press.

Danielson, D. R. (2002). Web navigation and the behavioral effects of constantly visible maps. *Interacting with Computers,* 14, 601-618.

Di Blas, N., Paolini, P., & Speroni, M. (2004). "Usable accessibility" to the Web for blind Users. *Proceedings UI4ALL*, Vienna.

Edwards, D.M. & Hardman, L. (1988). Lost in Hyperspace: Cognitive Mapping and Navigation in a Hypertext Environment. In. R.McAleese (Ed.) *Hypertext Theory in Practice* (pp 105-125). Aberdeen, UK.

Egger, F. N. (2003). From Interactions to Tranzactions: Designing the Trust Experience for Business-to-Consumer Electronic Commerce. Eindhoven, Technische Universiteit Eindhoven.

Eveland Jr., W. P. & Dunwoody, S. (1998). Users and navigation patterns of a science World Wide Web site for the public. *Public Understand. Sci*.,(7), 285-311.

Farris, J. S., Jones, K. S., & Elgin, P.D. (2002). Users' schemata of hypermedia: what is so 'spatial' about a website? *Interacting with Computers,* 14(5), 487-502.

Gerber, E., & Kirchner, C. (2001). Who's surfing? Internet access and computer use by visually impaired youths and adults. *Journal of Visual Impairment & Blindness*, 95, 176-181.

Goldman, S. R., Golden, R. M., & Van den Broek, P. (2006). Why are Computational Models of Text Comprehension Useful? in F. Schmalhofer & C. A. Perfetti (Eds.) *Higher level language processes in the brain: Inference and comprehension processes.* Mahwah, NJ.: L. Erlbaum Associates.

Goumi, A., Rouet, J.-F., & Aubert, D. (2003). The effects of content representation on information access in an instructional multimedia database. *EARLI*, Padova, Italy.

Gugerty, L., Treadaway, C., & Rubinstein, J.S. (2006). Individual Differences in Internet Search Outcomes and Processes. *CHI 2006*, Montreal, Quebec, Canada, ACM.

Hardman, L., Bulterman, D., & van Rossum, G. (1993). Links in Hypermedia: the requirement for context. Hypertext'93 Proceedings (pp. 183-191) Seattle, Washington, United States: ACM.

Herder, E. (2005). Characterizations of User Web Re-visit Behavior. *Proc. of Workshop on Adaptivity and User Modeling in Interactive Systems ABIS 2005*, Saarbrücken, Germany.

Herder, E. (2006). Forward, Back and Home Again: Analyzing User Behavior on the Web. Doctoral Dissertation. Twente University, The Netherlands.

Herder, E. & Juvina, I. (2004). Discovery of Individual User Navigation Styles. *Adaptive Hypermedia AH2004*, Eindhoven.
Hilbert, D. M. & Redmiles, D. F. (2000). Extracting Usability Information from User Interface Events. *ACM Computing Surveys,* 32(4), 384-421.
Hofman, R. & Van Oostendorp, H. (1999). Cognitive effects of a structural overview in a hypertext. *British Journal of Educational Technology,* 30(2), 129-141.
Hornbaek, K., Bederson, B. B., & Plaisant, C. (2002). Navigation Patterns and Usability of Zoomable User Interfaces with and without an Overview. *ACM Transactions on Computer-Human Interaction,* 9(4), 362-389.
Hornof, A. J., & Halverson, T. (2003). Cognitive Strategies and Eye Movements for Searching Hierarchical Computer Displays. *CHI2003*, Ft. Lauderdale, Florida, USA., ACM.
Howes, A., Payne, S. J., & Richardson, J. (2002). An instance-based model of the effect of previous choices on the control of interactive search. *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*, Fairfax, Virginia.
Huang, T.-M., Kecman, V., & Kopriva, I. (2006). Kernel Based Algorithms for Mining Huge Data Sets, Supervised, Semi-supervised, and Unsupervised Learning. Berlin, Heidelberg, Springer-Verlag.
Instone, K. (2002). An Open Discussion on Web Navigation, retrieved November 13th , 2002 from http://user-experience.org/uefiles/presentations//KEI-WebNav-20020916.pdf
ISO (1991). ISO 9241 - 11 Ergonomic Requirements for office Work with VDT's - Guidance on Usability.
Ivory, M. Y., Sinha, R.R., & Hearst, M.A. (2001). Empirically Validated Web Page Design Metrics. *ACM SIGCHI'01*, Seattle, WA, ACM Press.
Ivory, M., & Hearst, M. (2001). State of the Art in Automating Usability Evaluation of User Interfaces. *ACM Computing Surveys,* 33(4), 1-47.
Jameson, A. (2003). Adaptive Interfaces and Agents. In J. Jacko & A. Sears (Eds.). *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*. (pp. 305-330). Mahwah, NJ, Lawrence Erlbaum Associates.
Jones, K.S., Farris, J.S., Elgin, P.D., Anders, B.A., & Johnson, B.R. (2002). Examining Information Searching on the World Wide Web with a Screen Reader: a Verbal Protocol Analysis. *Proceedings of Human Factors and Ergonomics Society 46th Annual Meeting*. Pittsburgh, PA, USA.
Juvina, I. & Herder, E. (2005). The Impact of Link Suggestions on User Navigation and User Perception. *UM2005 User Modeling:*

Proceedings of the Tenth International Conference, Edinburgh, Springer-Verlag.

Juvina, I. & van Oostendorp, H. (2003). Human Factors in Web-assisted Personal Finance. *Human-Centred Computing. Cognitive, Social and Ergonomic Aspects. HCI International 2003 - Conference Proceedings*. Crete, Greece. D. Harris, Duffy, V., Smith, M., Stephanidis, C. Mahwah, NJ., Lawrence Erlbaum Associates, Publishers, 3, 477-481.

Juvina, I. & van Oostendorp, H. (2004). Predicting user preferences - from semantic to pragmatic metrics of Web navigation behavior. *SIGCHI-NL*, Amsterdam, ACM Digital Library.

Juvina, I. & van Oostendorp, H. (2005). Cognitive Model Working Alongside the User. *British HCI2005*, Edinburgh, UK.

Juvina, I., & van Oostendorp, H. (2006). Individual differences and behavioral metrics involved in modeling Web navigation. *Universal Access in Information Society*, Springer.

Juvina, I., Trausan-Matu, S., Iosif, Gh., Van der Veer, G., Marhan, A-M., & Chisalita, C. (2002). Analysis of Web Browsing Behavior - a great potential for psychological research. *TAMODIA 2002 - Task Models and Diagrams for User Interface Design*, Bucharest, INFOREC Publishing House.

Kickmeier, M. & Albert, D. (2002). Usability of Web-based navigation systems: The traditional Tree Browser and the alternative Star Tree. *EARLI SIG on Multimedia Comprehension*, Poitiers, France.

Kickmeier, M. & Albert, D. (2003). The effects of scanable items on information retrieval from computer screens. *EARLI*, Padova, Italy.

Kim, H. R., & Chan, P. K. (2003). Learning Implicit User Interest Hierarchy for Context in Personalization. *IUI03*. Miami, Florida, USA.

Kintsch, W. (1998). Comprehension: A Paradigm for Cognition. Cambridge University Press.

Kintsch, W. (2005). An Overview of Top-Down and Bottom-Up Effects in Comprehension: The CI Perspective. *Discourse Processes,* 39(2&3), 125-129.

Kitajima, M. & Polson, P. G. (1997). A comprehension-based model of exploration. *Human-Computer Interaction,* 12, 345-389.

Kitajima, M., Blackmon, M.H., & Polson, P.G. (2000). A Comprehension-based Model of Web Navigation and Its Application to Web Usability Analysis. *People and Computers XIV*, 357-373.

Krug, S. (2000). Don’t Make Me Think! A Common Sense Approach to Web Usability, Indianapolis, Indiana: New Riders Publishing.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.

Landauer, T. K., Foltz, P.W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes,* 25, 259-284.

Larson, K. & Czerwinski, M. (1998). Web Page Design: Implications of Memory, Structure and Scent for Information Retrieval. *Proceedings of CHI98*.

Lazar, J. (2003). The World Wide Web. In J. Jacko and A. Sears (Eds.). *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*. (pp. 714-730) Mahwah, NJ, Lawrence Erlbaum Associates.

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. *Journal of Educational Computing Research*, 20(2), 189-206.

McDonald, S., & Stevenson, R. J. (1999). Spatial versus conceptual maps as learning tools in hypertext. *Journal of Educational Multimedia and Hypermedia, 8*(1), 43-64.

Miller, C. S. & Fuchs, S. (2005). Simulating individual differences in Web Navigation. *Paper presented at the 15th annual meeting of the Society for Text and Discourse*, Amsterdam, Netherlands.

Miller, C. S. & Remington, R. W. (2001). Modeling an opportunistic strategy for information navigation. *Proceedings of the Twenty-Third Conference of the Cognitive Science Society*, Edinburgh, Scotland. Lawrence Erlbaum Associates.

Miller, C. S., & Remington, R. W. (2004). Modelling Information Navigation: Implications for Information Architecture. *Human-Computer Interaction,* 19(3), 225-271.

Miron, E. A. M. (2003). Analyzing the role of students' goal orientation during their interaction with an ILE. *HCT2003 - The 7th Human Centred Technology Postgraduate Workshop*, Falmer, Brighton, UK., HCT.

Montanero, M., Lucero, M., & Mendez, J.M. (2004). Influence of rhetorical structure revision vs. graphic support in learning from text. *EARLI SIG2 Meetings on "Comprehension of Text and Graphics: Basic and Applied Issues"*, Valencia, Spain, Universitat de Valencia.

Morrison, J. B., Pirolli, P., & Card, S.K. (2001). A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions. *CHI 2001*, Seattle.

Neerincx, M. A., Pemberton, S., & Lindenberg, J. (1999). U-WISH Web usability: methods, guidelines and support interfaces. Soesterberg, TNO Human Factors Research Institute.

Neerincx, M.A., Lindenberg, J. & Pemberton, S. (2001). Support concepts for Web navigation: a cognitive engineering approach*.*

Proceedings Tenth World Wide Web Conference, pp. 119-128. New York, NY: ACM Press.

Neerincx, M.A., Lindenberg, J., Rypkema, J.A. & Van Besouw, N.J.P. (2000). A practical cognitive theory of Web-navigation: Explaining age-related performance differences*. Position paper Workshop "Basic Research Symposium"* CHI2000. (Paper was also included in Seminar Proceedings "Towards a user-friendly Web", Amsterdam: www.internetusability.com).

Nielsen, J. (1989). The matters that really matter for hypertext usability. *Hypertext'89 Proceedings*.

Nielsen, J. (2006). Users Interleave Sites and Genres, *Jacob Nielsen's Alertbox*. 2006.

Nilsson, R. M., & Mayer, R.E. (2002). The effects of graphic organisers giving cues to the structure of a hypertext document on users' navigation strategies and performance. *Int. J. Human-Computer Studies,* 57, 1-26.

Olston, C. & Chi, E. H. (2003). ScentTrails: Integrating Browsing and Searching on the Web. *ACM Transactions on Computer-Human Interaction,* 10(3), 177–197.

Paganelli, L. & Paterno, F. (2002). Intelligent Analysis of User Interactions with Web Applications. *IUI 2002*, San Francisco, California, USA, ACM.

Paterno, F. (2002). Tools for Task Modelling:Where we are, Where we need to go. *TAMODIA*, Bucharest.

Pettijohn, T. (2002). Locus of Control Scale, accessed 12 Nov. 2002 from http://www.dushkin.com/connectext/psy/ch11/survey11.mhtml.

Pirolli, P. & Card, S. K. (1999). Information foraging. *Psychological Review,* 106(4), 643-675.

Pirolli, P., & Card, S. K. (1998). Information Foraging Models of Browsers for Very Large Document Spaces. *AVI'98*, L'Aquila, Italy, ACM Press.

Pirolli, P., Card, S. K. (1995). Information Foraging in Information Access Environments. *Proceedings CHI95*, ACM Press.

Pirolli, P. & Fu, W.-T. (2003). SNIF-ACT: A Model of Information Foraging on the World Wide Web. *Proceedings of the Ninth International Conference on User Modeling*, Johnstown, Pennsylvania, USA LNAI 2702, Springer-Verlag Berlin Heidelberg, 2003, pp. 45-54., Springer-Verlag Berlin Heidelberg.

Rada, R. (1991) *HYPERTEXT, From Text to Expertext*, McGraw-Hill.

Rauterberg, M. (1996). How to Measure Cognitive Complexity in Human-Computer Interaction. *Proceedings of the Thirteenth European Meeting on Cybernetics and Systems Research*, Vienna, Austria, Austrian Society for Cybernetic Studies.

Reason, J. (1990). *Human error*. Cambridge, Cambridge University Press.

Ritter, F. E., Baxter, G. D., Jones, G. & Young, R.M. (2000). Supporting Cognitive Models as Users. *ACM Transactions on Computer-Human Interaction,* 7(2), 141-173.

Roy, M. & Chi, M. T. H. (2003). Gender Differences in Patterns of Searching the Web. *Journal of Educational Computing Research*. 29(3), 335-348.

Salmerón, L., Cañas, J.J., Kintsch, W. & Fajardo, I. (2005). Reading strategies and hypertext comprehension. *Discourse Processes*, 40 (3), 171-191.

Sanchez, C. A., & Wiley, J. (2005). Working memory, spatial ability, and learning from the Web*. Paper presented at the 15th annual meeting of the Society for Text and Discourse*, Amsterdam, Netherlands.

Segal, R. B., & Kephart, J.O. (2000). Incremental learning in SwiftFile. *Machine Learning Conference*, San Francisco.

Sjolinder, M., (1998) Individual differences in spatial cognition and hypermedia navigation, in *Towards a Framework for Design and Evaluation of Navigation in Electronic Spaces*. Swedish Institute of Computer Science.

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

Stanney, K.M., & Salvendy, G., (1995). Information visualization; assisting low spatial individuals with information access tasks through the use of visual mediators. *Ergonomics*, 38(6), 1184-1198.

Sterrett, S. G. (2002). Physical Models and Fundamental Laws: Using One Piece of the World to TellAbout Another. *Mind and Society,* 3(5), 51-66.

Sumi (2002). The Software Usability Measurement Inventory. Accessed 14 April 2002 from http://www.ucc.ie/hfrg/questionnaires/sumi/.

Tabachnick, B. G. & Fidell, L. S. (2001). Using multivariate statistics. Boston: Allyn and Bacon.

Tamborello, F. P. I. & Byrne, M. D. (2005). Information Search: The Intersection of Visual and Semantic Space. *CHI 2005*, Portland, Oregon, USA., ACM.

Tauscher, L. & Greenberg, S. (1997). How people re-visit Web pages: empirical findings and implications for the design of history systems. *Int. J. Human-Computer Studies,* 47, 97-137.

Tavanti, M. & Lind, M. (2001). 2D vs. 3D, Implications on spatial memory. *Proceedings of IEEE's Information Visualization*. Washington DC, USA: IEEE Computer Society.

Trausan-Matu, S., Iosif, G., Juvina, I., & Marhan, A.-M. (2002). Task Modeling in Intelligent Tutoring Systems on the Web. *Proceedings of 1st International Workshop on Task models and Diagrams for user interface design TAMODIA'02*, Bucharest, INFOREC Publishing House.

Tucker, P., & Warr, P. (1996). Intelligence, elementary cognitive components, and cognitive styles as predictors of complex task performance. *Personality and Individual Differences,* 21(1), 91-102.

Van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating Memory-Based and Constructionist Processes in Accounts of Reading Comprehension. *Discourse Processes,* 39(2&3), 299-316.

Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The Landscape Model of Reading: Inferences and the On-line Construction of a Memory Representation. In H. van Oostendorp, & S.R. Goldman (Eds.). *The construction of mental representations during reading*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Van Hooijdonk, C., Maes, A., & Ummelen, N. (2006). 'I have been here before' - an investigation into spatial verbalizations in hypertext navigation. *Information Design Journal,* 14(1), 8-21.

Van Oostendorp, H. (2002). Satisfaction with a Website. *Multimedia Comprehension. Earli-SIG2 Meeting*, University of Poitiers, France.

Van Oostendorp, H. & Walbeehm, B. J. (1995). Towards modelling exploratory learning in the context of direct manipulation interfaces. *Interacting with Computers,* 7(1), 3-24.

Van Zwol, R. & Van Oostendorp, H. (2004). Google's "I'm feeling lucky", Truly a Gamble? *Web Information Systems - WISE 2004*, Proceedings of the 5th International Conference on Web Information Systems Engineering, Brisbane, Australia, Springer.

Velleman, E. & Snetselaar, H. (2000). Site Seeing. The development of an accessible Web site or Web based multimedia product for the sighted, partially sighted and the blind. Zeist, The Netherlands: Bartimeus Instituut.

Wammi (2002). Web Usability Questionnaire. Retrieved 18 March 2002 from http://www.wammi.com/

Wen, J. (2003). Post-Valued Recall Web Pages: User Disorientation Hits the Big Time. *IT&Society*, 1(3), 184-194.

Westerman, S.J. (1995). Computerized information retrieval: Individual Differences in the use of spatial Vs nonspatial navigational information. *Perceptual and Motor Skills*, 81, 771-786.

Young, R. M. (1998). Rational Analysis of Exploratory Choice. In M. Oaksford & N. Chater (Eds.). *Rational Models of Cognition.* Oxford: Oxford University Press.

Zampa, V., & Lemaire, B. (2002). Latent Semantic Analysis for User Modeling. *Journal of Intelligent Information Systems,* 18, 15-30.

Annex

A simplified version of the website How Stuff Works together with coefficients used by the ACT-R model presented in Section 5.2. The user intention is to find out how coffee works. The numbers presented on the first line below each option are goal relevancies – semantic similarities between the word "coffee" and each link label, respectively. Numbers appearing on the second line are called back-coherence1 coefficients – semantic similarities between the previous selection and each option, respectively. Back-coherence2 coefficients are shown on the third line and they are calculated as semantic similarities between a two steps back selection and each option, respectively. Back-coherence3 and back-coherence4 coefficients are calculated in an analog way.

Home-page

auto	computer	electronics	home	health
.002	.007	.008	.039	.088
science	money	travel	entertainment	shop
.069	.038	.000	.079	.142

Shop-page

cars	computers	electronics	garden	jewelry	games
.044	.007	.008	.125	.049	.008
.44	.20	.08	.33	.14	.29

Health-page

diseases	drugs	fitness	care	nutrition	pregnancy
.015	.030	.009	.056	.010	.011
.48	.47	.06	.59	.19	.09

Science-page

engineering	life	supernatural
.009	.043	.019
.44	.24	.02

Entertainment-page

arts	games	movies	music	sports	television	toys
.038	.008	.001	.005	.006	.004	.116
.11	.24	.36	.26	.29	.44	.19

Drugs-page

performance	enhancing	nicotine
.034	.018	.129
.02	.00	.03
.02	.04	.01

Care-page

diet	hangover
.036	.051
.12	.08
.27	.04

Toys-page

doodle	radio
.079	.027
.15	.02
.01	.22

Garden-page

décor	supplies	tools	irrigation
.052	.104	.069	.011
.05	.13	.01	.06
.09	.22	.23	.02

Tools-page

mowers	shears	clippers
.045	.004	.008
.06	.01	.06
.06	.03	.01
.00	.02	.02

Nicotine-page

brain	addictive	withdrawal	toxicity	cancer
.027	.064	.049	.044	.029
.04	.35	.12	.05	.00
.31	.02	.05	.01	.38
.20	.01	.02	.02	.41

Hangover-page

alcohol	nausea	aspirin
.082	.011	.052
.19	.06	.08
.13	.08	.05
.20	.05	.05

Doodle-page

pen	eraser
.015	.024
.05	.19
.03	.03
.00	.02

Decor-page

fountains	plaques	vanes
.024	.014	.006
.07	.01	.01
.04	.06	.03
.00	.01	.05

Supplies-page

pool	accessories	filters	pumps
.126	.067	.028	.039
.08	.21	.13	.21
.19	.04	.03	.04
.10	.21	.04	.05

Addictive-page

caffeine	marijuana
.147	.023
.24	.10
.28	.17
.09	.22
.05	.01

Alcohol-page

effects	warning	abuse
.003	.045	.029
.01	.01	.35
.04	.01	.09
.19	.10	.20
.29	.10	.18

Aspirin-page

headache	heart
.010	.014
.03	.03
.13	.01
.01	.43
.03	.19

Pool-page

Covers	heaters	filters	pumps
.024	.021	.028	.039
.00	.01	.01	.04
.24	.00	.13	.21
.14	.03	.03	.04
.18	.04	.04	.05

Accessories-page

fence	spa
.014	.019
.06	.03
.05	.04
.05	.03
.03	.07

Summary

The focus of the PhD project reported here was Web navigation. The term “Web navigation” was used in a broad sense referring to users’ orientation in an information space, locating information and progressing from one information source to another. The Web has brought not only the opportunity of a nonlinear access to information sources but also the challenges of *cognitive overload* and *disorientation* (Conklin, 1987; Edwards & Hardman, 1988). Involving spatial features (syntax) in processing contents (semantics) was recognized as a distinguishing characteristic of Web navigation (Di Blas, Paolini, & Speroni, 2004).

As stated in *Chapter 1* – Introduction, the objective of the research presented here was to build a cognitive model that predicts and explains human performance in Web-assisted tasks. The research was driven by the following questions:

- What are the most important factors determining success in Web-assisted tasks?
    - How can one measure or estimate these factors in an automated way?
- What are the explanatory cognitive mechanisms for the identified factors?
    - How can one implement these mechanisms in a (computational) cognitive model?
- What kind of Web navigation support can be conceived based on the knowledge gained from the two previous questions?
    - What impact has this support on users?

Chapter 2 presented the main behavioral and cognitive perspectives on Web navigation as reported in the literature of the field. *Chapter 3* discussed methodological issues involved throughout this research project: attention was paid simultaneously to theory, method and real-world applicability; Web navigation was grounded in the theories of Cognitive Science, in particular Text Comprehension, and Information Science, in particular Human-Computer Interaction; methodological paradigms of experimentation, statistics and modeling have been applied; the whole research was placed as much as possible in realistic settings and practical needs of Web engineering and design has always guided research decisions.

This research has documented how real Web applications are used (*Chapter 4*). A large number of users have been invited to the Usability Lab, where they have performed Web tasks of various sorts from several domains (Juvina & Van Oostendorp, 2003). Their behavior was recorded by various means: Web-logging, screen captures, video, audio,

and paper-and-pencil. A large amount of data has been analyzed and interpreted in order to discover the regularities that can inform scientists and practitioners. The most important findings in this category are:

- Using the Web can be seen as a dialogue: users inform Web applications about their choices and Web applications "reply" with content. Thus, interaction paradigms such as reading/writing and talking/listening can be applied in understanding Web navigation behavior.
- In order to adequately characterize Web navigation behavior, a complete set of criteria is needed, including objective (performance), subjective (satisfaction) and undesirable aspects of task execution (disorientation).
- Users re-visit pages not only because they forget what they have seen before, but also to get acquainted with the structure of the information space, which in turn helps them in preventing disorientation.
- Using the Web via a screen reader by visually impaired persons is more taxing from a cognitive point of view, and users are more vulnerable to dissatisfaction and disorientation if they have to visit a large number of pages to reach their goal.

Observational and empirical research has revealed the most important user-related factors that determine performance in Web-assisted tasks (*Section 4.2*). A large number of factors have been analyzed together in relation to a comprehensive range of outcomes of Web navigation tasks. A limited number of significant predictors were identified, and their relative contribution to explaining task outcomes was estimated. Since factors were studied together and the stepwise method of regression analysis was employed, it was possible to rule out factors that were only marginally significant or confounded with one another. A sequence of repeated studies have shown that a combination of two factors is the most important determinant of human performance in Web-assisted tasks: a structure-related factor (spatial ability) and a content-related factor (domain expertise). Spatial cognition is involved in representing the structure of the information space while domain knowledge is necessary for understanding and selecting relevant content.

Factors such as spatial ability can be measured only with specialized tests, which cannot be implemented in realistic Web applications. For this reason, using Web-logging data to calculate metrics of Web navigation behavior has been proposed (*Section 4.1.2*). Metrics referring to the structure of user navigation were called syntactic metrics, whereas metrics referring to the visited content were called semantic metrics. By summarizing raw Web-logging data, such as use of navigation actions, page visits and re-visits, links followed and duration of visits, *first-order metrics* have been computed, such as:

- Path length
- Amount of re-visits
- Back button usage
- View time per page
- Fan degree
- Number of cycles
- Net density
- Average connected distance

Second-order metrics were computed as linear combinations of the first-order metrics by the aid of principal component analysis. They were completely specified (numerically) by first-order metrics. However, interpreting their meaning and labeling them was based on their correlations with user characteristics and task outcomes. The interpreted second-order metrics were labeled *navigation styles*. Two of these navigation styles are described below:

- **Flimsy Navigation** is a parsimonious navigation style. The navigation path was not very elaborate with most of the navigation taking place around the homepage. Time was spent more with processing content than with figuring out the hyperstructure that showed where the relevant information was.
- **Laborious Navigation** involves intensive use of navigational infrastructure provided by the site. Users seemed to employ a trial and error strategy. They followed links just to see if they were useful or not. They figured out quite fast when paths were not leading towards their goal and returned. Re-visits were quite numerous but they were not redundant: once a page was re-visited a different link was followed, it was just another trial.

A *semantic metric* called *Path adequacy* was calculated based on navigation data and the task description that subjects were provided with at the beginning of a task. A navigation path was considered to be a concatenation of semantic objects that the user has encountered in her/his way. Page titles and link labels have been used as semantic objects. Link labels were better than page titles in characterizing the user's navigation path and computing semantic metrics because they convey more specific information. A navigation path was used in simulations of Web navigation as an indicator of contextual information involved in selecting specific navigation actions. Path adequacy was determined as a coefficient of semantic similarity between a navigation path and a task description. Semantic similarity was calculated with Latent Semantic Analysis (LSA). Path adequacy calculated at the end of a particular task was significantly correlated with *spatial ability* ($r=0.36$), and *task performance* ($r=0.47$). Path adequacy calculated at each step of a navigation session was used in simulations of Web navigation as a coherence criterion involved in selecting specific navigation actions.

Thus, different types of knowledge about users can be inferred based on the kind of information that is extracted from Web-logging data: *syntactic* (structural) information indicated mainly users' navigation styles, for example, if they re-visit pages rather than viewing new pages, if they return using the back button or just by following links, etc.; and *semantic* information indicated if users were effective in pursuing their goals independent of their navigation styles. These navigation metrics can be used in building adaptive Web applications, such as recommender systems (Juvina & Van Oostendorp, 2004).

Well-established theories of text comprehension (Kintsch, 1998), memory and cognition (Anderson, 1983), and working memory (Baddeley, 1986) have been used in the model development process. While using these theories to explain Web navigation behavior, specific aspects have been noticed. For example, in Web navigation a spatial representation of the information space is much more important than in the case of reading plain text. Users build and update a mental representation of the information space being navigated. This representation has a spatial character (in the sense that positions and distances are relevant, but not in the sense that is visual or three-dimensional), and it is relatively independent of the contents (semantics) being represented. Users make assessments of relevance and decisions to select particular contents based on both prior knowledge they have about those contents, and knowledge they gain from the local context of those particular contents (i.e., what contents they link to). Even in the case of using the Web by visually impaired persons (VIPs) spatial aspects are essential. For VIPs spatial aspects of Web use could refer to: temporal position of a particular information element in a sequence, category membership, etc.

A cognitive model (labeled CoLiDeS+) has been presented (*Section 4.3*) in which assessments of relevance are made based on both prior knowledge, modeled by an LSA semantic space (on the basis of the concept 'information scent'), and contextual information, modeled by users' past selections (on the basis of the concept 'path adequacy'). CoLiDeS+, an augmented version of CoLiDeS (Kitajima, Blackmon, & Polson, 2000), has been shown to account for important aspects of user navigation behavior such as: considering contextual information when judging goal-relevance and employing navigation strategies. This was accomplished by including the user's navigation path in the model and allowing the model to backtrack and reconsider past selections (e.g., next-best strategy). The model has been empirically tested for how well it simulates the actual user behavior and whether it is useful in generating Web navigation support. Although it does not simulate the user behavior particularly well, CoLiDeS+ was shown to perform better that its previous version (CoLiDeS). CoLiDeS+ was used to generate navigation support and this support had a positive impact on user behavior and task outcomes. A number of limitations of CoLiDeS+ have been identified, such as: low accuracy in simulating the users' behavior, caused mainly be its reliance on LSA to compute semantic similarities, and a high amount of hand-coding required for running simulations.

An ACT-R model of Web navigation has been presented (*Section 5.2*) in order to demonstrate the possibility of overcoming some of the CoLiDeS+ limitations. This model offered computational solutions to implement key features of Web navigation behavior as reported in literature and found in our empirical research. Some of these features were shared with previous models – selections based on goal relevance ('information scent') (SNIF-ACT and CoLiDeS); backtracking, threshold values and opportunistic strategies (MESA); and back coherence ('path adequacy, CoLiDeS+). Other aspects were implemented here for the first time – the intertwining between conservative and explorative strategies, and 'post-valued recall'.

It has been suggested that a cognitive model of Web navigation can be used as generator of Web navigation support, particularly when the model is specified in computational terms and can be run as a computer program (Juvina & Van Oostendorp, 2005). In a series of empirical studies, several ways to deliver model-based navigation support have been tested (*Sections 4.4, 4.5, and 5.1*). Suggestions of goal-relevant links via voice have been shown to increase task performance. In addition, users with low spatial abilities had a higher performance increase than users with high spatial abilities. It seems that the offered navigation support prevented users from spending time and cognitive resources on those navigating actions that are not directly effective but are usually employed in order to accurately represent the information structure. Users engage in apparently useless navigation actions in order to get acquainted with the context of a particular piece of information, which is eventually useful in judging the value of that particular information. It follows that users with low spatial abilities are probably less able to represent the information space and this is why they benefit considerably when the cognitive model is doing this job for them. However, suggestions via voice were not well received from a subjective point of view, users found them annoying and manipulative.

Graphical suggestions in the form of small red arrows pointing at the relevant link (Juvina & Herder, 2005) were not only effective but also well received from a subjective point of view. Men receiving support showed a decreased level of perceived disorientation as compared with men in the control condition, whereas such a difference was not found in women. Navigation support in the form of graphical link suggestions changed the structure of users' navigation behavior. In the support condition, participants used the back button less and the average connected distance in the navigation path was higher than in the control group. Thus, link suggestions caused the subjects to navigate in a more linear manner and reduced the amount of backtracking. High performers tend to take fewer suggestions than average and low performers. However, within each performance level, taking more suggestions is associated with increased task performance.

In the case of using the Web via screen readers, suggestions were implemented by changing the order of items on Web pages, in the sense that relevant items ('hyperlinks') were placed higher so that they are read sooner in a sequence of links. This manipulation was not successful, most probably because changing the order of items on Web pages breaks the coherence established by the content authors.

This research does not support extracting the relevant information from its original context as in the case of search engines. Instead, it has been shown that emphasizing the relevant information in its original context helps users discern between relevant and irrelevant information, and compensates for their deficient spatial abilities (when this was the case for the user).

Performing ecologically valid research was an important desideratum of this project; real websites and realistic tasks have been used as much as possible (*Chapter 6*). A trade-off between modeling accuracy and practical relevance has guided the research. CoLiDeS+ was conceived with the aim of building model-based navigation support. For this reason, some of the methodological criteria of cognitive modeling were relaxed. The simulation of user behavior was not complete. For example, the model did not have a mechanism to identify target content pages. Such a mechanism would have been extremely difficult to build and it was not necessary for the intended use of the model. The model was meant to work alongside the user and suggest links that are relevant to a given user goal. The user was assumed to take those suggestions or not and stop when the target page was reached. Performance of cognitive models in the field of Web navigation depends on progress made in other fields such as machine learning and natural language processing. For cognitive modeling, working in a task domain as weakly structured and knowledge-intensive as the one proposed here was a great challenge. A cognitive model of Web navigation needs to handle natural language, large knowledge networks, and a great deal of sub-symbolic computations. These aspects are not part of the traditional work in cognitive modeling research (Gluck & Pew, 2005) but are becoming increasingly prominent in the cognitive modeling community (Pirolli, 2005). A foreseeable issue for future research will be handling the computational complexity required by up-scaling cognitive models to be included in adaptive Web applications as generators of navigation support.

Samenvatting

De focus van dit proefschrift is Webnavigatie. De term "Webnavigatie" wordt in algemene zin gebruikt en verwijst naar de gebruikersoriëntatie in een informatieruimte, het lokaliseren van informatie en het zich voortbewegen vanuit een bepaalde informatiebron of plek naar een andere. Het Web heeft niet alleen de mogelijkheid geboden om op een niet-lineaire manier toegang te krijgen tot informatiebronnen, maar het heeft ook geleid tot problemen op het gebied van *cognitieve overbelasting* en *desoriëntatie* (Conklin, 1987; Edwards & Hardman, 1988). Het gebruik van ruimtelijke kenmerken (syntax) bij het verwerken van inhoud (semantiek) is onderkend als een belangrijke karakteristiek van Webnavigatie (Di Blas, Paolini, & Speroni, 2004).

In *Hoofdstuk 1 - Introductie* wordt het doel van het onderzoek gepresenteerd: het ontwikkelen van een cognitief model waarmee de prestaties van mensen die een ‘webgeassisteerde taak’ uitvoeren kunnen worden voorspeld en verklaard. Hierbij staan de volgende onderzoeksvragen centraal:

- Wat zijn de meest belangrijke factoren die het succes bepalen bij webgeassisteerde taken?
    - Hoe kunnen deze factoren geautomatiseerd worden gemeten of bij benadering worden bepaald?
- Hoe kunnen deze factoren begrepen worden in termen van cognitieve mechanismen?
    - Hoe kunnen deze mechanismen geïmplementeerd worden in een computationeel cognitief model?
- Wat voor ondersteuning voor Webnavigatie kan ontwikkeld worden op basis van de kennis die bij het beantwoorden van de vorige twee vragen wordt opgedaan?
    - Wat is hierbij de impact op de ondersteuning van gebruikers?

Hoofdstuk 2 geeft een literatuuroverzicht van de belangrijkste ideeën over gedrags- en cognitieaspecten bij Webnavigatie. In *Hoofdstuk 3* komen de methodologische aspecten van dit onderzoeksproject aan bod: er is hierbij zowel aandacht geschonken aan theorievorming als aan methodiek en praktische toepasbaarheid. Webnavigatie is gefundeerd in de theorievorming binnen de Cognitie wetenschap (met name over tekstbegrip) en de Informatiekunde (met name mens-machine interactie). De methodologische paradigma's van het experiment, statistiek en modelvorming zijn toegepast. Het onderzoek als geheel is zoveel mogelijk gepositioneerd in een realistische omgeving en bij onderzoeksbeslissingen hebben praktische behoeften van Web ontwikkeling en ontwerp steeds als leidraad gediend.

In *Hoofdstuk 4* is beschreven hoe bestaande Web applicaties worden gebruikt. Een groot aantal gebruikers zijn getest in een Usability Lab, waar zij verschillende Web taken uit verschillende domeinen hebben uitgevoerd (Juvina & Van Oostendorp, 2003). Het gedrag van deze gebruikers is op een aantal manieren geregistreerd middels Web-logging, screen captures, video, audio en op papier. Een grote hoeveelheid data is geanalyseerd en geïnterpreteerd om regelmatigheden te ontdekken die van belang kunnen zijn voor zowel wetenschappers als mensen uit de praktijk. De meest belangrijke bevindingen in deze categorie zijn:

- Het gebruik van het Web kan beschouwd worden als een dialoog: gebruikers informeren Web applicaties over hun keuzes en Web applicaties "beantwoorden" met het verstrekken van bepaalde informatie. Interactie paradigma's zoals lezen/schrijven en praten/luisteren kunnen dus toegepast worden bij het begrijpen van Webnavigatie gedrag.
- Om Webnavigatie gedrag adequaat te kunnen karakteriseren is een volledige set van criteria noodzakelijk, waarin ruimte is voor objectieve aspecten (prestaties), subjectieve aspecten (tevredenheid van gebruikers) en ongewenste aspecten bij het uitvoeren van taken (desoriëntatie).
- Gebruikers komen niet zozeer terug bij eerder bezochte pagina's omdat ze zijn vergeten wat ze eerder hebben gezien, maar vooral om bekend te raken met de informatieruimte. Dit helpt gebruikers om desoriëntatie te voorkomen.
- Het gebruik van het Web via zg. beeldscherm lezers ('screen readers') door mensen met een visuele beperking is belastend vanuit cognitief standpunt. Deze gebruikers neigen dan meer tot ontevredenheid en desoriëntatie als ze een groot aantal pagina's moeten bezoeken om hun doel te bereiken.

Met behulp van observatie en empirisch onderzoek zijn de meest belangrijke gebruikersgerelateerde factoren geïdentificeerd die de prestaties op webgeassisteerde taken bepalen (*sectie 4.2*). Een groot aantal factoren is geanalyseerd in samenhang met een uitgebreide reeks van uitkomsten op

Webnavigatie taken. Een beperkt aantal significante voorspellers is hierbij geïdentificeerd en hun relatieve bijdrage bepaald. Aangezien alle factoren tegelijkertijd zijn bestudeerd met behulp van multipele lineaire regressie analyse (via de methode ‘stepwise’), is het mogelijk om factoren buiten beschouwing te laten die slechts marginaal significant waren of sterk met elkaar samenhingen. Een serie experimenten heeft aangetoond dat twee factoren de meest belangrijke voorspellers zijn van de gebruikersprestaties bij webgeassisteerde taken: een structuur gerelateerde factor (ruimtelijk inzicht) en een inhoudelijk gerelateerde factor (domein kennis). Ruimtelijk inzicht is van belang bij het representeren van de structuur van de informatieruimte en domein kennis is nodig om de relevante onderwerpen te selecteren.

Factoren zoals ruimtelijke vaardigheden kunnen alleen met gespecialiseerde testen worden gemeten en deze kunnen niet worden geïmplementeerd in realistische Web applicaties. Om deze reden is voorgesteld om Web-logging data te gebruiken om maten te berekenen voor Webnavigatie gedrag (*sectie 4.1.2*). Maten die betrekking hebben op de structuur van gebruikersnavigatie worden *syntactische* maten genoemd en maten die betrekking hebben op de bezochte onderwerpen worden *semantische* maten genoemd.
Ruwe Web-logging data zijn gebaseerd op het aantal navigatie acties van gebruikers, het aantal bezochte en opnieuw bezochte pagina, het aantal gevolgde links en de tijdsduur waarmee pagina’s bekeken werden. Op basis van deze ruwe gegevens worden *eerste-orde variabelen* berekend. Voorbeelden hiervan zijn:

- Pad lengte
- Aantal terugkeringen
- Het gebruik van de “back” knop
- Kijktijd per pagina
- Mate van verspreiding (‘fan degree’)
- Aantal cycli
- Net dichtheid (‘net density’)
- Gemiddelde verbonden afstand (‘average connected distance’)

Tweede-ordevariabelen zijn berekend als lineaire combinaties van de eerste-orde variabelen met behulp van principale component analyse. Deze tweede-orde variabelen worden (numeriek) volledig gespecificeerd door de eerste-orde variabelen. Het benoemen van de tweede-ordevariabelen en de interpretatie van hun betekenis is echter gebaseerd op de correlaties met gebruikers karakteristieken en de resultaten op de taken. De geïnterpreteerde tweede-orde variabelen worden "navigatie stijlen" genoemd. Twee navigatiestijlen worden hieronder beschreven:

- **Oppervlakkige Navigatie** (**'Flimsy Navigation')** is een spaarzame navigatiestijl. Het navigatiepad is niet erg uitgebreid en het merendeel van de navigatie vindt plaats rondom de startpagina. Er wordt meer tijd gestoken in het verwerken van de inhoud dan aan het begrijpen van de hyperstructuur die aangeeft waar de relevante informatie te vinden is.
- **Arbeidszame Navigatie** (**'Laborious Navigation')** houdt in dat er intensief gebruik wordt gemaakt van de navigatie infrastructuur van de site. Gebruikers lijken een trial en error strategie toe te passen. Ze volgen links alleen om te kijken of ze bruikbaar zijn of niet. Ze merken vrij snel als paden niet naar het juiste resultaat leiden en keren dan terug. Het opnieuw bezoeken van eerdere pagina's komt veelvuldig voor maar deze bezoeken zijn niet overbodig: als een pagina opnieuw wordt bezocht dan is dat vanuit een andere link dan de eerste keer, het was simpelweg een nieuwe poging.

Een *semantische* maat die *Pad adequaatheid* wordt genoemd, is berekend op basis van navigatie data en de taakbeschrijving die proefpersonen kregen aan het begin van hun taak. Een navigatiepad wordt gedefinieerd als een optelsom van semantische objecten die de gebruiker tegenkomt op zijn/haar weg. Pagina titels en link labels zijn gebruikt als semantische objecten. Link labels waren beter dan pagina titels bij het karakteriseren van het gebruiker's navigatiepad omdat ze meer specifieke informatie bevatten. Navigatiepaden zijn gebruikt in simulaties als indicator van contextuele informatie die van belang is bij het selecteren van specifieke navigatie acties. Pad adequaatheid wordt gedefiniëerd door de coëfficiënt van semantische overeenkomst tussen een navigatiepad en een taakomschrijving. Semantische overeenkomst wordt berekend met behulp van Latent Semantic Analysis (LSA). Pad adequaatheid zoals die berekend werd na afloop van een specifieke taak, correleerde significant met ruimtelijk inzicht (r=0.36) en taakprestaties (r=0.47). Pad adequaatheid, berekend bij elke stap in een navigatie sessie, is gebruikt in de simulaties van Webnavigatie als coherentie criterium dat van belang is bij het selecteren van specifieke navigatie acties.

De verschillende soorten kennis over gebruikers die afgeleid kunnen worden van informatie verkregen uit Web-logging data zijn dus s*yntactische* informatie en *semantische* informatie. S*yntactische* informatie wordt voornamelijk bepaald door de navigatiestijl van gebruikers, bijvoorbeeld of ze liever pagina's herbezoeken en nieuwe pagina's bekijken, of ze terugkeren met behulp van de "back" knop of dat ze liever links volgen, etc. *Semantische* informatie vormt een indicatie of gebruikers hun doelen effectief verwezenlijken, onafhankelijk van hun navigatiestijl. Deze navigatie maten kunnen gebruikt worden bij het ontwikkelen van adaptieve Web applicaties, zoals aanbevelingssystemen ('recommender systems')(Juvina & Van Oostendorp, 2004).

Algemeen gangbare theorieën over tekstbegrip (Kintsch, 1998), geheugen en cognitie (Anderson, 1983) en werkgeheugen (Baddeley, 1986) zijn toegepast bij het ontwikkelen van ons model voor Webnavigatie. Bij het toepassen van deze theorieën om Webnavigatie gedrag te verklaren vallen een aantal specifieke aspecten op. Bijvoorbeeld, bij Webnavigatie is een ruimtelijke representatie van de informatieruimte van veel groter belang dan bij het lezen van normale teksten. Gebruikers maken en actualiseren een mentale representatie van de informatieruimte waarin ze navigeren. De representatie heeft een ruimtelijk karakter (in die zin dat positie en afstanden relevant zijn, niet in de zin van visuele of driedimensionale aspecten) en is relatief onafhankelijk van de inhoud (de semantiek) die gerepresenteerd wordt. Gebruikers maken inschattingen van relevantie en nemen beslissingen om bepaalde informatie te selecteren. Dit doen ze op basis van hun bestaande kennis over een bepaald onderwerp en op basis van de informatie die ze vergaren uit de lokale context van een bepaalde informatiebron of plek. Ook in het geval van slechtzienden die gebruik maken van het Web zijn ruimtelijke aspecten essentieel. Voor slechtzienden kunnen ruimtelijke aspecten betrekking hebben op de temporele positie van een bepaald informatie element in een opeenvolging, een item in een categorie, etc.

Een cognitief model (genaamd CoLiDeS+) wordt gepresenteerd (*sectie 4.3*) waarin inschattingen van relevantie gemaakt worden op basis van zowel eerdere kennis (gemodelleerd door een LSA semantische ruimte, met behulp van het begrip 'information scent') als contextuele informatie (gemodelleerd door eerdere keuzes van de gebruiker met behulp van het begrip pad adequaatheid). CoLiDeS+, een uitbreiding van CoLiDeS (Kitajima, Blackmon, & Polson, 2000) heeft aangetoond dat het belangrijke aspecten van navigatiegedrag van gebruikers kan verklaren, zoals het gebruik van contextuele informatie bij het beoordelen van doelrelevantie en het toepassen van navigatie strategieën. Dit is gedaan door het navigatiepad van de gebruiker op te nemen in het model (met behulp van pad adequaatheid) en het model toe te staan terug te keren en eerdere selecties te heroverwegen (‘next-best strategy’). Dit model is empirisch getoetst om te beoordelen hoe goed het feitelijk gebruikersgedrag simuleert en hoe bruikbaar het model is om ondersteuning bij Webnavigatie te genereren. Hoewel het model feitelijk gebruikersgedrag niet heel erg goed simuleert, bleek CoLiDeS+ wel beter te presteren in vergelijking met de eerdere versie van het model (CoLiDeS). CoLiDeS+ is verder gebruikt om navigatieondersteuning te genereren en deze ondersteuning heeft een positieve invloed op het gebruikersgedrag en op de resultaten bij de taken. Een aantal beperkingen van CoLiDeS+ zijn geïdentificeerd, zoals een geringe nauwkeurigheid bij het simuleren van gebruikersgedrag. Dit wordt voornamelijk veroorzaakt door de afhankelijkheid van LSA bij het berekenen van semantische overeenkomst en de grote hoeveelheid handmatige coderingen die vereist zijn bij het draaien van de simulaties.

Een ACT-R model van Webnavigatie wordt gepresenteerd (*sectie 5.2*) om de mogelijkheden te demonstreren om beperkingen van CoLiDeS+ op te lossen. Het ACT-R model biedt computationele oplossingen om kernaspecten van Webnavigatie gedrag zoals genoemd in de literatuur en gevonden in ons empirisch onderzoek te implementeren. Een aantal van deze aspecten wordt gedeeld met eerdere modellen – selecties gebaseerd op doelrelevantie (‘information scent’) (SNIF-ACT en CoLiDeS); terugkeringen, drempelwaarden en opportunistische strategieën (MESA); ‘back coherence’ (pad adequaatheid, CoLiDeS+) – , andere aspecten werden hier voor het eerst geïmplementeerd – het vervlechten van conservatieve en exploratieve strategieën en ‘post-valued recall’.

Er wordt voorgesteld dat een cognitief model van Webnavigatie gebruikt kan worden om ondersteuning te genereren voor Webnavigatie, vooral als het model gespecificeerd wordt in computationele termen en hierdoor gebruikt kan worden als een computerprogramma (Juvina & Van Oostendorp, 2005). In een serie empirische studies zijn verschillende manieren getest waarop navigatie ondersteuning die gebaseerd is op dit model, kan worden aangeboden (*secties 4.4, 4.5, en 5.1*).

Mondelinge aanbevelingen van relevante links om een bepaald doel te bereiken bleken te leiden tot verbetering van taakprestaties. Bij gebruikers met laag ruimtelijk inzicht namen de prestaties meer toe dan bij gebruikers met een hoog ruimtelijk inzicht. Het lijkt erop dat de aangeboden navigatieondersteuning gebruikers ervan heeft weerhouden om tijd en cognitieve vermogens te besteden aan navigatie acties die niet direct effectief zijn maar die normaal worden toegepast om een nauwkeurige representatie te vormen van de informatieruimte. Gebruikers houden zich bezig met ogenschijnlijk zinloze navigatieacties om bekend te raken met de context van een specifieke informatie-eenheid. Deze acties zijn uiteindelijk toch nuttig bij het bepalen van de waarde van een bepaald stuk informatie. Hieruit volgt dat gebruikers met laag ruimtelijk inzicht waarschijnlijk minder goed in staat zijn om een representatie te maken van de informatieruimte en daarom hebben zij meer baat bij een cognitief model dat hen ondersteunt bij het vormen van die representatie. Echter, de gesproken aanbevelingen werden vanuit subjectief oogpunt niet goed ontvangen. Gebruikers vonden de aanbevelingen vervelend en manipulatief.

Grafische aanbevelingen in de vorm van kleine rode pijlen die naar relevante links wezen (Juvina & Herder, 2005) bleken niet alleen effectief, maar werden ook positief beoordeeld vanuit subjectief standpunt. Mannelijke gebruikers die op deze wijze ondersteuning kregen vertoonden een vermindering van waargenomen desoriëntatie vergeleken met mannelijke gebruikers in de controlegroep, maar dit verschil trad niet op bij vrouwelijke gebruikers. Ondersteuning bij het navigeren in de vorm van grafische suggesties voor te volgen links veranderde ook de structuur van het navigatiegedrag van gebruikers. Deelnemers die ondersteuning kregen, gebruikten de "back" knop minder vaak en de 'gemiddelde verbonden afstand' in de navigatiepaden was groter dan bij deelnemers uit de controlegroep. Link aanbevelingen zorgen er dus voor dat proefpersonen meer lineair navigeerden en ook verminderde het aantal keren terugkeringen. Proefpersonen die hoog presteerden neigden naar het minder vaak overnemen van aanwijzingen dan gebruikers die gemiddeld of laag presteerden. Echter, voor alle prestatieniveaus geldt: het volgen van meer aanwijzingen is geassocieerd met een toename van de taakprestatie. In het geval van het gebruik van scherm lezers ('screen readers') voor het Web, zijn de suggesties geïmplementeerd door de volgorde van items ('hyperlinks') op een webpagina zodanig te veranderen dat relevante links hoger geplaatst werden zodat ze in een opeenvolging (van links) eerder werden gelezen. Deze manipulatie bleek niet succesvol, waarschijnlijk omdat het veranderen van de volgorde van links op een webpagina de door de auteurs aangebrachte coherentie verstoort.

Dit onderzoek biedt geen ondersteuning voor het extraheren van relevante informatie vanuit de originele context zoals dat bijvoorbeeld bij zoekmachines gebeurt. In plaats daarvan is aangetoond dat het benadrukken van relevante informatie binnen de originele context gebruikers helpt om onderscheid te maken tussen relevante en niet-relevante informatie en dat het compensatie kan bieden voor het gebrek aan ruimtelijk inzicht (als dit bij de gebruiker van toepassing was).

Het uitvoeren van ecologisch valide onderzoek was een belangrijk desideratum bij dit project; zoveel mogelijk is gebruik gemaakt van echte websites en realistische taken (*Hoofdstuk 6*). Het vinden van de juiste balans tussen de nauwkeurigheid van het model en de praktische relevantie is een leidraad geweest bij dit onderzoek. CoLiDeS+ is ontworpen met als doel om een model-gebaseerde navigatie ondersteuning te ontwikkelen. Om deze reden zijn een aantal methodologische criteria van cognitief modelleren versoepeld. Het simuleren van gebruikersgedrag was niet volledig. Het model heeft bijvoorbeeld geen mechanisme om de inhoud van doelpagina's te identificeren. Zo'n mechanisme zou extreem moeilijk te ontwikkelen zijn en het was niet noodzakelijk voor het doel van het model. Het model was bedoeld om naast de gebruiker te werken en om links aan te bevelen die relevant zijn voor het bereiken van een bepaald gebruikersdoel. Er is hierbij verondersteld dat de gebruiker deze aanbevelingen kan overnemen of niet en dat de gebruiker stopt op het moment dat de doelpagina is bereikt. De prestaties van cognitieve modellen op het gebied van Webnavigatie hangen af van de voortgang die wordt geboekt op andere gebieden, zoals machine learning en natuurlijke taalverwerking. Het was een grote uitdaging om cognitief modelleren toe te passen op een taakdomein dat zo weinig gestructureerd en kennis intensief is als het domein dat in dit project is gebruikt. Een cognitief model van Webnavigatie moet om kunnen gaan met natuurlijke taal, grote kennis netwerken en een grote hoeveelheid sub-symbolische berekeningen. Deze aspecten zijn geen onderdeel van het traditionele onderzoek op het gebied van cognitief modelleren (Gluck & Pew, 2005), maar ze worden meer en meer prominent op het vakgebied van cognitief modelleren (Pirolli, 2005). Een voor de hand liggend onderwerp voor toekomstig onderzoek is het omgaan met de computationele complexiteit die vereist is wanneer cognitieve modellen uitgebreid worden zodanig dat ze ingebouwd kunnen worden in adaptieve Web applicaties als generator van navigatie ondersteuning.

Curriculum Vitae

Ion Juvina was born in 1967 in the suburbs of Giurgiu - a small town in the south of Romania, next to the Danube River. He did his high-school and university studies in Bucharest.

His professional experience includes working as a

- Research and teaching assistant, for University of Bucharest, "Politehnica" University, and Institute of Psychology, all in Bucharest.
- Industrial psychologist, for Romanian Railway Authority, Institute of Aviation Medicine, and Ministry of Interior, all in Bucharest.
- Consultant in Human Resources and Management.

He is a founding member of ROCHI, the Romanian chapter of ACM SIGCHI – a special interest group in Human-computer Interaction. As a member of this association, he has been involved since 2001 with voluntary organizational, editorial and scientific work.

Since February 2002, Ion Juvina has been hired by Utrecht University as a Research and Teaching Assistant. His research work has been directed toward obtaining a doctoral degree and it has been reported in this dissertation.

List of recent publications

Herder, E. & Juvina, I. (2004). Discovery of Individual User Navigation Styles. *Adaptive Hypermedia AH2004*, Eindhoven.

Juvina, I. & Herder, E. (2005). The Impact of Link Suggestions on User Navigation and User Perception. *UM2005 User Modeling: Proceedings of the Tenth International Conference*, 483-493, Edinburgh: Springer.

Juvina, I. & van Oostendorp, H. (2003). Human Factors in Web-assisted Personal Finance. In D. Harris, V. Duffy, M. Smith, & C. Stephanidis, *Human-Centred Computing. Cognitive, Social and Ergonomic Aspects. HCI International 2003 - Conference Proceedings*. Crete, Greece, 3, 477-481, Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Juvina, I., & van Oostendorp, H. (2004). Predicting user preferences - from semantic to pragmatic metrics of Web navigation behavior. *SIGCHI-NL*, Amsterdam: ACM Digital Library.

Juvina, I. & van Oostendorp, H. (2005). Bringing Cognitive Models in the Domain of Web Accessibility. *HCI International*, Las Vegas.

Juvina, I. & van Oostendorp, H. (2005). Cognitive Model Working Alongside the User. *British HCI2005*, 409-420, Edinburgh: Springer.

Juvina, I., & van Oostendorp, H. (2006). Individual differences and behavioral metrics involved in modeling Web navigation. *Universal Access in Information Society*. (4)3, 258-269, Springer.

Juvina, I., & van Oostendorp, H. (2006). Enhancing internet experience of visually impaired persons by means of dynamic highlighting and selective reading. Information Design Journal, (14)1, 71-81.

Juvina, I., Trausan-Matu, St., Iosif, Gh., Van der Veer, G., Marhan, A-M., & Chisalita, C. (2002). Analysis of Web Browsing Behavior - a great potential for psychological research. *TAMODIA 2002 - Task Models and Diagrams for User Interface Design*, 170-178, Bucharest, INFOREC Publishing House.

Van Oostendorp, H. & Juvina, I. (2003). Role of Icons and Chatboxes in Computer Supported Collaborative Learning. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for Change in Networked Learning Environments*. 275-279, Dordrecht: Kluwer Academic Publishers.

SIKS Dissertatiereeks

1998-1 Johan van den Akker (CWI)
DEGAS - An Active, Temporal Database of Autonomous Objects
1998-2 Floris Wiesman (UM)
Information Retrieval by Graphically Browsing Meta-Information
1998-3 Ans Steuten (TUD)
A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective
1998-4 Dennis Breuker (UM)
Memory versus Search in Games
1998-5 E.W.Oskamp (RUL)
Computerondersteuning bij Straftoemeting
1999-1 Mark Sloof (VU)
Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products
1999-2 Rob Potharst (EUR)
Classification using decision trees and neural nets
1999-3 Don Beal (UM)
The Nature of Minimax Search
1999-4 Jacques Penders (UM)
The practical Art of Moving Physical Objects
1999-5 Aldo de Moor (KUB)
Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems
1999-6 Niek J.E. Wijngaards (VU)
Re-design of compositional systems
1999-7 David Spelt (UT)
Verification support for object database design
1999-8 Jacques H.J. Lenting (UM)
Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation.
2000-1 Frank Niessink (VU)
Perspectives on Improving Software Maintenance
2000-2 Koen Holtman (TUE)
Prototyping of CMS Storage Management
2000-3 Carolien M.T. Metselaar (UVA)
Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief.
2000-4 Geert de Haan (VU)
ETAG, A Formal Model of Competence Knowledge for User Interface Design
2000-5 Ruud van der Pol (UM)
Knowledge-based Query Formulation in Information Retrieval.
2000-6 Rogier van Eijk (UU)
Programming Languages for Agent Communication

2000-7 Niels Peek (UU)
Decision-theoretic Planning of Clinical Patient Management
2000-8 Veerle Coup, (EUR)
Sensitivity Analyis of Decision-Theoretic Networks
2000-9 Florian Waas (CWI)
Principles of Probabilistic Query Optimization
2000-10 Niels Nes (CWI)
Image Database Management System Design Considerations, Algorithms and Architecture
2000-11 Jonas Karlsson (CWI)
Scalable Distributed Data Structures for Database Management
2001-1 Silja Renooij (UU)
Qualitative Approaches to Quantifying Probabilistic Networks
2001-2 Koen Hindriks (UU)
Agent Programming Languages: Programming with Mental Models
2001-3 Maarten van Someren (UvA)
Learning as problem solving
2001-4 Evgueni Smirnov (UM)
Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets
2001-5 Jacco van Ossenbruggen (VU)
Processing Structured Hypermedia: A Matter of Style
2001-6 Martijn van Welie (VU)
Task-based User Interface Design
2001-7 Bastiaan Schonhage (VU)
Diva: Architectural Perspectives on Information Visualization
2001-8 Pascal van Eck (VU)
A Compositional Semantic Structure for Multi-Agent Systems Dynamics.
2001-9 Pieter Jan 't Hoen (RUL)
Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes
2001-10 Maarten Sierhuis (UvA)
Modeling and Simulating Work Practice
BRAHMS: a multiagent modeling and simulation language for work practice analysis and design
2001-11 Tom M. van Engers (VUA)
Knowledge Management:
The Role of Mental Models in Business Systems Design
2002-01 Nico Lassing (VU)
Architecture-Level Modifiability Analysis
2002-02 Roelof van Zwol (UT)
Modelling and searching Web-based document collections
2002-03 Henk Ernst Blok (UT)
Database Optimization Aspects for Information Retrieval
2002-04 Juan Roberto Castelo Valdueza (UU)

The Discrete Acyclic Digraph Markov Model in Data Mining
2002-05 Radu Serban (VU)
The Private Cyberspace Modeling Electronic
Environments inhabited by Privacy-concerned Agents
2002-06 Laurens Mommers (UL)
Applied legal epistemology; Building a knowledge-based ontology of the legal domain
2002-07 Peter Boncz (CWI)
Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications
2002-08 Jaap Gordijn (VU)
Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas
2002-09 Willem-Jan van den Heuvel(KUB)
Integrating Modern Business Applications with Objectified Legacy Systems
2002-10 Brian Sheppard (UM)
Towards Perfect Play of Scrabble
2002-11 Wouter C.A. Wijngaards (VU)
Agent Based Modelling of Dynamics: Biological and Organisational Applications
2002-12 Albrecht Schmidt (Uva)
Processing XML in Database Systems
2002-13 Hongjing Wu (TUE)
A Reference Architecture for Adaptive Hypermedia Applications
2002-14 Wieke de Vries (UU)
Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems
2002-15 Rik Eshuis (UT)
Semantics and Verification of UML Activity Diagrams for Workflow Modelling
2002-16 Pieter van Langen (VU)
The Anatomy of Design: Foundations, Models and Applications
2002-17 Stefan Manegold (UVA)
Understanding, Modeling, and Improving Main-Memory Database Performance
2003-01 Heiner Stuckenschmidt (VU)
Ontology-Based Information Sharing in Weakly Structured Environments
2003-02 Jan Broersen (VU)
Modal Action Logics for Reasoning About Reactive Systems
2003-03 Martijn Schuemie (TUD)
Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy
2003-04 Milan Petkovic (UT)

Content-Based Video Retrieval Supported by Database Technology
2003-05 Jos Lehmann (UVA)
Causation in Artificial Intelligence and Law - A modelling approach
2003-06 Boris van Schooten (UT)
Development and specification of virtual environments
2003-07 Machiel Jansen (UvA)
Formal Explorations of Knowledge Intensive Tasks
2003-08 Yongping Ran (UM)
Repair Based Scheduling
2003-09 Rens Kortmann (UM)
The resolution of visually guided behaviour
2003-10 Andreas Lincke (UvT)
Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture
2003-11 Simon Keizer (UT)
Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks
2003-12 Roeland Ordelman (UT)
Dutch speech recognition in multimedia information retrieval
2003-13 Jeroen Donkers (UM)
Nosce Hostem - Searching with Opponent Models
2003-14 Stijn Hoppenbrouwers (KUN)
Freezing Language: Conceptualisation Processes across ICT-Supported Organisations
2003-15 Mathijs de Weerdt (TUD)
Plan Merging in Multi-Agent Systems
2003-16 Menzo Windhouwer (CWI)
Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouses
2003-17 David Jansen (UT)
Extensions of Statecharts with Probability, Time, and Stochastic Timing
2003-18 Levente Kocsis (UM)
Learning Search Decisions
2004-01 Virginia Dignum (UU)
A Model for Organizational Interaction: Based on Agents, Founded in Logic
2004-02 Lai Xu (UvT)
Monitoring Multi-party Contracts for E-business
2004-03 Perry Groot (VU)
A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving
2004-04 Chris van Aart (UVA)
Organizational Principles for Multi-Agent Architectures

2004-05 Viara Popova (EUR)
Knowledge discovery and monotonicity
2004-06 Bart-Jan Hommes (TUD)
The Evaluation of Business Process Modeling Techniques
2004-07 Elise Boltjes (UM)
Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes
2004-08 Joop Verbeek(UM)
Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politi‰le gegevensuitwisseling en digitale expertise
2004-09 Martin Caminada (VU)
For the Sake of the Argument; explorations into argument-based reasoning
2004-10 Suzanne Kabel (UVA)
Knowledge-rich indexing of learning-objects
2004-11 Michel Klein (VU)
Change Management for Distributed Ontologies
2004-12 The Duy Bui (UT)
Creating emotions and facial expressions for embodied agents
2004-13 Wojciech Jamroga (UT)
Using Multiple Models of Reality: On Agents who Know how to Play
2004-14 Paul Harrenstein (UU)
Logic in Conflict. Logical Explorations in Strategic Equilibrium
2004-15 Arno Knobbe (UU)
Multi-Relational Data Mining
2004-16 Federico Divina (VU)
Hybrid Genetic Relational Search for Inductive Learning
2004-17 Mark Winands (UM)
Informed Search in Complex Games
2004-18 Vania Bessa Machado (UvA)
Supporting the Construction of Qualitative Knowledge Models
2004-19 Thijs Westerveld (UT)
Using generative probabilistic models for multimedia retrieval
2004-20 Madelon Evers (Nyenrode)
Learning from Design: facilitating multidisciplinary design teams
2005-01 Floor Verdenius (UVA)
Methodological Aspects of Designing Induction-Based Applications
2005-02 Erik van der Werf (UM))
AI techniques for the game of Go
2005-03 Franc Grootjen (RUN)
A Pragmatic Approach to the Conceptualisation of Language
2005-04 Nirvana Meratnia (UT)
Towards Database Support for Moving Object data

2005-05 Gabriel Infante-Lopez (UVA)
Two-Level Probabilistic Grammars for Natural Language Parsing
2005-06 Pieter Spronck (UM)
Adaptive Game AI
2005-07 Flavius Frasincar (TUE)
Hypermedia Presentation Generation for Semantic Web Information Systems
2005-08 Richard Vdovjak (TUE)
A Model-driven Approach for Building Distributed Ontology-based Web Applications
2005-09 Jeen Broekstra (VU)
Storage, Querying and Inferencing for Semantic Web Languages
2005-10 Anders Bouwer (UVA)
Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments
2005-11 Elth Ogston (VU)
Agent Based Matchmaking and Clustering - A Decentralized Approach to Search
2005-12 Csaba Boer (EUR)
Distributed Simulation in Industry
2005-13 Fred Hamburg (UL)
Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen
2005-14 Borys Omelayenko (VU)
Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics
2005-15 Tibor Bosse (VU)
Analysis of the Dynamics of Cognitive Processes
2005-16 Joris Graaumans (UU)
Usability of XML Query Languages
2005-17 Boris Shishkov (TUD)
Software Specification Based on Re-usable Business Components
2005-18 Danielle Sent (UU)
Test-selection strategies for probabilistic networks
2005-19 Michel van Dartel (UM)
Situated Representation
2005-20 Cristina Coteanu (UL)
Cyber Consumer Law, State of the Art and Perspectives
2005-21 Wijnand Derks (UT)
Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics
2006-01 Samuil Angelov (TUE)
Foundations of B2B Electronic Contracting
2006-02 Cristina Chisalita (VU)
Contextual issues in the design and use of information technology in organizations

2006-03 Noor Christoph (UVA)
The role of metacognitive skills in learning to solve problems
2006-04 Marta Sabou (VU)
Building Web Service Ontologies
2006-05 Cees Pierik (UU)
Validation Techniques for Object-Oriented Proof Outlines
2006-06 Ziv Baida (VU)
Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling
2006-07 Marko Smiljanic (UT)
XML schema matching -- balancing efficiency and effectiveness by means of clustering
2006-08 Eelco Herder (UT)
Forward, Back and Home Again - Analyzing User Behavior on the Web
2006-09 Mohamed Wahdan (UM)
Automatic Formulation of the Auditor's Opinion
2006-10 Ronny Siebes (VU)
Semantic Routing in Peer-to-Peer Systems
2006-11 Joeri van Ruth (UT)
Flattening Queries over Nested Data Types
2006-12 Bert Bongers (VU)
Interactivation - Towards an e-cology of people, our technological environment, and the arts
2006-13 Henk-Jan Lebbink (UU)
Dialogue and Decision Games for Information Exchanging Agents
2006-14 Johan Hoorn (VU)
Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change
2006-15 Rainer Malik (UU)
CONAN: Text Mining in the Biomedical Domain
2006-16 Carsten Riggelsen (UU)
Approximation Methods for Efficient Learning of Bayesian Networks
2006-17 Stacey Nagata (UU)
User Assistance for Multitasking with Interruptions on a Mobile Device
2006-18 Valentin Zhizhkun (UVA)
Graph transformation for Natural Language Processing
2006-19 Birna van Riemsdijk (UU)
Cognitive Agent Programming: A Semantic Approach
2006-20 Marina Velikova (UvT)
Monotone models for prediction in data mining
2006-21 Bas van Gils (RUN)
Aptness on the Web
2006-22 Paul de Vrieze (RUN)

Fundaments of Adaptive Personalisation
2006-23 Ion Juvina (UU)
Development of Cognitive Model for Navigating on the Web