AUTHOR'S QUERY SHEET

Author(s): C. Katsanos et al. THAM 461026
Article title:
Article no:

Dear Author

The following queries have arisen during the editing of your manuscript and are identified on the proofs. Unless advised otherwise, please submit all corrections using the CATS online correction form.

AQ1 Is the running head ok?

AQ2 Please spell out [ALT] and [CWW] in full at the first mention.

AQ3 Please provide publisher details.

AQ4 Please provide page numbers.

AQ5 Please provide city/state of publisher.

New Review of Hypermedia and Multimedia,
Vol. 16, No. 1, April 2010, 1–20

Taylor & Francis
Taylor & Francis Group

Evaluating website navigability: validation of a tool-based approach through two eye-tracking user studies

CHRISTOS KATSANOS†*, NIKOLAOS TSELIOS†,‡ and NIKOLAOS AVOURIS†

†HCI Group, Department of Electrical and Computer Engineering, University of Patras, 26500 Rio, Patras, Greece

‡ICT in Education Group, Department of Educational Sciences and Early Childhood Education, University of Patras, 26500 Rio, Patras, Greece

(Received 14 December 2008; final version received 7 January 2010)

Following information scent has been established as a metaphor to describe user's behaviour while navigating an information space by successively selecting hyperlinks. This metaphor suggests that users assess the profitability of following a particular hyperlink based on its perceived semantic association with their goal. The purpose of this paper is to study how information scent, this important attribute of hypermedia navigability, influences concurrently four aspects of users' behaviour while exploring a website: (1) distribution of attention; (2) confidence in choice of link; (3) efficiency; and (4) effectiveness. It was found that in webpages with high scent, users were significantly more focused, confident of their choices, efficient and effective compared to webpages with ambiguous scent. The findings of the study are discussed in comparison with results obtained from a previously conducted analysis using InfoScent Evaluator (ISEtool), a tool that has been proposed to facilitate scent evaluation of websites. This comparison provided support for the effectiveness of ISEtool in indicating potential scent-related navigability problems. We argue that such a tool-based approach can facilitate hypermedia design by reducing the resources and expertise required, and by providing the necessary flexibility for practitioners.

Keywords: Information scent; Eye-tracking study; Automated tool; Latent Semantic Analysis

1. Introduction

The web has become part of many peoples' everyday life activities. Reading the news, booking a hotel, planning a trip, seeking information in an encyclopaedia and shopping are just some of the activities many people do mostly online. A prime objective of websites' designers is people to be able to access, find and use the information and services easily.

*Corresponding author. Email: ckatsanos@ece.upatras.gr

New Review of Hypermedia and Multimedia
ISSN 1361-4568 print/ISSN 1740-7842 online © 2010 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/13614561003605179

Users access a website with some information needs and expectations about where to seek information. Web navigation, "a process during which users must continually make decisions regarding strategies for reaching a goal and determining whether they have reached that goal" (Furnas 1997), should facilitate achieving user goals. Kalbach (2007) argues that design of web navigation is "the systematic organization of links to provide access to information and to make meaningful associations in a way that enhances understanding, reflects brand, and lends to overall credibility of a site". Navigability has been long considered as an important aspect of hypermedia design and remains an issue of concern (Chen and Czerwinski 1997, Spool *et al.* 1999, Otter and Johnson 2000).

Navigation in a website is goal driven, dominated by local decisions. Link-labels come immediately before a decisive point in navigation: the transition from one page to another. Users scan the links of a webpage in an attempt to associate their labels with the content they are seeking and assess the profitability of following a particular link. Furnas (1997) coined the term "residue" to describe the hint that a link's description holds of what lays behind it. Residue was recast and refined by Pirolli (1997) as "information scent" and defined in Card *et al.* (2001) as a users "imperfect perception of the value, cost, or access path of information sources obtained from proximal cues, such as WWW links". Theoretical analysis (Pirolli 1997, Pirolli and Card 1999, Kitajima *et al.* 2000, Fu and Pirolli 2007) and empirical usability research (Spool *et al.* 2004 cited in Blackmon *et al.* 2005, Pirolli 2007, Brumby and Howes 2008) suggest that information scent is the most important factor in web navigation. Dillon and Vaugan (1997) eloquently stated that in hypertext navigation "the journey is the destination" in order to emphasise the importance of hyperlinks' semantics. Miller and Remington (2004) and Kalbach (2007) are in line with the above point, arguing that getting link-labels right is vital in web navigation.

However, people use a surprisingly great variety of words to refer to the same thing (Furnas 1997). Morville (2005) eloquently stated that "while the web's architecture rests on a solid foundation of code, its usefulness depends on the slippery slope of semantics. It's all about words; words as labels, words as links". Lack of appropriate links' descriptions can cause various navigability problems and deteriorate the overall information interaction experience. Users can become lost while following links with ambiguous labels, which were perceived to be semantically similar to their goal but did not lead to the targeted information. In addition, poorly labelled links can result to users falsely assuming that the information sought is not available at all. Previous research (Card *et al.* 2001, Blackmon *et al.* 2002, 2005, Pirolli 2007, Brumby and Howes 2008) has demonstrated that people have lower success rates and require more time to complete the task when they are presented with weak scent compared to high scent. Ceaparu and Shneiderman (2004) report study results in which users' performance nearly doubled and frustration levels dropped when the wording of link-labels was changed.

The goal of the research presented in this paper is to investigate the effect of information scent on web navigation behaviour. The contribution of this paper is that it investigates how the manipulation of information scent influences concurrently four aspects of users' behaviour: (1) distribution of attention; (2) confidence in choice of link; (3) efficiency; and (4) effectiveness. The findings of this study are discussed in comparison with results obtained from a previously conducted analysis using a tool-based approach for identifying scent-related, navigability problems. In this paper, the following two scent-related, navigability problems are addressed, which are extensively discussed in web navigation literature (Pirolli and Card 1999, Card *et al.* 2001, Blackmon *et al.* 2002, 2005, Withrow 2002, Chi *et al.* 2003, Fu and Pirolli 2007, Spool *et al.* cited in Pirolli 2007, Brumby and Howes 2008):

(1) A *weak-scent or low-scent link* occurs when the link that leads to the correct path does not have sufficient information scent. Therefore, it is perceived to be semantically similar to the user's goal at the same or at a lesser degree than the rest of the links. In such cases, users are expected to face a cognitive dead-end as no link seems to be suitable for their goal. Card *et al.* (2001) argue that "when information scent becomes low, users switch to another site or search engine". Usually, users encounter inadequate scent in webpages that use ambiguous and/or very short link-labels. Possible solutions to this problem include rewording of the correct link and/or the addition of links that create alternative paths to the targeted information (Kitajima *et al.* 2000).
(2) A *competing or confusing links* problem arises when a webpage contains more than one links that seem to have comparable information scent. In such cases, users are expected to be confused about their next click (Withrow 2002). Potential causes for this problem can be the choice of either semantically close or highly general link-labels. A solution to such cases is greater specificity and clarity that makes individual links distinct from each other or the creation of alternative paths to the information sought by following any of the competing links (Kitajima *et al.* 2000).

It should be noted that there are also other aspects apart from the quality of hyperlinks' descriptions that affect navigability of a website, including content structure and visual design. In this paper, however, we focus on the identification of navigability issues related to the semantic appropriateness of hyperlinks' descriptions. This decision was inspired and formulated by previous research findings showing that choice of words in headings and links is more important than content structure of a website (Miller and Remington 2004, Resnick and Sanchez 2004). In addition, Otter and Johnson (2000) present study results where 45% of the participants identified poorly or ambiguously labelled links as the main cause of their disorientation, while the location of the links and their visual characteristics received only a 9 and 5%, respectively. Thus, two studies were designed through which the information scent of the links of a controlled set of websites was evaluated using user behaviour data and a tool-based approach.

The tool used in the studies was InfoScent Evaluator tool (ISEtool; Katsanos *et al.* 2006), which has been proposed as a way to facilitate scent evaluation of websites and identify potential scent-related navigability problems. The tool is based on the Information Foraging model of users' behaviour when engaged in information search tasks (Pirolli and Card 1999), and uses semantic similarity measures to computationally exploit the notion of information scent. A textual description of the user's goal is compared with the textual description of each link in a webpage using semantic similarity measures. Currently, ISEtool uses Latent Semantic Analysis (LSA; Landauer and Dumais 1997).

The LSA served also as a computational model of information scent in the Cognitive Walkthrough for the Web (Blackmon *et al.* 2002, 2005), a method based on the Cognitive Walkthrough Evaluation (Wharton *et al.* 1994) and a comprehension-based model of web navigation, named Colides (Kitajima *et al.* 2000). Since the results presented are very promising, we also used LSA combined with a web data extraction algorithm and a stepwise evaluation approach in order to develop ISEtool. This tool is aimed at providing the necessary flexibility and efficiency to practitioners, and can be used both during the initial design and redesign of websites. Dickinson (2007) investigated how LSA-based tools evaluating navigability problems might be welcomed by practitioners in the "real-world". He reports research findings in which practitioners favoured ISEtool for its flexibility, intuitiveness, ease of use and overall usability, stating that "ISEtool is close to a form that could be used by developers with little training".

In the rest of the paper, the underlying mechanisms and a typical usage scenario of ISEtool are briefly described first. Next, the methodology of the two user studies is presented, followed by the analysis of the results. In the user studies, we investigated the impact of ambiguous information scent on four measured aspects of users' behaviour and compared the findings with the results obtained using ISEtool. Finally, we discuss the implications, future directions and conclusions of the presented research.

2. InfoScent Evaluator tool (ISEtool): a tool-based method to identify scent-related navigability problems

Evaluation of navigability is a necessary and repeated step during the lifecycle of a website. User testing and usability inspections are the two most common types of methods currently used to evaluate the navigation model of a website. Usability testing is based on the analysis of observational data collected from representative users performing typical tasks with the site, whereas inspection techniques involve a pool of experts identifying usability issues against a set of established guidelines or heuristics.

However, both approaches have some inherent limitations. Usability testing is expensive, takes a lot of time and requires the availability of suitable subjects to complete the tests, whereas inspection methods require expertise, produce many false-alarms and are inevitably subjective. In addition, both

types of methods lack adequate reliability, since the usability findings can vary widely when different evaluators study the same website, even if the same evaluation technique is used; a phenomenon known as the evaluator's effect (Hertzum and Jacobsen 2001). Automating aspects of the design and evaluation process is critical given the quantity, frequency of updates and sheer size of websites being produced (Brinck and Hofer 2002). Automation in usability evaluation reduces the level of expertise required, increases consistency of findings, increases coverage of the evaluation and allows the comparison of alternative solutions.

Various automated approaches have been proposed in the literature. Many of them are based on: (1) the analysis of usage patterns in log-files or (2) the assessment of static HTML according to specified criteria. Examples of the former category include ANTS (Gonzalez-Rodriguez *et al.* 2007) and TEA (Obendorf *et al.* 2004). NIST's WebSat, available at http://zing.ncsl.nist.gov/WebTools/tech.html, assesses static HTML according to a user-selected set of guidelines, whereas WebTango (Ivory and Hearst 2002) evaluates a website according to empirically validated quantitative attributes.

In addition, simulation-based approaches for evaluating websites have been also proposed in the literature. For example, Chi *et al.* (2003) have developed the Bloodhound tool to predict success rates of simulated users based on the Web User Flow by Information Scent algorithm (WUFIS; Chi *et al.* 2001). Furthermore, recent work-in-progress research, such as CogTool-Explorer (Teo and John 2008) and Visual Saliency Maps (Veksler and Gray 2007) attempt to combine models of visual search with semantic similarity measures to improve the quality of predictions.

ISEtool has been proposed as such a simulation-based automated approach to identify potential navigability problems related to the semantic appropriateness of hyperlinks' descriptions. Currently, the tool is made freely available upon request at http://hci.ece.upatras.gr/isetool. ISEtool combines a web data extraction algorithm and semantic similarity measures to computationally exploit the notion of information scent. This is achieved by deriving a quantitative estimation of the semantic similarity between a textual description of a typical user goal and each link-label in the evaluated webpage.

Various approaches to estimate semantic similarity between texts have been proposed in the literature (Miller 1995, Landauer and Dumais 1997, Rhode *et al.* 2004). Currently, ISEtool employs the LSA technique (Landauer and Dumais 1997). The adequacy of LSA's reflection of human meaning-based judgements has been established in a variety of cases, such as judging essay quality (Landauer *et al.* 2003), structuring the information space of a website (Katsanos *et al.* 2008), identifying navigability problems (Blackmon *et al.* 2005) and providing real-time navigation support (van Oostendorp and Juvina 2007). However, given that there is much ongoing research targeted at understanding which of the semantic similarity measures perform better in the context of information scent modelling (Turney 2001, Kaur and Hornof 2005, Budiu *et al.* 2007), ISEtool has been

built on a software framework that allows easy integration of alternative algorithms.

Next, a short description of LSA is presented first, followed by a typical scenario of using the tool and a brief discussion of the tool's major improvements compared to its initial version.

2.1 Latent Semantic Analysis (LSA)

The underlying idea of LSA is that "the totality of information about all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and set of words to each other" (Landauer *et al.* 1998). First, LSA parses suitable, large text corpora that represent a given population's understanding of words and produces a type-document matrix of each type's frequency of occurrence. Types are the individual components making up a document and are typically terms (Martin and Berry 2007). Next, each cell is weighted by a function that expresses both the type's importance in the particular document and the degree to which it carries information in the domain of discourse in general. Subsequently, LSA applies singular value decomposition (SVD), the mathematical generalisation of factor analysis, to find the least-squares best fit number of dimensions. In this way, the initial matrix is transformed into a semantic space, in which the degree of semantic similarity between any pair of texts, such as the description of a user goal and a link-label, is measured by the cosine of the corresponding two vectors. Each cosine value lies between $+1$ (identical) and -1 (opposite). Near-zero values represent unrelated texts.

2.2 InfoScent Evaluator tool (ISEtool)—a typical usage scenario

A typical scenario of applying the proposed tool-based methodology is the following. First, the evaluator identifies the typical users of the website. Subsequently, a set of user goals is compiled including typical activities that each user group is likely to try to accomplish on the evaluated site. Next, the evaluator provides a typical user goal, specifies the URL of a webpage and selects an appropriate semantic space representative of the user profile under evaluation (see figure 1a). Currently, ISEtool employs LSA in a web-service-like approach and the semantic spaces available at the LSA website (lsa.colorado.edu) are automatically retrieved and provided as a list to select from.

Subsequently, ISEtool runs an automated analysis that combines a web data extraction algorithm with the selected similarity estimation algorithm to compute the information scent for all the links of the page. The web data extraction algorithm collects the labels of all textual hyperlinks and the alternative texts (i.e. ALT tags) of all graphical hyperlinks. It is assumed that graphical hyperlinks are represented by their alternative texts, but since ALT tags may be totally missing in some cases, the tool offers the possibility to add

AQ2

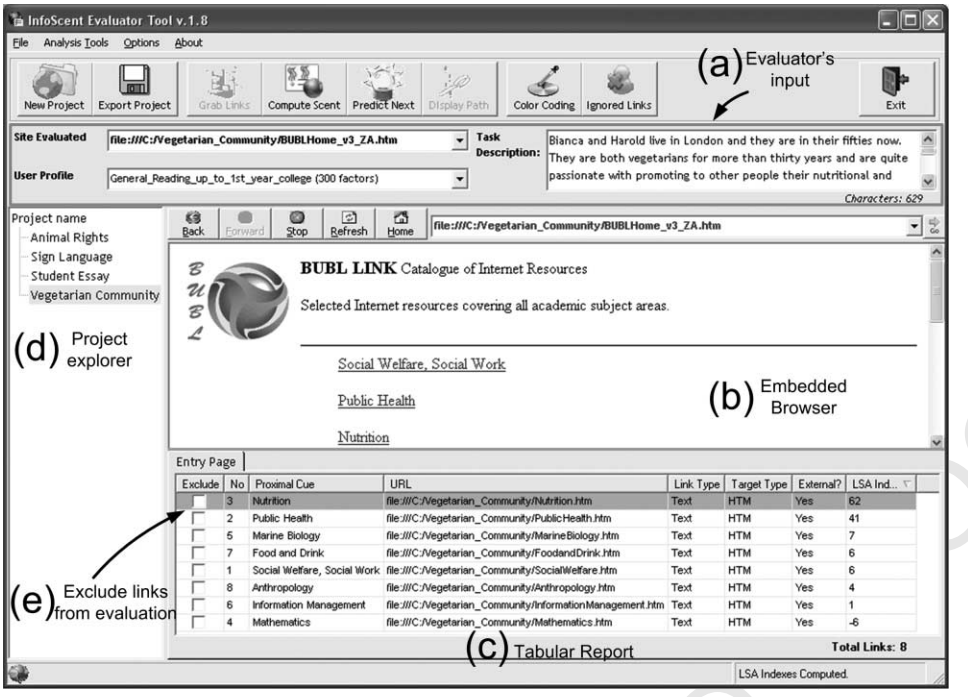

Figure 1. Using ISEtool to evaluate information scent of a webpage and identify potential navigability problems.
Note: The competing-links homepage of the websites used in the first study is displayed. The potential competing link is coloured red (darker in greyscale).

descriptions of graphical hyperlinks manually. Similarly, scripts are handled by asking from the user of the tool to optionally provide a description. In addition, various other attributes of the links are also collected, such as the type of the links' target (e.g. downloadable file or webpage).

Currently, ISEtool uses LSA one-to-many analysis in a web-service-like approach in order to estimate the semantic similarity of the user goal against all the retrieved textual descriptions of hyperlinks. Specifically, ISEtool transparently fills and submits the one-to-many LSA analysis form using: (1) the semantic space selected by the evaluator; (2) the LSA document-to-document comparison type; (3) the provided user goal; and (4) the list of extracted links, separated by blank lines. Next, the LSA indexes for all the goal-link pairs are retrieved from the LSA one-to-many analysis results webpage and stored in the tool's database.

The output of the tool is a tabular report presenting the collected attributes and the information scent of the links in the evaluated webpage (see figure 1c). The evaluator can sort the results by descending or ascending computed information scent. Potential scent-related navigability problems are colour-coded based on empirically validated heuristics, such as "a link is classified as competing when it has at least 80% similarity to the goal that the correct link has" (Blackmon *et al.* 2002). Both the heuristics and the colours used to denote

problems can be easily adjusted through the tool's interface. The whole process can be iterated by having the tool automatically follow the link with the highest scent. It is worth mentioning that the tool offers a number of additional options to the evaluator, such as exporting the results in various formats and visualising the simulated user trail.

The typical usage scenario described so far concerns an initial version of the tool, which was introduced in a previous paper (Katsanos *et al.* 2006). Since then, ISEtool has been iteratively improved based on valuable insight gained from a study that investigated its usage by practitioners (Dickinson 2007), from comments of the tool's users and from authors' own experience of integrating it in their usability evaluation practices. First, an internal browser has been embedded in its interface, which can allow the designer to take also into account the visual layout of the webpage, while interpreting the results (see figure 1b). This internal browser is synchronised with the tool's tabular report; that is when a link is selected in the tabular report it is auto-focused and highlighted in the browser. In addition, the flexibility of the tool has been increased by extending its functionalities in order to allow the practitioner to: (1) conduct multiple evaluations in the same ISEtool-project (see figure 1d); (2) choose any link as the next step in the simulated path instead of automatically picking the one with the highest computed information scent; and (3) exclude links from the current evaluation (see figure 1e).

3. Studies

Two studies were conducted to investigate the effect of information scent on users' link-selection behaviour. In this context, the quality of results obtained by ISEtool was compared to user study data. In the next sections, first, the common elements of the two studies are described, and then, each study is presented separately, followed by a discussion and future directions of the presented research.

3.1 Methodology and procedures

In both studies, the navigability of experimental websites was first evaluated using the proposed tool-based method, and then through an eye-tracking user study. All the experimental websites consisted of nine pages organised in a two-level hierarchy. Each homepage presented a navigation menu with eight links that lead to a webpage in the second level. Figure 2a presents a representative example of such a homepage. A task description was associated with each website. The homepages were designed to present different scent-conditions for their associated tasks. These scent-conditions were produced as follows: first, one of the authors compiled a list of navigation menus and task descriptions by visiting a number of real-world websites, such as www.howstuffworks.com, to get inspired. Next, the other two authors selected the pairs of tasks and menus that they perceived to

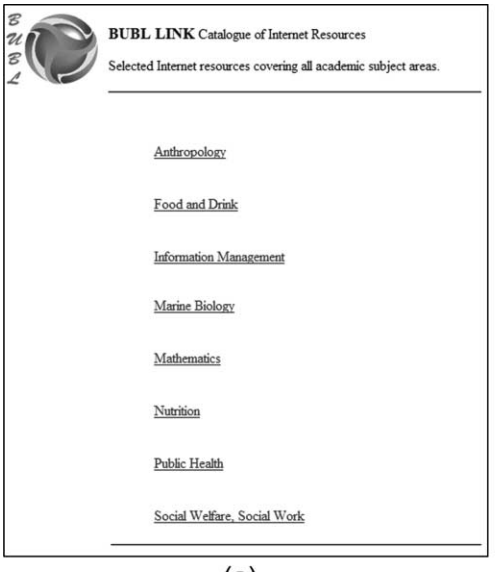

Figure 2. (a) Representative example of the homepage of a website used in the study. (b) Associated typical user goal.

produce the desired scent-conditions for each study. Subsequently, three colleagues, who were not involved in the study, verified these scent-conditions and the plausibility of the associated tasks.

First, the websites were evaluated using ISEtool. Detailed descriptions of the typical user goals were provided as input to the tool and the "General Reading up to 1st year college" LSA semantic space was selected as the representative user profile. Typical tasks were expressed through personas (Cooper 1999). Figure 2b presents a representative example of such a task description. LSA is a text-based, similarity-estimation algorithm and therefore longer and contextually richer text descriptions of the user's goal should increase its reflection of human meaning-based judgements. Blackmon *et al.* (2002, 2005, 2007) used detailed goal statements with CWW (i.e. automatically generated summaries of the target encyclopaedia article) to increase the accuracy of LSA's predictions.

Subsequently, the tool run an automated analysis that: (1) collected the links of the homepage; (2) transparently filled and submitted the one-to-many LSA analysis form using the collected links, the selected semantic space and the provided user goal; and (3) extracted the LSA indexes for all goal-link pairs from the LSA results webpage. Next, the tabular output of the tool was sorted by descending information scent and navigability problems were detected based on the empirically validated criteria reported in Blackmon *et al.* (2002). Figure 1 presents such an example, in which a competing link is identified as a link with at least 80% of the correct-link's LSA index (i.e.

"Public Health") and is coloured red. In this case, the page was classified as a competing-links homepage.

The same websites were evaluated through eye-tracking user studies. Participants were students of a University Department and were offered partial course credit for taking part in the study. All participants were proficient in English (the websites' language) and experienced internet users. Participants were asked to perform the tasks using a typical web browser and their behaviour was monitored. The presentation of the websites and the order of links in the homepages were counter-balanced to avoid serial order effects. An unobtrusive 17″ Tobii T60 eye tracker with minimum fixation duration set to 100 ms and screen resolution set to 1024×768 pixels was used to record participants' eye movements.

The procedure of the studies was the following: first, participants were asked to complete a short, pre-test questionnaire recording their sex, age, internet experience and English proficiency. Subsequently, they were provided with the first task and read it at their own convenience. Afterwards, they were presented with the corresponding website and asked to complete the task at hand. Participants were instructed to look at all links before selecting the one that felt more appropriate for each task in order to control for the effects of the visual search behaviour of satisficing (Brumby and Howes 2004), according to which people select the first "good enough option" and not the "optimal solution". The end of the task was marked by either the participant raising her hand or expiration of the three-minute time limit. After the task, participants were asked to rate on a 1–10 scale the confidence level for their first click; that is, how sure they felt they would find the requested information following the selected link. The same process was repeated for the next task until all available tasks were completed by all participants.

Four measures capturing different aspects of participants' information interaction behaviour were collected. Effectiveness of the participants' link-selection behaviour was measured on a binomial level for their first link-click (*first-click success*), whereas their efficiency was captured by the time required to make their first link-selection (*time for first click*). Furthermore, a self-reported measure of the participants' confidence in their first link-selection (*first-click confidence*) was measured. Spool *et al.* (2004 cited in Pirolli 2007) argue that the best way to detect scent is to measure users' confidence, since they have observed that when the scent is not clear users doubt their choices and often report that they are making "wild guesses". In addition, the total number of participants' observations on the available links (*total link-observations*) was measured as an indicator of participants' distraction. A *link-observation* was defined as every time a person has looked within an Area of Interest (AOI) including a link, starting with a fixation within the AOI and ending with a fixation outside the AOI. Rectangular AOIs were centred on the description of each link expanding 20 pixels from each side of the link's text. A large number of total link-observations indicates that the user sampled and re-sampled the available links prior to selecting one, as if distracted or

hindered from isolating the target (Goldberg and Kotval 1999). Link-observations were used instead of simple on-link fixations to avoid bias of higher fixations counts due to lengthier text descriptions across the homepage-conditions. Table 1 summarises the hypotheses evaluated in the two studies.

There are many potential other methods we could have used instead of eye-tracking to collect data on users' link-selection behaviour. One way to more closely monitor a user's behaviour during a task is through the think-aloud technique (Nielsen 1993). This technique, however, requires extra effort on the part of the user and affects task performance measures. Alternatively, retrospective think-aloud (Nielsen 1993) could have been used. However, unlike eye-tracking, this method can at best give indirect evidence of users' behaviour since it involves users' subjective interpretation and people tend to underestimate or overestimate events that happened in the past.

3.2 Study 1

In the first study, three experimental websites were evaluated for three associated tasks. One homepage was designed to present high scent, one to present weak scent and one to present a competing-links problem. The results from the ISEtool evaluation verified the problems we designed. In the high-scent homepage, the correct link had a scent value of 0.52 and the rest of the links had values below 0.13. In the competing-links homepage, the correct link had 0.41, the competing link 0.62 and the rest of the links below 0.07. In the weak-scent homepage, the correct link had –0.02 and the rest of the links had values ranging from 0.09 to −0.07. Thirty-one participants, 23 males, aged 21–44 with a mean age of 23, took part in the eye-tracking user study that evaluated the navigability of these three websites. The procedure described in the previous section was followed.

Table 2 presents descriptive statistics of the dependent variables measured in the two studies. Figure 3 presents heatmaps of participants' total duration of fixations and representative gaze-plots of participants' fixations sequence in each homepage. These data indicate that there were differences in the observed participants' behaviour due to the variety in the quality of the presented links.

In the high-scent page, attention was mainly focused in the area containing the correct link and participants made a selection quickly after their first scan of the available links (10.5 sec on average). This pattern indicates a focused

Table 1. Hypotheses evaluated in the two studies.

Effectiveness	$\mathbf{H_1}$	Participants' first-click success will be the same in all homepage-conditions
Efficiency	$\mathbf{H_2}$	Participants' time for first-click will be the same in all homepage-conditions
Confidence	$\mathbf{H_3}$	Participants' first-click confidence will be the same in all homepage-conditions
Distraction	$\mathbf{H_4}$	Participants' total link-observations will be the same in all homepage-conditions

Table 2. Descriptive statistics of the dependent variables measured in the two studies.

		First-click success (%)	Average time for first-click (sec)	Average first-click confidence (1–10)	Average total link-observations
Study 1	High scent (1)	87±12	10.5±1.7	9.4±0.3	31±06
	Competing links (1)	16±13	15.9±3.1	6.8±0.3	51±10
	Weak scent (1)	06±10	23.5±4.7	2.9±0.8	67±11
Study 2	High scent (3)	91±10	14.1±1.8	8.3±0.4	14±02
	Weak scent (3)	17±13	21.4±2.8	6.4±0.4	20±02

and efficient search, which can be also observed in participants' high first-click success percentage (87%) and self-reported confidence (9.4 sec on average). In the page with the competing-links problem, there was a slight distribution of attention indicating lack of confidence regarding the correct choice. Participants fixated and refixated some of the presented links and required more time before making their choice (15.9 sec on average). This lack of confidence can be also observed in the users' lower self-reported confidence (6.8 sec on average) compared to the high-scent page. Finally, in the weak-scent page, attention was diffused across all the links indicating increased uncertainty. Participants scanned and rescanned the list of links trying to identify one that was appropriate for the given goal. This increased uncertainty is also depicted in the participants' delay in making a choice (23.5 sec on average), very low first-click success percentage (6%) and very low confidence (2.9 sec on average) compared to the other pages.

Proportion analysis indicated a significant effect of the ISEtool-identified problem's type on participants' *first-click success*, Cochran's $Q(3) = 39.93$, $p < 0.001$. Therefore $\mathbf{H_1}$ was rejected (see table 1). Follow-up McNemar's tests using the Bonferroni adjustment criterion revealed a significant difference for the high-scent page versus the weak-scent page ($p < 0.001$) and the competing-links page ($p < 0.001$), respectively. On the contrary, no significant difference was found for the weak-scent page versus the competing-links page ($p = 0.453$).

A repeated measures MANOVA was applied to evaluate the effect of the ISEtool-identified problem's type on participants' *time for first click*, *first-click confidence* and *total link-observations*. A significant main effect was found, $F(6, 25) = 40.03$, $p < 0.001$. Follow-up univariate ANOVAs were used to analyze further the results. Whenever the assumption of sphericity was violated (i.e. Machly's test was significant), degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity. The univariate analyses indicated a significant effect of the scent-condition on the time participants required to make their first link-choice, their self-reported confidence for their choice and the total number of link-observations: $F(1.67, 50.13) = 27.70$, $p < 0.001$, $F(1.67, 50.12) = 94.71$, $p < 0.001$ and $F(2, 60) = 13.04$, $p < 0.001$, respectively. Therefore $\mathbf{H_2}$, $\mathbf{H_3}$ and $\mathbf{H_4}$ were also rejected (see table 1).

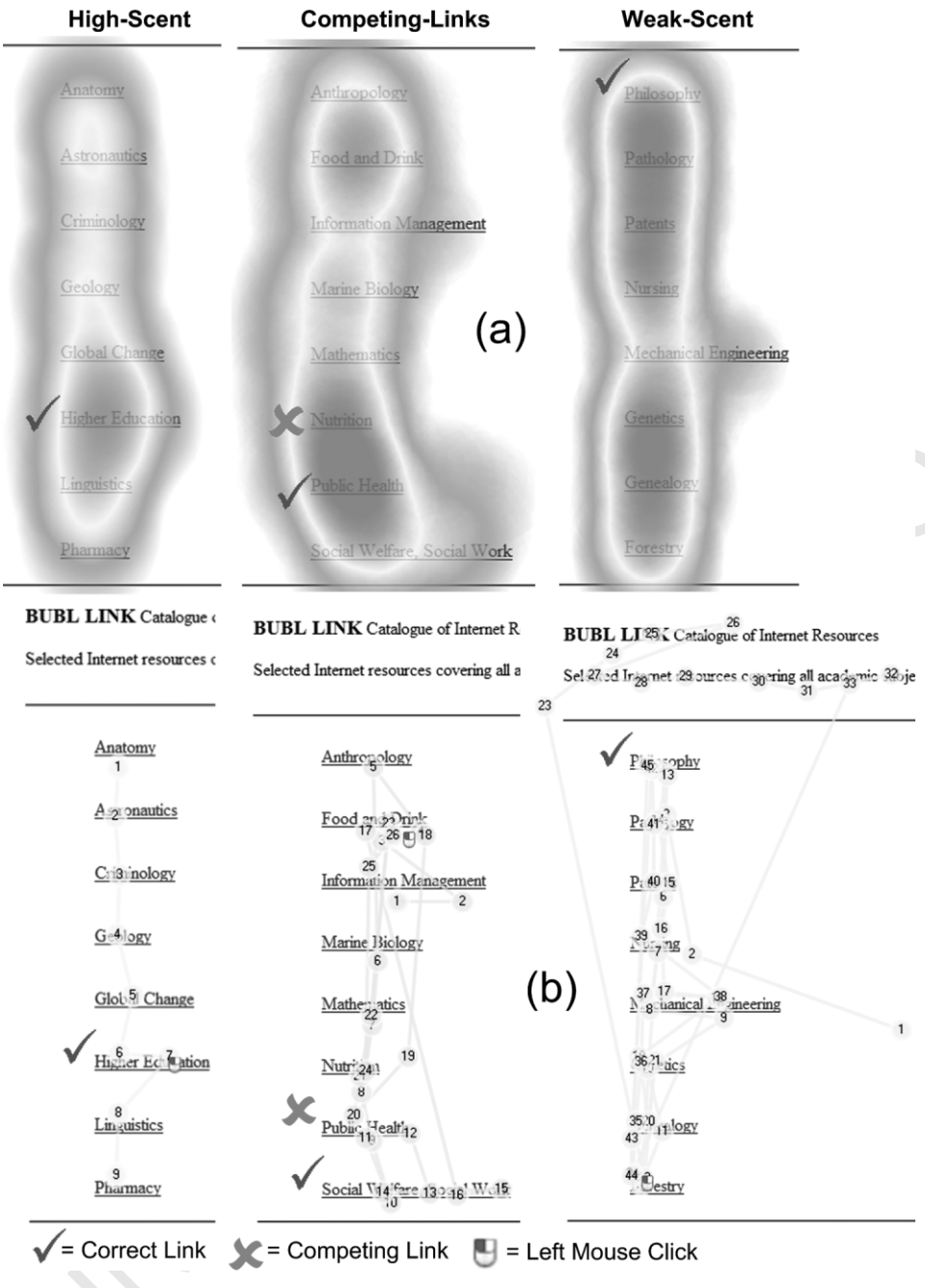

Figure 3. (a) Heatmaps of the total duration of participants' fixations on the links presented in the scent-conditions of the first study. Areas where participants looked the most are coloured red (darker in greyscale). (b) Representative gaze-plots of participants' fixations sequence in each webpage.

Planned contrasts were conducted to investigate pair-wise differences. In the weak-scent homepage, participants required significantly more time to click on a link, were significantly less confident for their choice and were significantly more distracted compared to the high-scent homepage: $F(1, 30) = 38.36$, $p < 0.001$, $\omega^2 = 0.75$, $F(1, 30) = 225.48$, $p < 0.001$, $\omega^2 = 0.94$ and $F(1, 30) = 29.02$, $p < 0.001$, $\omega^2 = 0.70$, respectively. Similarly, when competing links existed, participants' time for first click, confidence and level of distraction differed significantly compared to the high-scent homepage: $F(1, 30) = 14.97$, $p < 0.001$, $\omega^2 = 0.57$, $F(1, 30) = 39.82$, $p < 0.001$, $\omega^2 = 0.75$ and $F(1, 30) = 9.18$, $p < 0.005$, $\omega^2 = 0.48$, respectively. Furthermore, the weak-scent problem was found to be significantly more serious since all three measures differed significantly compared to the competing-links page: $F(1, 30) = 18.65$, $p < 0.001$, $\omega^2 = 54.62$, $F(1, 30) = 47.43$, $p < 0.001$, $\omega^2 = 0.78$ and $F(1, 30) = 4.20$, $p < 0.05$, $\omega^2 = 0.35$, respectively.

3.3 Study 2

In the second study, six experimental websites were evaluated. Three of the websites presented a high-scent homepage and the rest three a weak-scent homepage. The results from the ISEtool evaluation also identified these scent-conditions. In the high-scent homepages, the correct link had an average scent value of 0.60 and the rest of the links had values below 0.14. In the weak-scent homepages, the correct link had an average scent value of 0.13 and the rest of the links were below 0.08. An eye-tracking user study was conducted to evaluate the navigability of the same six websites. Thirty-five participants, 29 males, aged 21–25 with a mean age of 22 participated. The same procedure followed in the previous study was also used in this study.

The six websites were grouped into two scent-conditions for the analysis of the results; that is three websites with a high-scent homepage and three with a weak-scent homepage. Table 2 presents descriptive statistics of the measured dependent variables. Figure 4 presents heatmaps of participants' total duration of fixations and representative gaze-plots of participants' fixations sequence in each homepage. These data indicate that information scent had an influence on the participants' attention-focusing and link-selection behaviour.

In the high-scent pages, attention appeared to be mainly focused in the area containing the correct link and participants made a selection quickly after their first scan of the available links (14.1 sec on average). This pattern indicates a focused and efficient search, which can be also observed in the representative gaze-plots of participants and in their high first-click success (91%), and self-reported confidence (8.3 sec on average). In the weak-scent pages, attention was distributed across most of the links indicating increased uncertainty. Participants scanned and rescanned the list of links trying to identify one that was appropriate for the given goal. This increased uncertainty is also depicted in the participants' more time required to select a link (21.4 sec on average), lower first-click success rate (17%) and lower

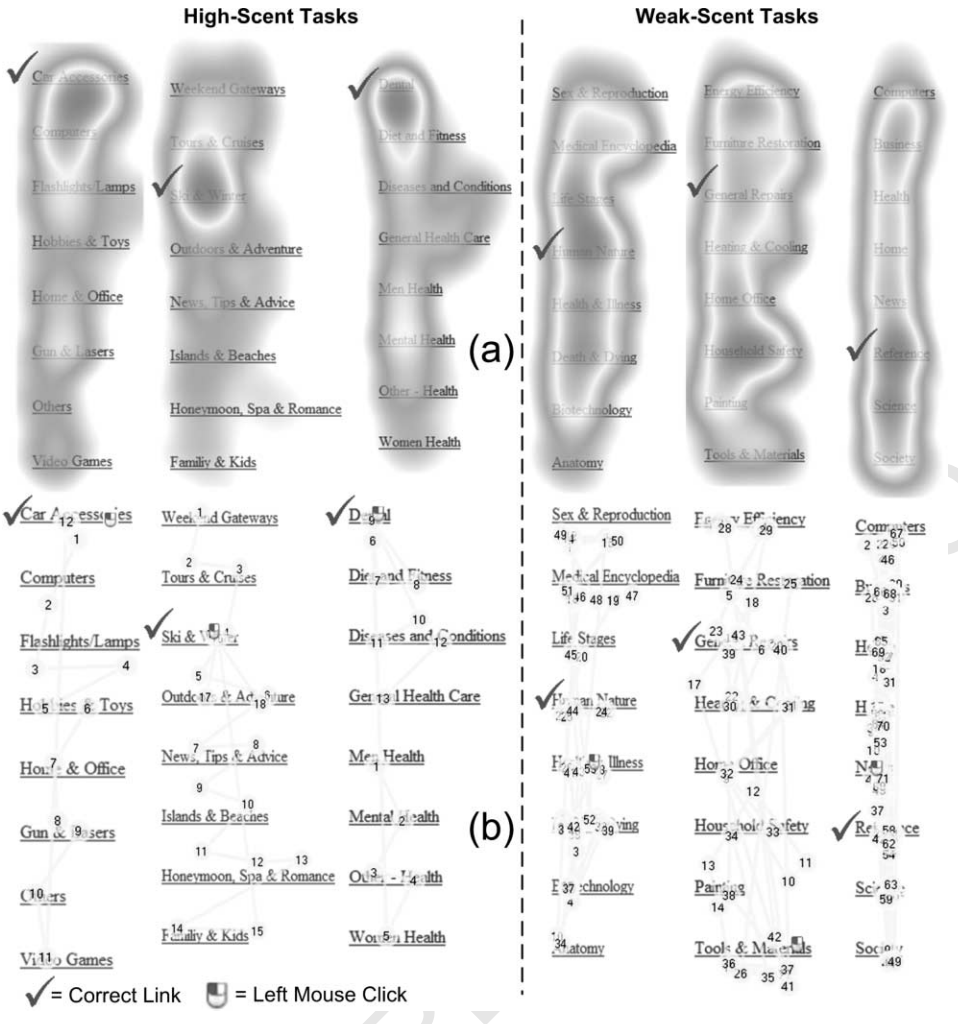

Figure 4. (a) Heatmaps of the total duration of participants' fixations on the links in the scent-conditions of the second study. Areas where participants looked the most are coloured red (darker in greyscale). (b) Representative gaze-plots of participants' fixations sequence in each webpage.

confidence (6.4 sec on average) compared to the high-scent pages. These findings replicate the ones obtained in the first study.

A McNemar's test was applied to evaluate the effect of the ISEtool-identified problem's type on participants' *first-click success.* Participants were significantly less successful with their first click in the weak-scent condition ($p < 0.001$). Thus, $\mathbf{H_1}$ was rejected (see table 1). A repeated measures MANOVA evaluated the effect of the scent-related problem's type on participants' *time for first click*, *first-click confidence* and *total link-observations.* A significant main effect was found, $F(3, 32) = 17.33$, $p < 0.001$. Follow-up analysis indicated that participants required significantly more

time to click on a link, were significantly less confident for their choice and were significantly more distracted in the weak-scent webpages; $F(1, 34) = 30.59$, $p < 0.001$, $\omega^2 = 0.69$, $F(1, 34) = 49.33$, $p < 0.001$, $\omega^2 = 0.77$ and $F(1, 34) = 29.02$, $p < 0.001$, $\omega^2 = 0.68$, respectively. Therefore, **H_2, H_3** and **H_4** were also rejected (see table 1).

4. Discussion and conclusions

In this paper, the results of two eye-tracking user studies are reported in which the impact of ambiguous information scent was measured on four aspects of users' link-selection behaviour: (1) distribution of attention; (2) confidence; (3) efficiency; and (4) effectiveness. Using four dependent measures provide a more complete picture of exactly how variations in the pattern of information scent on a webpage influence users' attention and link-selection behaviour. To the best of our knowledge, this is the first paper to do so. In this context, the validity of a tool-based approach to facilitate scent evaluation of websites was also investigated by comparing its results to user study data.

Table 2 presents descriptive statistics of the four dependent variables measured in the two studies. It was found that in webpages with high-scent users had significantly more focused attention–allocation patterns, reported significantly higher levels of confidence for their choices, were significantly more correct with their first click, and required significantly less time to select a link compared to pages with competing links or weak scent. In particular, participants navigating a page with high scent were on average 6.8 times more successful with their first click, 1.6 times faster in their selection, twice less distracted and reported being 1.6 times more confident compared to pages containing any of the aforementioned problems[1]

Figures 3a and 4a show heatmaps of the total duration of participants' fixations on the links presented in each webpage. It was found that in high-scent webpages participants' attention was mainly focused in the area containing the correct link. This indicates a focused and efficient search, which is also obvious in the representative gaze-plot of participants' fixations sequence (figures 3b and 4b). In the pages with any of the aforementioned scent-related problems, participants appeared more distracted than in the high-scent webpages.

Furthermore, it was found that participants encountered much difficulty in webpages with weak scent. Averaging across the two studies, in high-scent tasks 89% of participants selected the correct link in 12.3 sec and were very confident for their choice (8.9 out of 10). In contrast, for weak-scent tasks participants were on average 12% successful, required 22.5 sec to select a link, and reported being rather unconfident for their choice (4.7 out of 10).

The findings of the presented studies are discussed in comparison with results obtained from a previously conducted analysis using ISEtool, a computational tool that facilitates the evaluation of hyperlinks' information scent based on semantic similarity measures. This comparison provided

support for the effectiveness of ISEtool in indicating potential scent-related navigability problems. Such a tool-based approach automates aspects of the evaluation process and therefore decreases the time and human resources required. This level of automation is critical for practitioners since they often work under strict time and cost constraints. In addition, the proposed approach can enable comparisons between alternative designs, increase coverage of the evaluation, reduce the level of expertise required and increase consistency of usability findings. Furthermore, ISEtool can be used to aid in the identification of problems sometimes neglected in user testing or expert-based methods. Potential useful applications of the tool include its integration with content management systems, an idea introduced by "real-world" practitioners in Dickinson (2007), and its application in accessibility evaluation since research identifies misleading links as one of the top reasons of frustration for users with disabilities (Lazar *et al.* 2007).

However, we must caution against too much interpretation of the results. Although the findings of the two studies are promising, additional studies with "real-world" websites are required before any safe claims can be made about the tool's usefulness in hypermedia navigation design. In the present studies, we designed simplified versions of webpages containing mainly the links to be tested and instructed participants to consider all the available links before making a selection. However, actual websites are rarely that simple and users navigating them often employ much more complex strategies than an assess-all approach (Brumby and Howes 2008). Future work includes investigating ideas that would allow ISEtool to model the effect of both a website's layout and user's link-skipping behaviour in order to further increase the accuracy of its predictions. For instance, Veksler and Gray (2007) have introduced the idea of "visual saliency maps", which essentially combines visual characteristics of a page's layout, such as colour and clutter, with semantic similarity measures to predict the user's link-selection behaviour.

In addition, ISEtool is currently based on the assumption that the proximal cue of a textual hyperlink is only its link-label. However, this assumption does not always hold true in actual websites. For instance, in many online newspapers links are grouped into blocks with headings that are not links or include short descriptions of an article with a "more ..." link. Eye-tracking studies could be a useful tool to better understand users' attention-focusing processes while seeking information in the web and investigate such research questions.

Future research should also investigate ways of producing goal descriptions that capture the domain knowledge of the task (also discussed in Chi *et al.* 2003). In this paper, we have used personas for this purpose, which we argue that can provide the necessary contextual information that is required for accurate scent modelling. Other researchers (Blackmon *et al.* 2007) have used automatically produced summaries of the target article to address this aspect. The latter approach can further automate the evaluation process, but such goal descriptions may only model users with an exceptional amount of

knowledge about the target information. In contrast, personas can represent varying knowledge and skill levels of the typical users of a website.

Finally, it should be stressed that despite the advantages of the presented automated approach, the value of established techniques should not be neglected. Instead, such automated approaches confirm the usefulness of traditional user methods, which can be applied in a more focused way, with significantly less effort and in a more systematic manner. In addition, such tools provide bridges between research and Web design practice and make it more likely for Web practitioners to embrace and employ them in order to improve the navigability of their websites, even when there are strict time and cost constraints or lack of the required expertise.

Note

[1] These ratios are derived from table 2 by first averaging the collected measures across the two studies for each scent-condition, and then grouping the conditions with scent-related problems and comparing them to the high-scent condition.

References

M.H. Blackmon, M. Kitajima and P.G. Polson, "Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs", in *SIGCHI Conference on Human Factors in Computing Systems*, 2–7 April 2005, Portland, OR/New York: ACM, pp. 51–58, 2005.

M.H. Blackmon, D.R. Mandalia, P.G. Polson and M. Kitajima, "Automating usability evaluation: Cognitive walkthrough for the web puts LSA to work on real-world design problems", in *Handbook of Latent Semantic Analysis*, T.K. Landauer, D.S. McNamara, D. Simon, and W. Kintsch, (Eds), Mahwah, NJ: Lawrence Erlbaum, pp. 345–375, 2007.

M.H. Blackmon, P.G. Polson, M. Kitajima and C. Lewis, "Cognitive walkthrough for the web", in *SIGCHI Conference on Human Factors in Computing Systems*, 20–25 April 2002, Mineapolis, MN/New York: ACM, pp. 463–470, 2002.

T. Brinck and E. Hofer, "Automatically evaluating the usability of web sites", in *CHI'02 Extended Abstracts on Human Factors in Computing Systems*, 20–25 April 2002, Minneapolis, MN/New York: ACM, pp. 906–907, 2002.

D. Brumby and A. Howes, "Good enough but I'll just check: Web-page search as attentional refocusing", in *6th International Conference on Cognitive Modeling*, 30 July–1 August 2004, Pittsburgh, PA: Carnegie Mellon University/University of Pittsburgh, pp. 46–51, 2004.

D. Brumby and A. Howes, "Strategies for guiding interactive search: An empirical investigation into the consequences of label relevance for assessment and selection", *Human–Computer Interaction*, 23(1), pp. 1–46, 2008.

AQ3,4 R. Budiu, C. Royer and P. Pirolli, "Modeling information scent: A comparison of LSA, PMI and GLSA similarity measures on common tests and corpora", in *8th RIAO Conference*, 30 May–1 June 2007, Pittsburgh, PA, 2007.

S. Card, P. Pirolli, M. Wege, J. Morrison, R.W. Reeder, P. Schraedley and J. Boshart, "Information scent as a driver of web behavior graphs: Results of a protocol analysis method for web usability", in *SIGCHI Conference on Human Factors in Computing Systems*, 31 March–5 April 2001, Seattle, WA/New York: ACM, pp. 498–505, 2001.

I. Ceaparu and B. Shneiderman, "Finding governmental statistical data on the Web: Three empirical studies of the FedStats topics page", *Journal of the American Society for Information Science and Technology*, 55(11), pp. 1008–1015, 2004.

C. Chen and M. Czerwinski, "Spatial ability and visual navigation: An empirical study", *New Review of Hypermedia and Multimedia*, 3(1), pp. 67–89, 1997.

E. Chi, P. Pirolli, K. Chen and J. Pitkow, "Using information scent to model user information needs and actions and the web", in *SIGCHI Conference on Human Factors in Computing Systems*, 31 March–5 April 2001, Seattle, WA/New York: ACM, pp. 490–497, 2001.

E. Chi, A. Rosien, G. Supattanasiri, A. Williams, C. Royer, C. Chow, E. Robles, B. Dalal, J. Chen and S. Cousins, "The Bloodhound Project: Automating discovery of web usability issues using the InfoScent Simulator", in *SIGCHI Conference on Human Factors in Computing Systems*, 5–10 April 2003, Fort Lauderdale, FL/New York: ACM, pp. 505–512, 2003.

AQ5 A. Cooper, "*The Inmates are Running the Asylum*", SAMS, 1999.

M. Dickinson, *A real world evaluation of two LSA-based usability tools for website navigation*, Thesis (Msc), University of Teesside, UK, 2007.

A. Dillon and M. Vaughan, "It's the journey and the destination': Shape and the emergent property of genre in evaluating digital documents", *New Review of Hypermedia and Multimedia*, 3(1), pp. 91–106, 1997.

W-T. Fu and P. Pirolli, "SNIF-ACT: A cognitive model of user navigation on the World Wide Web", *Human–Computer Interaction*, 22, pp. 355–412, 2007.

G. Furnas, "Effective view navigation", in *SIGCHI Conference on Human Factors in Computing Systems*, 22–27 March 1997, Atlanta/New York: ACM, pp. 367–374, 1997.

H.J. Goldberg and X.P. Kotval, "Computer interface evaluation using eye movements: Methods and constructs", *International Journal of Industrial Ergonomics*, 24, pp. 631–645, 1999.

M. Gonzalez-Rodriguez, F. Diez-Diaz and A. Vidau, "Remote navigability testing using data gathering agents", in *4th International Conference on Universal Access in Human–Computer Interaction*, 22–27 July 2007, Beijing, China: Springer, pp. 28–33, 2007.

M. Hertzum and N.E. Jacobsen, "The evaluator effect: A chilling fact about usability evaluation methods", *International Journal of Human–Computer Interaction*, 13(4), pp. 421–443, 2001.

M. Ivory and M. Hearst, "Statistical profiles of highly-rated web sites", in *SIGCHI Conference on Human Factors in Computing Systems*, 20–25 April 2002, Mineapolis, MN/New York: ACM, pp. 367–374, 2002.

AQ5 J. Kalbach, "*Designing Web Navigation: Optimizing the User Experience*", O'Reilly Media, 2007.

C. Katsanos, N. Tselios and N. Avouris, "InfoScent evaluator: A semi-automated tool to evaluate semantic appropriateness of hyperlinks in a web site", in *20th Conference of the Computer–Human Interaction Special Interest Group (CHISIG) of Australia on Computer–Human Interaction*, 20–24 November 2006, Sydney, Australia/New York: ACM, pp. 373–376, 2006.

C. Katsanos, N. Tselios and N. Avouris, "AutoCardSorter: Designing the information architecture of a web site using latent semantic analysis", in *SIGCHI Conference on Human Factors in Computing Systems*, 5–10 April 2008, Florence, Italy/New York: ACM, pp. 875–878, 2008.

I. Kaur and A.J. Hornof, "A comparison of LSA, WordNet and PMI–IR for predicting user click behavior", in *SIGCHI Conference on Human Factors in Computing Systems*, 2–7 April 2005, Portland, OR/ New York: ACM, pp. 51–60, 2005.

M. Kitajima, M.H. Blackmon and P.G. Polson, "A comprehension-based model of web navigation and its application to web usability analysis", in *People and Computers XIV – Usability or Else! (Proceedings of HCI 2000)*, S. McDonald, Y. Waern, and G. Cockton, (Eds), Vienna, Austria: Springer, pp. 357–373, 2000.

T.K. Landauer and S. Dumais, "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge", *Psychological Review*, 104(2), pp. 211–240, 1997.

T.K. Landauer, P.W. Foltz and D. Laham, "Introduction to latent semantic analysis", *Discourse Processes*, 25, pp. 259–284, 1998.

T.K. Landauer, D. Laham and P. Foltz, "Automatic essay assessment", *Assessment in Education: Principles Policy and Practice*, 10(3), pp. 295–308, 2003.

J. Lazar, A. Allen, J. Kleinman and C. Malarkey, "What frustrates screen reader users on the web: A study of 100 blind users", *International Journal of Human–Computer Interaction*, 22(3), pp. 247–269, 2007.

D.I. Martin and M.W. Berry, "Mathematical foundations behind latent semantic analysis", in *Handbook of Latent Semantic Analysis*, T.K. Landauer, D.S. McNamara, D. Simon, and W. Kintsch, (Eds), Mahwah, NJ: Lawrence Erlbaum, pp. 35–55, 2007.

G.A. Miller, "WordNet: A lexical database for English", *Communications of the ACM*, 38(11), pp. 39–41, 1995.

C.S. Miller and R.W. Remington, "Modeling information navigation: Implications for information architecture", *Human–Computer Interaction*, 19(3), pp. 225–271, 2004.

AQ5 P. Morville, *Ambient Findability: What We Find Changes Who We Become*, O'Reilly Media, 2005.

J. Nielsen, *Usability Engineering*, Boston, MA: Academic Press, 1993.

H. Obendorf, H. Weinreich and T. Hass, "Automatic support for web user studies with SCONE and TEA", in *SIGCHI Conference on Human Factors in Computing Systems*, 24–29 April 2004, Vienna, Austria/New York: ACM, pp. 1135–1138, 2004.

M. Otter and H. Johnson, "Lost in hyperspace: Metrics and mental models", *Interacting with Computers*, 13(1), pp. 1–40, 2000.

P. Pirolli, "Computational models of information scentfollowing in a very large browsable text collection", in *SIGCHI Conference on Human Factors in Computing Systems*, 22–27 March 1997, Atlanta/New York: ACM, pp. 3–10, 1997.

AQ5 P. Pirolli, *Information Foraging Theory: Adaptive Interaction with Information*, USA: Oxford University Press, 2007.

P. Pirolli and S. Card, "Information foraging", *Psychological Review*, 106(4), pp. 643–675, 1999.

M.L. Resnick and J. Sanchez, "Effects of organizational scheme and labeling on task performance in product-centered and user-centered retail web sites", *Human Factors*, 46(14), pp. 104–117, 2004.

D.L. Rhode, L.M. Gonnerman and D.C. Plaut, An introduction to COALS [online]. Available online at: http://dlt4.mit.edu/ ~dr/COALS/Coals.pdf (accessed 4 February 2008), 2008.

AQ5 J. Spool, T. Scanlon, W. Schroeder, C. Snyder and T. DeAngelo, *Web Site Usability: A Designer's Guide*, Morgan–Kaufmann, 1999.

L. Teo and B. John, "Cogtool-explorer: Towards a tool for predicting user interaction", in *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, 5–10 April 2008, Florence, Italy/New York: ACM, pp. 2793–2798, 2008.

P.D. Turney, "Mining the web for synonyms: PMI–IR versus LSA on TOEFL", in *Twelfth European Conference on Machine Learning*, 3–7 September 2001, Freiburg, Germany/Berlin: Springer, pp. 491–502, 2001.

H. van Oostendorp and I. Juvina, "Using a cognitive model to generate web navigation support", *International Journal of Human–Computer Studies*, 65(10), pp. 887–897, 2007.

V. Veksler and W. Gray, "Mapping semantic relevancy of information displays", in *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, 28 April–3 May 2007, San Jose, CA/New York: ACM, pp. 2729–2734, 2007.

AQ5 C. Wharton, J. Rieman, C. Lewis and P. Polson, "The cognitive walkthrough method: A practitioner's guide", in *Usability Inspection Methods,* J. Nielsen and R.L. Mack (Eds), Wiley, pp. 105–140, 1994.

J. Withrow, "Do your links stink? Techniques for good web information scent", *Bulletin of the American Society for Information Science and Technology*, 28(5). Available online at: http://www.asis.org/Bulletin/Jun-02/withrow.html (accessed 15 November 2008), 2002.