World Scientific
www.worldscientific.com

A SURVEY OF TOOLS SUPPORTING DESIGN AND EVALUATION OF WEBSITES BASED ON MODELS OF HUMAN INFORMATION INTERACTION

CHRISTOS KATSANOS

HCI Group, Dept. Of Electrical and Computer Engineering, University of Patras
Rio, Patras, 26500, Greece
ckatsanos@ece.upatras.gr

NIKOLAOS TSELIOS[*,#]

[*]*HCI Group, Dept. Of Electrical and Computer Engineering, University of Patras*
[#]*ICT in Education Group, Dept. of Educational Sciences and Early Childhood Education, University of Patras*
Rio, Patras, 26500, Greece
nitse@ece.upatras.gr

NIKOLAOS AVOURIS

HCI Group, Dept. Of Electrical and Computer Engineering, University of Patras
Rio, Patras, 26500, Greece
avouris@upatras.gr

This paper presents and discusses a survey of existing tools to support design and evaluation of websites, with special emphasis on improving the information navigation process. The amount of information of today's websites, the continuous evolution of the medium and the heterogeneity of typical users' profiles, make the website design task particularly hard. The presented tools are mainly based on recent models of Web usage behavior, and involve various natural language and semantic similarity modeling methods. Validation studies of the presented tools have shown that they can support effectively various phases of the website design lifecycle including information structuring, hyperlink evaluation and assessment of alternative designs. In this paper, existing techniques are discussed, the aspects of Web design that (should and) could be better supported are identified and suggestions are made on extensions of existing approaches to better support the usability evaluation process.

Keywords: Semantic similarity algorithms; cognitive models; automated tools; Web design; usability evaluation.

1. Introduction

An exponential growth in the number of available websites and applications has been observed for almost twenty years now. However, the design of an effective website that meets user needs still remains a hard objective to achieve. Adoption of systematic user centered approaches and tools to support the website design and evaluation process has been suggested as a means for tackling the problem.

In order to gain insight in the design process, Ivory[1] summarized various studies of the adopted practices by professional Web designers. Several approaches are reported, including paper sketches, site maps depicting the organization of a site, storyboards depicting representative tasks and non-interactive graphical layouts. Results from a study with 169 Web practitioners unveiled that 36% always follow guidelines, 52% adopt approaches to optimize webpage download, 21% examine accessibility and 28% always conduct some form of usability evaluation testing[2]. However, only 15% report that they always use automated evaluation tools since they think that they have limited functionality and are difficult to use. On this issue, practitioners also report their belief that the available tools do not help them to produce better sites. Such findings indicate that further work is required to produce useful, solid and easy to adopt tools in order to support the website design lifecycle.

Artificial intelligence (AI) methods and modeling techniques are gaining increasing attention in this context. The usefulness of such methods is significant, since the Web by its very nature is a self-organized and continuously evolving system. In addition, the Web infrastructure itself greatly facilitates the process of acquiring data related to its structure and usage. Thus, a plethora of AI techniques have been applied in Web modeling and analysis, such as network traffic, structure, search engines and Web usage behavior[3]. Despite its stochastic nature, studies[4, 5] suggest that the Web seems to impose some regular patterns of usage. Such studies increased the belief that Web usage behavior is ruled, at least to a certain extent, by some patterns and provided a fertile ground for investigating models of human behavior while browsing on the Web.

The latter hypothesis is of paramount importance for systematic understanding of human–web interaction, given that despite the long tradition of the Human Computer Interaction (HCI) field, the research area still suffers from lack of sound theoretical descriptive or prescriptive models that can lead to effective design tools. Indeed, this was a fundamental problem of the HCI field since its early days, 30 years ago. Card, Moran and Newell[6] point that "although modern cognitive psychology contains a wealth of knowledge of human behavior, it is not a simple matter this knowledge to bear on the practical problems of design". A recent development concerns modeling of user behavior while searching for information, a typical Web related activity.

The task of evaluating and improving the usability of websites can be daunting given the quantity of sites being produced, the frequency of updates and the size of many sites[7]. An increased level of automation in the usability evaluation process, based on the aforementioned models could be proven beneficial, especially to practitioners, since they function under strict time constraints[8]. In this context, an important aspect of the Web design lifecycle, that can greatly influence the user's experience with a website, is the elicitation of proper *information architecture*[9, 10]. Information architecture is the practice of structuring information (knowledge or data) for a purpose. Although there might be some Web users that have well-formulated goals and abundant relevant knowledge while browsing a website, frequently this is not the case. Therefore, proper structure of

information may greatly determine the efficiency of the user to find meaningful information in the context of his current goal.

The premise of realizing automated tools to support design and evaluation is evident since the early days of the HCI field. Initially, in their pioneering work[6] Card, Moran and Newell proposed a metaphor based upon a view of basic cognitive processes, such as short and long term memory. However, the large differences across contexts, led to results of poor generalization. A survey of the most prominent automated usability evaluated approaches is presented by Ivory and Hearst[11]. In this paper, we focus on more recent tools that support the design and evaluation of websites' information architecture based on validated models of goal-directed Web navigation behavior and advanced probabilistic methods. Typically, such tools semi-automate the evaluation of the quality of hyperlinks' descriptions, an issue of fundamental importance, since it seems to be a strong determinant of users' satisfaction, even more than proper content organization[12–14].

The rest of this paper is organized as follows: First, the background on modeling of goal-directed user behavior while interacting with a website is provided. Next, an overview of tools that employ to some extent AI techniques to support design and evaluation of websites based on the aforementioned models is presented. The presented tools cover two aspects of the Web design lifecycle, namely *information structure elaboration* and *links' appropriateness evaluation*. Some approaches may also substitute, at least to an extent, established HCI techniques such as Cognitive Walkthrough and Card Sorting. Other approaches[15] are aimed to partially substitute techniques such as eye tracking studies by providing predictions of users' distribution of attention. Finally, we discuss the conclusions, implications and future directions of the presented research area and tools. We argue that the development of HCI tools that exploit AI could substantially aid the Web design lifecycle and could help both practitioners and researchers to manage the available resources in a more efficient way.

2. Background: Models of Goal-directed Web Navigation

In their pioneering work on *Information Foraging theory*[16], Pirolli and Card applied ideas from optimal foraging theory to understand how human users search for information. The fundamental underlying assumption of the theory is that when searching for information, humans use inherent, cognitive foraging mechanisms that evolved to help our animal ancestors find food. The analogy to the food foraging behavior is the following: Animals use scent to assess the possibility of finding prey in the current area and guide them to other beneficial patches. Similarly, humans rely on various cues in the information environment to estimate how much useful information they are likely to get on a given path, and afterwards they compare the actual outcome with their expectations. According to the Information Foraging theory, Web users constantly make decisions on what type of information to look for, which information path to follow, whether to continue seeking information at a specific site or to move on to another site and when to stop their search. These decisions are triggered by an inherent cost–benefit analysis mechanism, through

which the user examines the provided information gain against the amount of effort required to obtain it.

Users become recipients of many and multiple "segments" of information, while exploring and searching for information in the Web. While navigating through different information clusters, users assess the appropriateness of following a particular path by considering a representation, usually textual or graphical, of the distant content. Furnas[17], coined the term "residue" to describe the hint that a representational object holds (e.g. a hyperlink) of what lays behind it. Residue definition was refined by Pirolli[18] as information "scent" and defined as a user's "(imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues, such as WWW links". The process of information scent's assessment is exemplary described by Withrow[19]: "The mechanisms supporting information scent likely draw on the semantic networks that are unique to each individual. The connections between nodes differ not only in terms of distance but also in strength, both of which represent our understanding of the concepts and their interrelationships. Each label or other visual cue on a website activates nodes in our networks that then activate connected nodes in a spreading activation pattern. As the activation spreads, it weakens, so that concepts further away from the point of origination receive less stimulation. Based on these patterns of activation we make our best judgment concerning the link to click when browsing".

Thus, information scent's assessment substantially influences the cost–benefit procedure and, as a result, the probability of following a particular hyperlink in a webpage[20–23]. On the contrary, when users no longer expect to find useful additional information, that is their information scent assessment is getting lower, they move to a different information source. A number of studies[20–22] validates the importance of proper hyperlinks' information scent. These studies suggest that users have lower success rates and require more time to complete their tasks when they are presented with navigation options of weak scent compared to high scent. In a relative study[13], it was found that for sites with good labels and links' descriptions there was no benefit to having one site structure over the other. One other study[24] investigated the effect of eight slightly different levels of information scent on users' interaction with a website. It was found that when scent increased, users were gradually getting more effective and efficient, had significantly more focused attention–allocation patterns (a finding consistent with the results obtained by Habuchi, Kitajima and Takeuchi[25]) and reported higher levels of selection confidence. Even small differences in the target link's scent could substantially affect users' behavior.

Scent-based Navigation and Information Foraging in the ACT–R architecture[a] (SNIF–ACT v1.0)[23] is the cognitive engineering model upon which information foraging builds to predict user behavior. It models how information scent is used in Web navigation, but makes the limiting assumption that all links from a Web page are attended

---

[a] ACT-R is a cognitive architecture that aims to define the basic and irreducible cognitive and perceptual operations that enable the human mind during task execution. (Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press).

prior to a decision about the next navigation action. Information scent serves as a key concept not only in Information Foraging theory but in other comparable approaches aimed to explain goal-directed Web navigation as well.

According to the *Comprehension-based Linked model of Deliberate Search (CoLiDeS)*[26–28], every action of the user is the result of a two-phase process, involving an attention phase and an action–selection phase. During the attention phase, the user creates a mental representation of the webpage by dividing it into a collection of subareas. Subsequently, the user focuses on a subarea that she/he believes is semantically closer to her/his goal (i.e. the area with the highest information scent). Following, during the action–selection phase, the user comprehends all the interface widgets in the focused subregion and chooses to act on the one whose description is perceived to be closer to her/his goal (i.e. the widget with the highest information scent).

Method for Evaluating Site Architectures (MESA)[13], is a cognitive engineered approach for calculating the navigation cost through alternative Web designs for a given task. It simulates users navigating through a website by modeling the interplay between the website's information structure (breadth x depth), the quality of the link labels (i.e. information scent), and human cognition limitations. MESA combines a threshold strategy with an opportunistic strategy to model Web navigation.

Recent studies have validated and improved the accuracy of these proposed models. *CoLiDeS+*[29] extends the CoLiDeS model by adding contextual information involved in making navigational decisions. *SNIF–ACT v2.0*[30] extends the initial version of the model by incorporating the Bayesian Satisficing Model in the evaluation of Web links. This model assumes that "instead of searching for the optimal choice, choices are often made once they are good enough based on some estimation of the characteristics of the environment"[30].

Other approaches emphasize that both cognitive and affective uncertainty influence information seeking behavior. The *Information Search Process (ISP)*[31] is a six stage model of the users' holistic experience in the process of information seeking. The basic premise of the theory is that affective aspects, such as uncertainty and confusion can influence relevance judgments as much as cognitive aspects.

3. Tools Survey

Given the discussed advances of theoretical and empirical research on understanding the cognitive processes taking place during search of information in a website, the next step is to proceed with development of tools for Web designers, usability specialists and researchers that permit the rapid exploration of hypotheses about complex interactions of user goals, user behaviors and website designs. For instance, despite the plethora of guidelines concerning the creation of appropriate link labels, it is crucial to offer an efficient and objective way to measure the appropriateness of hyperlinks' descriptions. In

line with this argument, various semantic similarity algorithms,[b] such as Latent Semantic Analysis (LSA)[32] and Pointwise Mutual Information for Information Retrieval (PMI–IR)[33], have been proposed as a computational model of information scent[34] and have been used to facilitate the task of evaluating information scent[20, 35].

In this section, we present a survey of tools that employ to some extent AI techniques in their models of human information interaction in order to support the design or evaluation of websites' information architecture. As discussed next, such tools with predictive and prescriptive power can greatly facilitate the design and evaluation of websites. The tools are presented in chronological order. The description of each tool is divided into the following subsections: a) overview, b) input required, c) algorithm employed, d) output produced, and e) validation studies.

3.1. *Bloodhound: A tool that employs agents to produce usability metrics*

3.1.1. *Overview*

Bloodhound[22] is a simulation-based tool, which automatically analyzes a website to produce a usability report and help the designer to identify navigability problems. The underlying model of Bloodhound is SNIF–ACT v1.0[23], which was briefly presented in section 2.

3.1.2. *Input required*

Bloodhound requires as input: one or more typical user tasks expressed as a series of keywords, the URL address of the target page for each task, and the URL address of the webpage in which exploration starts.

3.1.3. *Algorithm employed*

The tool uses the provided input to run a novel algorithm, namely Information Scent Absorption Rate (ISAR), which calculates the probability distribution of simulated users following links and backtracking, guided by information scent. First the algorithm, parses the website and produces an *adjacency matrix* of the hyperlink topology and a *content matrix* (word×document) in which each cell specifies how important a word is in each document according to its TF.IDF (Term Frequency–Inverse Document Frequency).[c] Then, it produces a matrix that contains proximal cue words associated with each link

---

[b] Semantic similarity algorithms are automated methods to estimate the semantic similarity between words, phrases or passages. These algorithms can be classified as either taxonomical, statistical or hybrid[34]. Taxonomical approaches use manually created lexical databases to derive a quantitative measure of similarity between terms, such as path length between two node–words. In statistical techniques, semantic relationships between terms are captured from the probability of their co–occurrence in a large collection of documents. Hybrid methods attempt to combine taxonomies of concepts with statistical properties of a text corpus.

[c] TF–IDF (Term Frequency–Inverse Document Frequency) is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus.

(*link cues matrix*) and a vector of keywords for each user goal defined by the designer (*goal vector*), both of which are weighted according to the importance of each word provided by the context matrix. For a given goal, the scent of each link is calculated by multiplying the link cues matrix with the corresponding goal vector. The product of this matrix multiplication is stored as a *proximal scent matrix*. Each entry $S(i,j)$ in the scent matrix is the calculated probability of the likelihood that a user will surf from page i to page j, given that she/he has the specified information goal. A modified version of this scent matrix, which also models users backtracking behavior, is used with a network flow algorithm called Spreading Activation[36] to generate a predicted usage log. The latter can be used to extract simulated user paths and infer the usability of a Web site.

3.1.4. *Output produced*

The output of the tool is presented in the form of a usability report which contains metrics such as: a) the success rate of the agents simulating users for each specified task, b) high traffic pages per task, which are pages that the agents visited most and therefore require extra care in their design, c) frequency of each wrong page reached by the simulated users, which allows the designer to see where people would go if they would get confused by the navigation options, and d) an overall rating for the website based on the average success rate of the agents for all the provided tasks and research-based heuristics of actual users' success rates (e.g. Ref. 37).

3.1.5. *Validation studies*

A study[22] validated the accuracy of Bloodhound predictions against human performance data. The study involved 244 users performing 32 tasks of varying difficulty for four sites. Participants' web navigation behavior was monitored. The frequency of webpages accessed by the users were compared to the predictions derived using Bloodhound. It was found that "in nearly all cases, Bloodhound was able to produce click streams that moderately correlate with user data, and in a third of the time, Bloodhound actually produced click streams that correlate strongly with user streams"[22]. In addition, the authors reported that Bloodhound appeared to be sensitive to the query keywords used to formulate the goal and that future research should investigate ways of producing keywords that capture the domain knowledge of the task; a potential issue for most of the tools presented in this paper.

3.2. *MESA: Method for evaluating site architectures*

3.2.1. *Overview*

Method for Evaluating Site Architectures (MESA)[13] simulates users navigating through a website by modeling the interplay between the information structure of a website, the quality of the link labels, and human cognition limitations (e.g. the model only focuses and evaluates one link phrase at a time). Although MESA has not been implemented in

the form of an operational design tool, it has been included in this survey due to its highly computational nature.

3.2.2. *Input required*

MESA requires as input: the breadth×depth information structure of the website, the perceived relevance of each link to the targeted information (i.e. information scent), and the target webpage–node in the provided information structure. It is worth mentioning that the developers of MESA argue that the information scent of each link can be quantified either by involving a number of human raters (experts or users), or by using semantic similarity algorithms to estimate the semantic similarity between a textual description of a goal and each link label.

3.2.3. *Algorithm employed*

MESA models a user navigating a website as an agent that follows a combination of a threshold strategy with an opportunistic strategy. The threshold strategy involves selection of the first evaluated link that exceeds a predefined, scent threshold. The opportunistic strategy lowers this threshold to a predefined value if none of the available links exceeds the original threshold and there is at least one marginally relevant link. Subsequently, the simulated user rescans the page and either visits a link that exceeds the new, lowered threshold or backs to the parent page. For the rest of the visits, the current, lowered, threshold value is used. In this way, the pages visited by the agent depend on the previously visited pages and evaluated links that have shaped the current threshold value. MESA reflects the limitations of human working memory by not remembering the previous threshold value after a certain amount of page visits. The MESA algorithm does not consider the layout of the webpages. However, it randomizes the order in which links are evaluated and averages the simulated performance to achieve results that are neutral to the position of links in the webpages. The developers of MESA argue that "any understanding of how page layout and design affect the user's focus could eventually be incorporated"[13] into the model.

3.2.4. *Output produced*

MESA produces a prediction of the mean navigation time (averaged across a specified number of simulations) for the provided information structure and link label quality. In addition, it produces a trace of all the link evaluations, link selections and back actions of the simulated users.

3.2.5. *Validation studies*

Miller and Remington investigated the validity of MESA through a series of simulations[13] comparing MESA predictions with human participants' observed navigation times. The experimental websites used were constructed using items and categories found in a discount department store. Hyperlinks' information scent values

used in the MESA simulations were provided by three judges. A total of 45 University students performed eight navigation tasks in three different structures (breadth×depth) of the websites and their behavior was monitored and then compared to the predictions of the model. The model was able to account for 70.7% of the variance in the human performance data when the variance in the ratings provided by the judges was also taken into account. As a result, Miller and Remington[13] argued that MESA could be used to test information structures and make design decisions. They also found that the quality of link labels is a greater predictor of navigation times than the structure of the pages; the targets with the best link labels were found faster that those with poor labels regardless of the structure.

3.3. *ACWW: Automated cognitive walkthrough for the Web*

3.3.1. *Overview*

Automated Cognitive Walkthrough for the Web (ACWW)[20] is a publicly available, web-based tool (Fig. 1) that supports practitioners to analyze one or more webpages for problems that may hinder successful navigation. The tool identifies four types of usability problems, described in the following. In addition, it produces a prediction of the mean total number of clicks on a webpage that a user would require to select the correct link as a measure of the ease of navigation. ACWW is a first step towards the automation of the *Cognitive Walkthrough for the Web* method (CWW)[38], a transformation of the *Cognitive Walkthrough*[d] usability inspection method for website evaluation. The underlying model of ACWW is CoLiDeS[26–28], which was briefly presented in section 2. However, many of the processes suggested by CoLiDeS have to be manually performed; most notably the parsing of a webpage into subregions and focusing of visitor attention.

3.3.2. *Input required*

To use ACWW, the designer needs to provide manually the following information: a) a detailed, narrative description (100–200 words) of each typical user goal, b) heading labels of all the manually identified subregions on a webpage, c) text labels of the links included in each subregion. Next, the designer needs to specify the correct region heading(s) and link(s) for each goal and define a set of parameters that relate to the information scent calculation algorithm employed by the tool, such as topic spaces representing the reading level of the modeled user group and text–elaboration options.

3.3.3. *Algorithm employed*

The algorithm of ACWW is based on indices provided by the LSA[32] semantic similarity algorithm. LSA parses suitable, large text corpora that represent a given user population's

---

[d] The Cognitive Walkthrough method is a usability inspection method used to identify usability issues in interactive systems, focusing on how easy it is for new users to accomplish a given task with the system. The method is known for its ability to identify usability problems quickly, applied in early prototypes of the system.

understanding of words and produces a term–document matrix of each word's frequency of occurrence. Subsequently, LSA applies Singular Value Decomposition (SVD), the mathematical generalization of factor analysis, to represent each document and each word as a vector of high dimensionality (typical 300 or more) that captures latent relationships in the document collection. LSA is actually exploiting the property of natural language that words with similar meaning tend to occur together. LSA produces an index that measures the degree of semantic similarity between any pair of texts by calculating the cosine of the corresponding two vectors. Similarly to a correlation coefficient, each cosine value lies between –1 and 1, where 1 represents that the texts have identical meaning and 0 represents no relationship between the texts.

ACCW uses this LSA index to measure both the semantic similarity between the goal description and the provided headings/links, and the semantic similarity of one heading/link to another. In addition, ACWW computes the LSA term vector length for all the headings/links, to evaluate if the user, as modeled by the collection of documents used to train LSA, has an adequate level of relevant background knowledge to be familiar with the headings and links. Next, the algorithm of ACWW applies a set of if–then empirically-validated rules on the calculated indexes and identifies four types of navigability problems, described in the following.

3.3.4. *Output produced*

ACWW identifies four types of navigability problems[20]: a) a *weak scent* link, referring to the situation "when a correct link is not semantically similar to the user goal and there are no other correct links that have moderate or strong similarity", b) an *unfamiliar* problem occurring "when typical users of the website lack sufficient background knowledge to comprehend a correct link or heading text", c) a *competing headings* problem arising "when any heading and its associated sub region is semantically very similar to the user goal but does not contain a correct link that leads to accomplishing the user goal", and d) a *competing links* problem occurring "when a correct or competing sub region contains one or more links that are semantically similar to the user goal but not on the solution path". The tool also produces the predicted number of clicks a user would require to select the correct link on a webpage for a particular typical goal based on the number and types of problems identified. All the results of the ACWW analysis are included in an Excel spreadsheet file, which is emailed to the evaluator (Fig. 1).

3.3.5. *Validation studies*

The accuracy of the ACWW tool was evaluated through a series of experiments[20, 38, 39]. In these experiments, participants were presented with a detailed description of a search goal for an encyclopedia article and were asked to find the appropriate link while browsing an experimental website of an encyclopedia with topic links grouped under generic headings. It was found that participants required more clicks to find the target webpage when it was located under unfamiliar links or weak scent links compared to cases in which the topic link was strongly related to the goal. In addition, it was found

that participants took more clicks if there were competing links on the webpage. By aggregating the results, a "mean predicted total clicks" formula was produced, which is reported by ACWW as a measure of problem severity. This formula was cross-validated in a separate experiment[20].
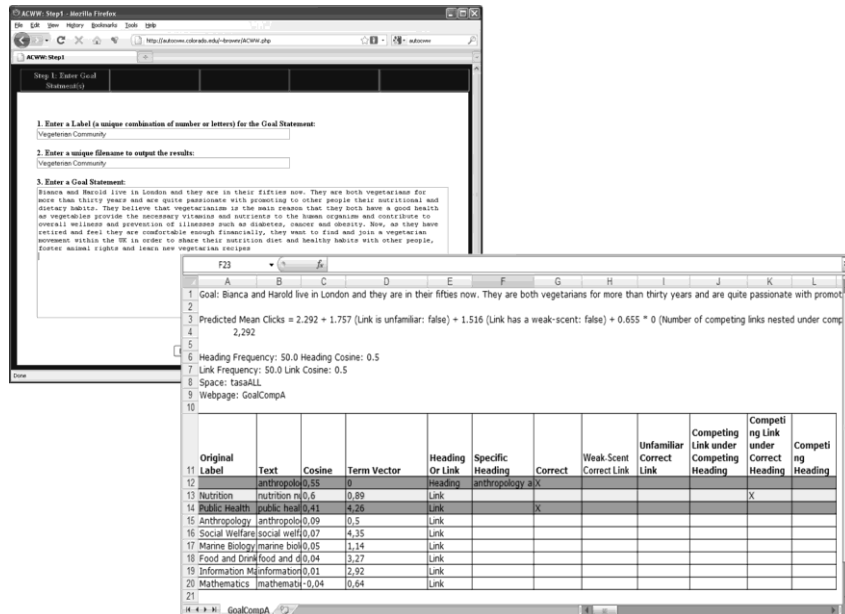

Fig. 1. Web-based interface of ACWW and an example of its output.

3.4. *ISEtool: InfoScent evaluator tool*

3.4.1. *Overview*

InfoScent Evaluator Tool (ISEtool)[35] has been proposed as a semi-automated tool to support the evaluation of information scent of single webpages, subsets of webpages or whole websites. Similarly to Bloodhound, ISEtool uses SNIF–ACT v1.0 as its underlying model of goal-directed web navigation. The tool supports an iterative evaluation process and offers a number of options to the designer which can be easily parameterized. A study[40], involving web designers, supported the flexibility, ease of use and overall usability of ISEtool.

3.4.2. *Input required*

In a typical usage scenario, the designer provides a textual description of a user goal, specifies the URL of a webpage, and selects a semantic similarity algorithm as the computational model of information scent (Fig. 2a). Currently, the LSA semantic similarity algorithm is the only available option. However, ISEtool is built on a software framework that allows the easy integration of alternative semantic similarity algorithms,

since there is much ongoing research targeted at understanding which one performs better in the context of information scent modelling[34, 41].
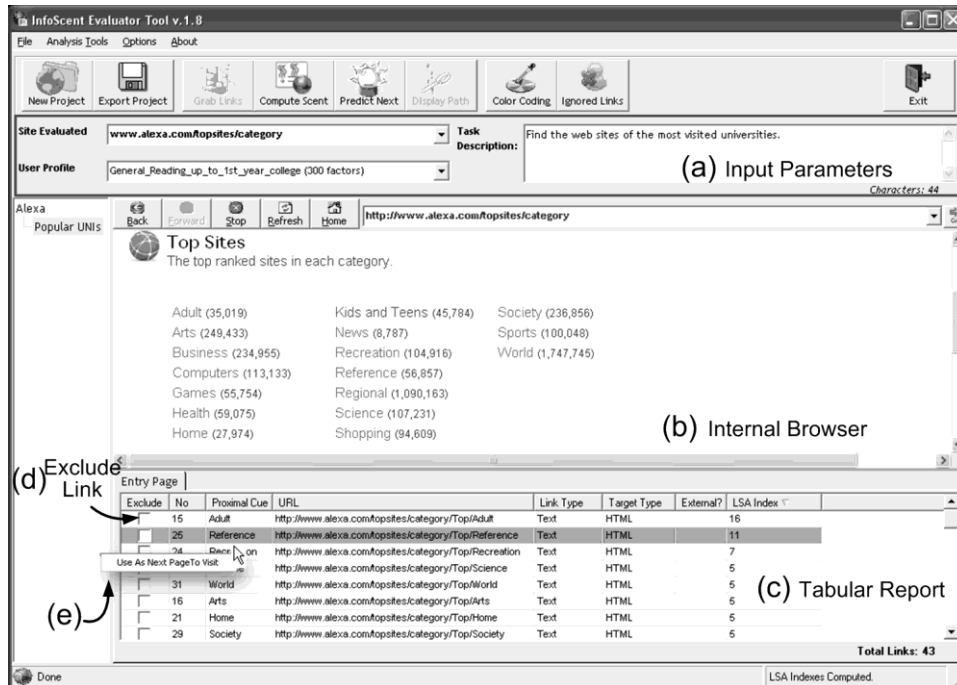

Fig. 2. Using ISEtool to evaluate the information scent in a webpage. Note: An identified weak scent problem is denoted in red (darker in grayscale).

3.4.3. *Algorithm employed*

ISEtool runs an automated analysis that combines a Web data extraction algorithm with the selected semantic similarity 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 adequately represented by their alternative texts as established usability and accessibility guidelines suggest. In cases in which these text equivalents are unavailable in the source code of the webpage, ISEtool asks from the designer to optionally provide such a textual description for the graphical hyperlinks. Next, the tool computes the semantic similarity between each available textual description for a link in the currently evaluated webpage and the provided textual description of a typical user goal. Currently, this is achieved by running a one–to–many analysis of the LSA algorithm in a transparent and automated way, using a publicly available service.[e] In addition, the algorithm of the tool finds and stores various other

---

[e] Service provided by the University of Colorado Boulder, Institute of Cognitive Science, http://lsa.colorado.edu

useful link properties, such as the type of the links' target (e.g. downloadable file or webpage) and whether a link is internal or external to the evaluated domain.

3.4.4. *Output produced*

The output of the tool is presented as a tabular report, including all the collected and calculated attributes of the links in the currently evaluated webpage (Fig. 2c). The evaluator can sort the results for any of these attributes. The default color coding in this tabular report visually groups the links into five scent–levels (weak, low, moderate, adequate, high), but it can be easily adjusted by the designer to serve different purposes. For instance, it can be used to identify weak–scent or competing links, based on empirically validated heuristics[20, 24]. In addition, the tool displays an embedded browser (Fig. 2b), which combined with the possibility to exclude any link from the output (Fig. 2d), can allow the designer to take also into account the visual layout of the webpage while interpreting the results. This embedded browser is synchronized 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. The evaluation process is iterative and the user of the tool can choose any of the available links as the next step (Fig. 2e). Finally, 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 visualizing the simulated user trail.

3.4.5. *Validation studies*

Two validation studies comparing the results of an ISEtool evaluation with data derived from human participants were conducted. In the first study, the scent–ratings collected in the context of a study[42] investigating the minimum number of raters required to reliably evaluate information scent were compared to the scent values produced by ISEtool using LSA and the "General reading up to first year college" profile. The dataset used was derived by 101 participants, who were asked to rate the semantic relevance of all the links in eight experimental webpages for eight associated goal descriptions on a 1–5 scale (1=poor relevance, 5=high relevance). The webpages presented navigation menus of actual websites related to specific tasks (e.g. buy a specific object). A high degree of correlation ($r = 0.58$, averaged across the tasks) was found between ISEtool scent values and scent–ratings of participants.

In the second study, 54 University students were asked to perform the same tasks on the same websites and their behavior was monitored.[f] Correlation analysis indicated a very high degree of correlation between the ISEtool identified scent level of the correct link and the observed participants' success ratio ($r=0.922$, $p=.001$), average time to select the first link ($r=-0.777$, $p=.023$), average self-rated confidence in the selection of the first link ($r=0.923$, $p=.001$), and number ($r=-0.853$, $p=.007$) and duration ($r=-0.798$, $p=.017$) of eye–observations on the links. The attention distribution and focusing patterns recorded using a 17'' Tobii T60 eye tracker provided further support for the validity of

---

[f] Data collected in a previous study[24] involving 19 University students were extended for this paper.

ISEtool (Fig. 3). In webpages that ISEtool classified as having higher scent levels, attention was mainly focused in the area containing the correct link, indicating a focused and efficient search. As ISEtool identified lower scent levels, attention was distributed across the rest of the available links, thus indicating an increasing level of uncertainty. This uncertainty is also depicted in other measures of participants' behavior. For instance, for the last three weaker scent tasks, the success ratio was on average 25%, whereas in the rest five higher scent tasks it was 70%.
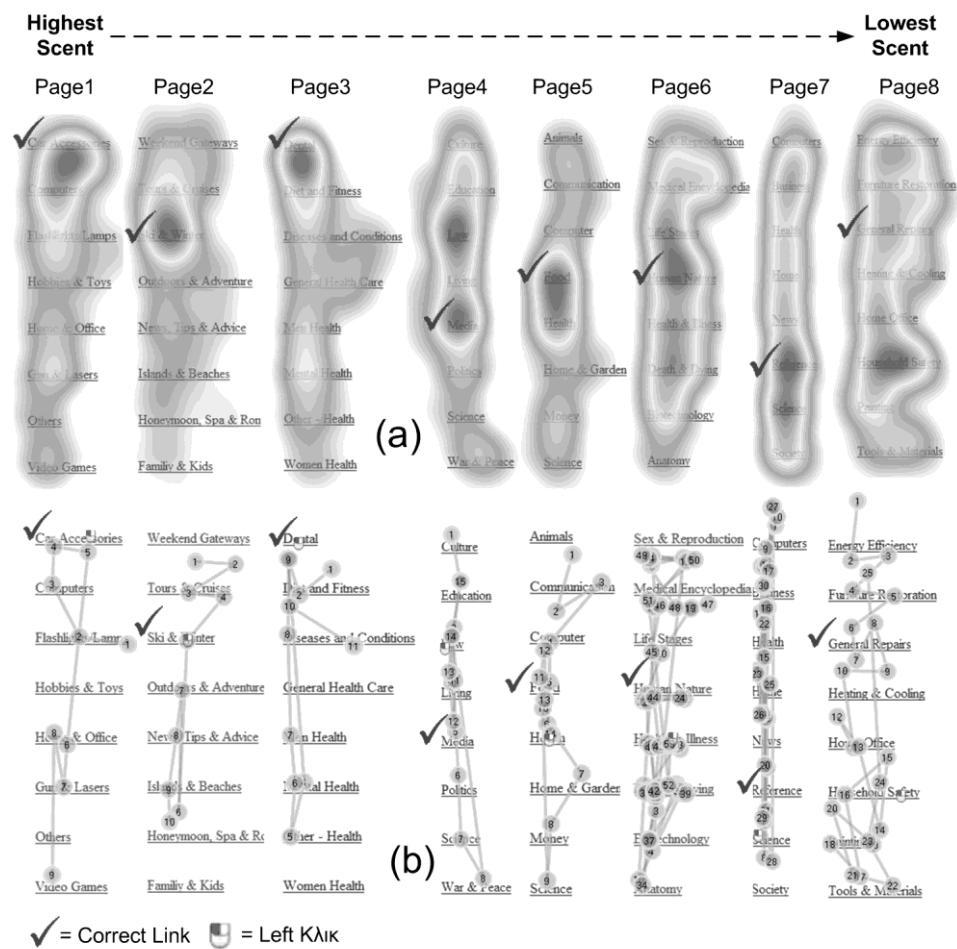

Fig. 3. (a) Heatmaps of participants' total duration of fixations for each webpage. (b) Representative gaze–plots of participants' sequence of fixations in each webpage. Note: Webpages are presented in descending information scent order based on their ISEtool mediated evaluation (left = higher scent, right = lower scent).

3.5. *CoLiDeS+: Combining link scent with navigation path relevancy*

3.5.1. *Overview*

CoLiDeS+[29] extends the CoLiDeS model by adding contextual information involved in making navigational decisions. In specific, CoLiDeS+ introduces the concept of *path adequacy*, which is the relevance of a navigation path to the user's goal, to account for the spatial cognition involved in Web navigation. According to CoLiDeS+ the user's selection of a link on a given webpage is not only influenced by the information scent of the currently available options, but also by the succession of links selected prior to reaching this specific webpage. Although CoLiDeS+ has not been implemented in the form of an operational design tool, it has been included in this survey due to its highly algorithmic nature.

3.5.2. *Input required*

Similarly to CoLiDeS, CoLiDeS+ requires as input a detailed task description of the user's goal, the heading labels of all the manually identified subregions on each webpage and text labels of the links included in each subregion. However, it does not require the specification of the correct region heading(s) and link(s).

3.5.3. *Algorithm employed*

CoLiDeS+, like CoLiDeS, parses each new webpage in several areas, focuses on the area perceived to be more relevant to the user goal, and uses LSA to estimate the relevancy of all the links in the focused subregion. However, it does not always choose the link with the highest information scent. Instead, it selects a link based on a comparison of the links' scent values in the focused sub region with the scent value of the previously selected link that led to the current webpage. If a link with greater scent value is found then it is followed. If not, the algorithm calculates a metric called path adequacy, which is the semantic similarity between a navigation path and the user's goal. If a link contributes to an increase in path adequacy compared to the adequacy of the current path then it is clicked. This models the situation in which a link may not have high scent by itself, but can still contribute to an increase in the scent of the entire path. Otherwise, an impasse is declared and CoLiDeS+ reacts with a 'next best' strategy similar to the opportunistic strategy of the MESA model, and if necessary backtracks. The algorithm stops when the current webpage contains the target information, which is not predefined as in CoLiDeS.

3.5.4. *Output produced*

CoLiDeS+ can be used to simulate users' link selections and backtracking behavior. In this way, 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)"[29].

3.5.5. *Validation studies*

Two studies[29] investigated the accuracy of CoLiDeS+ predictions. It was hypothesized that CoLiDeS+ would be able to simulate real users' navigation behavior and the navigation support generated based on simulations would have a positive influence on their navigation behavior and task performance. In both studies, CoLiDeS+ was first used to simulate web navigation tasks and identify successful paths and dead–ends. These simulation results were then used to implement a system that generated navigation support. Subsequently, participants were divided into two groups and were asked to perform these navigation tasks. Approximately half of the participants received the navigation support, whereas the rest did not. It was found that participants using the model-based navigation support system had better task performance, navigated in a more structured way, judged the system as more usable, and perceived themselves as less disoriented. In addition, based on their findings Oostendorp and Juvina[29] argued that "LSA (so far) is limited as a tool to model user's relevance assessments", but "users themselves are also limited in assessing relevance and they differ to a large extent among one another".

3.6. *CogTool–Explorer: Predicting Web user interaction*

3.6.1. *Overview*

CogTool–Explorer[43, 44], has been proposed as a tool to predict a Web user's goal-directed exploratory interaction with a website. Unlike the other tools discussed so far, CogTool–Explorer models computationally the effect of the layout position of links in a webpage to increase the accuracy of its predictions. The tool builds upon *CogTool*[45], a freely distributed user interface prototyping tool that can produce quantitative, model-based predictions of skilled performance time from tasks demonstrated on storyboard mock-ups of a user interface. However, the current version of CogTool (v.1.1.3) does not include a full implementation of the set of functionalities referred as CogTool–Explorer.

3.6.2. *Input required*

To use CogTool–Explorer the designer needs to provide a design prototype of the website, a textual description of a typical user goal, the URL address of the webpage in which exploration starts and the address of the correct webpage. The design prototype of the evaluated website can be produced either manually or automatically. In the first case, the designer works directly on the tool's drawing canvas dragging and dropping interactive elements from a palette of user interface widgets into frames that represent webpages, and specifying link transitions. In the automated alternative, the tool is directed to crawl an HTML implementation of the website and produce this design prototype (Fig. 4).
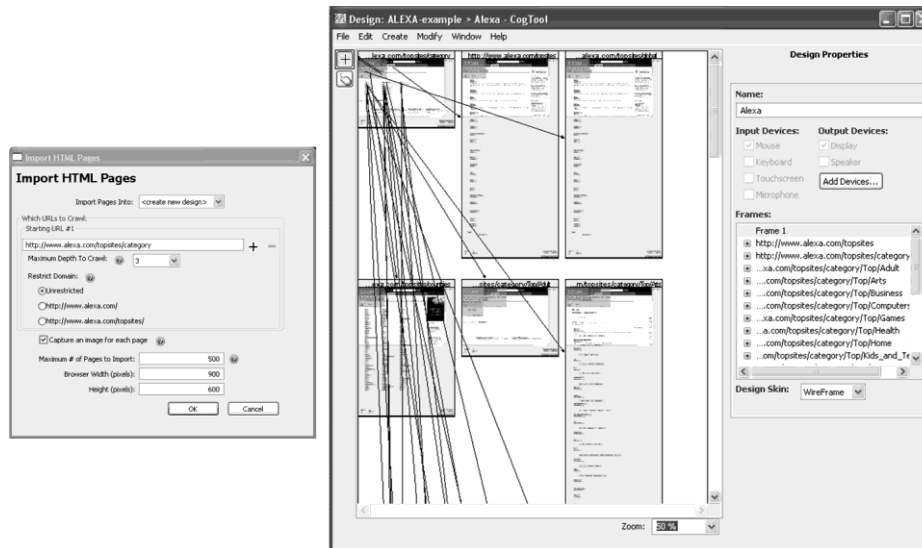
Fig. 4. Using CogTool–Explorer v1.1.3 "Import HTML Pages" functionality to automatically produce a representative model of a website.

3.6.3. *Algorithm employed*

CogTool–Explorer runs an algorithm that combines a visual search strategy adapted from the Minimal Model of Visual Search[46] with the SNIF–ACT v2.0 model to make predictions of goal-directed exploratory behavior. The algorithm moves a simulated eye's attention to a link, encodes the text label, and evaluates its semantic relatedness to the search goal using LSA. This simulated eye starts in the upper–left corner and looks at the link that is closest to the model's current point of visual attention (x–y coordinates), does not re-focus a link in the same visit and maintains its current point of visual attention when a new page is visited. After a link has been focused and its semantic relatedness to the provided goal evaluated, the algorithm chooses to either satisfice (i.e. select the best link evaluated so far) or continue evaluating additional links. This decision is not fixed and depends on the scent of the links that have been evaluated so far, which is moderated by the simulation parameters τ and k (τ is a noise function applied to the relatedness value that reflects one's variability when estimating the scent of a link for a goal and k reflects one's "readiness" to satisfice in a task). When a link is selected, the algorithm follows the link's transition to the next page and repeats the visual and scent evaluation cycles. It is worth noting that "each run of the model can be different because of the noise function, thus, the path of the model through the webpages on each run is analogous to predicting interaction choices of a single human trial"[44].

3.6.4. *Output produced*

CogTool–Explorer produces a predicted task time and a detailed script of the simulated user's cognitive, visual perception and motor processes. The results in the tool's output

can be viewed, saved for later reuse or modification, or exported to various common formats for further elaboration.

3.6.5. *Validation studies*

Teo and John compared the predictions of CogTool–Explorer to human performance data[44] from 22 of the tasks performed in a simulated encyclopedia presenting 32 topic links grouped in two columns (ACWW Experiment2[38]). In CogTool–Explorer, the design prototype of the evaluated experimental website was first automatically produced. Then the tool simulated the 22 tasks the same number of times as did the human participants. Although the magnitude of CogTool–Explorer's predictions was slightly larger than participants' performance, its predictions aligned with human performance. Also CogTool–Explorer captured the effect of target column; participants made significantly fewer clicks when the correct link was in the left than in the right column. In contrast, ACWW predicted far more clicks than the actual observations and did not predict any significant difference between search tasks across the two different columns upon which the links were grouped. This is due to the fact that CoLiDeS, the underlying model of ACWW, does not specify "how attention might move between several competing subregions and between links within a subregion"[44].

3.7. *AutoCardSorter: Automated card sorting tool*

3.7.1. *Overview*

Automated Card Sorting Tool (AutoCardSorter)[47] is a freely available tool that provides automated support for the design or evaluation of a website's structure. The tool combines semantic similarity measures, clustering algorithms and mathematical heuristics to simulate *Card Sorting*[9, 48]. Card Sorting is one of the main HCI methods used to elicit conceptual structures from participants and organize the content provided in a website in a way that increases findability. In a typical application of the method, 15–20 participants are asked to sort a stack of index cards, each containing a small description of the concepts to be grouped (e.g. webpages), into groups that make sense to them or pre-established groups specified by the designer. Open Card Sorting (with no pre-established groups) is used primarily to inform the design of new websites while the closed variation is used for adding content to an existing structure or validating the results of an open Card Sorting. However, the method is demanding in terms of time and human resources and can be daunting for both the participants and the designer when designing or evaluating large sites[48]. AutoCardSorter is offered as an automated alternative to open Card Sorting and can achieve proper structuring of a website, even when there are strict time and cost constraints or lack of the required expertise.

3.7.2. *Input required*

In a typical usage scenario of AutoCardSorter, the designer provides descriptions of the content items to be grouped (e.g. webpages of a site) (Fig. 5a), and specifies the

parameters of the algorithm employed by the tool; that is which semantic similarity and data clustering algorithm is going to be used (Fig. 5b). Currently, for the first parameter the only available option is LSA, whereas for the second parameter three types of hierarchical agglomerative clustering algorithms are offered: a) single–linkage, b) complete–linkage, and c) average–linkage. However, the tool is built on a software framework that allows easy integration of alternative semantic similarity algorithms or alternative unsupervised learning techniques.
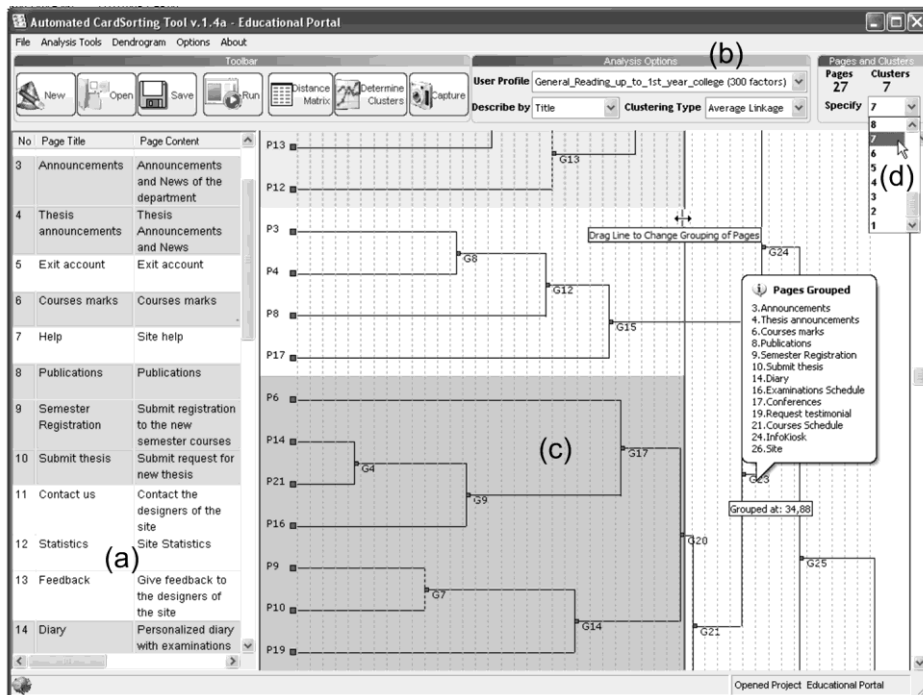

Fig. 5. Using AutoCardSorter to design the information structure of an educational portal.

3.7.3. *Algorithm employed*

AutoCardSorter runs an algorithm that first creates a matrix of the semantic similarities of each content–item's textual description to another, and then applies the selected clustering algorithm to produce groupings of semantically close items. The tool also implements mathematical techniques, such as Eigenvalue analysis, to determine the statistically-optimal number of categories.

3.7.4. *Output produced*

The output of the tool is an interactive dendrogram, in which the horizontal axis measures the semantic distance between groups of content–items; the more left a group is produced the more semantically closer its member are (Fig. 5c). The designer can cut off the dendrogram at various levels to produce different groupings by either dragging a vertical,

red line or by specifying explicitly the desired number of categories (Fig. 5d). In both cases, the tool reorganizes the results in real–time to present the groupings produced in different colors. In addition, AutoCardSorter provides two complementary ways to determine the statistically-optimal number of categories, namely Eigenvalue and Scree–plot analysis[49]. Finally, it should be mentioned that the tool offers increased input/output flexibility, since it allows both importing and exporting data either in XML or in common file formats (e.g. txt, csv).

3.7.5. *Validation studies*

Three independent studies[47] provided support for the validity and efficiency of AutoCardSorter. The studies compared the widely used open Card Sorting method and AutoCardSorter in the design or redesign of the information structure of websites for various domains and sizes. For each website, the tool-based method was first applied, followed by a Card Sorting experiment with 18 to 34 participants. The quality of the results produced by AutoCardSorter was investigated by three different types of comparisons with the results of the Card Sorting studies: a) similarity matrices correlation analysis, which compared the LSA similarity of each card–pair to the normalized frequency of these cards being placed together by participants, b) base–clusters comparison, an approach proposed[50] to objectively compare the distance of two dendrograms, and c) elbow-based navigation schemes comparison, which compared the structures obtained by the two approaches for the statistically-optimal number of categories. The total time required to design the website structure using each approach was also compared. It was found that the tool-based approach was on average 17 times faster compared to a typical Card Sorting study, providing at the same time highly similar results. In this paper, we also present the results of one new study that was conducted following the same methodology. Analysis of the results replicated the findings of the previous studies (Table 1), thus further supporting the validity and efficiency of AutoCardSorter.

Table 1. Summary of the results of four independent studies that compared in different contexts the validity and efficiency of AutoCardSorter against open Card Sorting with representative users.

	Previous Studies[47]			New Study
	Study 1 Health project site	Study 2 Educational portal	Study 3 Travel & tourism site	Study 4 News site
Similarity matrices correlation	0.80 (p < 0.01)	0.52 (p < 0.01)	0.59 (p < 0.01)	0.57 (p < 0.01)
Average amount of base–clusters separation	0%	7.5%	2.5%	7.5%
Elbow-based navigation schemes agreement	100%	93%	87%	89%
AutoCardSorter efficiency compared to Card Sorting	27 times faster	11 times faster	14 times faster	16.5 times faster

4. Discussion

Table 2 summarizes the main characteristics of the surveyed tools supporting design and evaluation of websites based on models of human information interaction. Most of the presented tools (Bloodhound, ACWW, ISEtool, CoLiDeS+, CogTool–Explorer) address the problem of creating link labels that facilitate scent following, whereas only two tools (AutoCardSorter, MESA) address the problem of appropriate information structure elaboration.

Table 2. Overview of the surveyed tools. Note: Tools freely available at the time this paper was written are denoted with an asterisk.

	Input	Output	Related model or HCI method	Algorithms or techniques used
Bloodhound[22]	• User goal (series of keywords) • Starting page URL • Target page URL	• Overall rating • Success per task (%) • High traffic pages • Number of wrong pages reached	• SNIF–ACT v1.0	• Information Scent Absorption Rate (ISAR)
MESA[13]	• Breadth × depth structure • Information scent of each link • Target webpage	• Mean predicted navigation time • Simulated cognitive user processes	• MESA	• Threshold-based strategy • Opportunistic strategy
*ACWW[20]	• User goal (detailed description) • Webpage subregions • List of headings/links • Correct heading/link	• Mean predicted clicks • Navigability problems: weak scent, competing headings & links, unfamiliar headings & links	• CoLiDeS • Cognitive Walkthrough	• LSA
*ISEtool[35]	• User goal (free text or persona) • Starting page URL	• Interactive tabular report • Navigability problems (parameterized)	• SNIF–ACT v1.0	• Web extraction algorithm • LSA
CoLiDeS+[29]	• User goal (detailed description) • Webpage subregions • List of headings/links	• Simulated cognitive user processes	• CoLiDeS+	• LSA
CogTool–Explorer[43]	• Goal (free text) • Website design prototype • Starting page URL • Target page URL	• Predicted task time • Simulated cognitive, visual perception, and motor processes	• SNIF–ACT v2.0	• Web crawler • LSA • Minimal model of visual search
*AutoCardSorter[47]	• Text description of content–items (e.g. webpages)	• Interactive dendrogram • Best-fit number of categories	• Card Sorting	• Hierarchical clustering • LSA • Eigenvalue analysis

AutoCardSorter appears to be the best tool to support the initial design of an information structure. The tool replicates open Card Sorting, a participatory design method which involves users to elicit a website's information structure. Extensive validation of the tool shows robust results, which in all reported cases were rather close to

the collected data from human participants. However, unlike open Card Sorting, AutoCardSorter does not provide any insight into the labels that should be chosen for the produced categories.

MESA is the only approach that models the interplay between label quality and information structure. A combined usage of AutoCardSorter and any of the link elaboration tools could also address this interplay. However, MESA appears to be the most efficient approach in terms of evaluating design alternatives for the breadth×depth information architecture of a website. Although MESA does not embed a computational model of label quality, the approach can be still efficient and effective using scent–ratings from 10 participants, as suggested by a recent related study[42]. The main disadvantages of MESA are that it requires the modeler to hard-code the simulated Web structure, and more importantly that it has not been implemented yet as a ready–to–use tool.

In terms of the tools proposed to support the design of link labels that facilitate scent–following, CogTool–Explorer appears to be the most complete modeling approach. Both Bloodhound and ISEtool do not model visual search and assume a global evaluation of all hyperlinks in a webpage, whereas both ACWW and CoLiDeS+ ask from the designer to manually parse the webpage into sub regions and do not specify "how attention might move between several competing sub regions and between links within a subregion"[45]. Thus, none of these tools addresses link positioning issues or more complex web navigation strategies, such as satisficing. In contrast, CogTool–Explorer is the only tool that simulates visual perception processes. Thus, CogTool–Explorer can provide greater insight into expected user behavior and produces more accurate predictions of users' performance[44]. Nevertheless, at the time this paper was written, the set of functionalities referred as CogTool–Explorer was ongoing work and there wasn't any freely available version of the tool. In addition, given that the tool builds upon CogTool, a rather complex user interaction modeling environment, its usability for web practitioners is questionable.

Currently, ISEtool and ACWW are the only fully implemented and freely available tools, which support the task of creating link labels that facilitate scent–following. ISEtool proposes an iterative information scent evaluation process, which real–world Web practitioners[40] found efficient, flexible and "close to a form that could be used by developers with little training". ACWW proposes a complete evaluation method that adapts the Cognitive Walkthrough inspection method for website evaluation. However, many of the processes suggested by CoLiDeS, the ACWW's underlying model of goal-directed web navigation behavior, have to be manually performed. In specific, the user of the tool has to manually divide the webpage into subregions, provide heading descriptions for these regions and add the descriptions of the links contained in each region. As a result, using ACWW to evaluate webpages with many hyperlinks or a whole website can be a quite challenging task. However, ACWW, unlike ISEtool, allows for the simultaneous evaluation of all links in a webpage against a set of different used goals. In this way, ACWW makes it easier to ensure that changing some link labels to resolve scent-related problems for a specific task will not create new navigability problems for a different task.

Despite their limitations, the surveyed AI tools can greatly facilitate the design or evaluation process of websites in a complementary way to existing HCI approaches. In specific, such tools could help to distribute resources in user-based evaluation approaches in a more efficient and effective manner. For instance, AutoCardSorter could be used to produce the information structure of a website, which is then evaluated in a user testing study along with other parameters affecting the user experience, such as credibility and aesthetics. In addition, the presented tools could be used to support the transition from user research to the design itself, which is a quite difficult task[51]. For example, one could use MESA to evaluate alternative breadth×depth information structures produced using data collected from an open Card Sorting study. The presented tools may be also used to resolve tradeoffs among guidelines. For instance, according to established usability guidelines link labels must be concise enough to support scannability but at the same time they need to be long enough to convey meaningful information. Any of the aforementioned scent evaluation tools could be used to evaluate a set of link labels with varying text length and choose the best compromise between scannability and information scent. Finally, the presented tools give practitioners the chance to experiment and form a better understanding of the related design issues.

5. Conclusions and Future Work

AI techniques have been successfully applied in various systems and contexts, such as text entry support and prediction for mobile devices[52]. Focusing on Web applications, AI techniques have been successfully used for personalization and user modeling, weblog analysis, semantic web agents, intelligent web search agents, and spam detectors.

In this paper, we presented a survey of AI tools primarily implementing models of goal-directed web navigation behavior to simulate expected users' behavior and support related design decisions. As derived from the presentation of their functionality, such tools could significantly assist various aspects of a website's design and evaluation lifecycle. Initial application and validation studies showed promising results, thus further strengthening our argument that practitioners can make use of this knowledge in assessing interaction design for the Web. Our understanding of the processes involved during web interaction and the models describing them, seem to evolve over time and the intelligent tools implementing them entering maturity, becoming useful for practitioners[40].

As more complete theories and cognitive models emerge, it is reasonable to expect continuous evolution and improvement in the utility and accuracy of AI tools for web design and evaluation. Currently, despite the encouraging results reported in the related validation studies, the completeness and accuracy of such model-based tools in non-experimental settings is still questioned. Future work should focus on the systematic study of web designers' practices as well as additional studies investigating the accuracy of the developed tools in real–world contexts. Also, as far as semantic similarity algorithms are concerned, the problems of directly comparing their results in specific contexts are difficult to handle due to the scarcity of freely available implementations and

the differences in the corpora used. Therefore, it remains unclear to what extent such metrics can accurately model human relatedness judgments[53]. In addition, current tools do not provide adequate support to objectively evaluate the appropriateness of graphical hyperlinks. AI techniques that can infer semantic information from images[54] could be used as a potential solution to this problem. Even affective aspects of the user interface such as credibility and aesthetics could be modeled using appropriate AI techniques[55]. Finally, further research is required towards studying phenomena relating with goal reformulation during web navigation and the identification of possible extensions to the existing models.

Concluding, despite the advantages of the presented AI tools, the value of established user-based techniques and use of existing evidence-based knowledge in the form of guidelines, the traditional tools of developers are not to be overlooked. Instead, such model-based tools could be used as a complementary part of an iterative design process in conjunction with existing HCI approaches, allowing deeper exploration of alternative solutions. More importantly, such tools provide bridges between cognitive modeling research and web design practice and make it more likely for practitioners to embrace and employ them in order to improve the usability of their websites, even when there are strict time and cost constraints or lack of the required expertise.

6. References

1. M. Y. Ivory, *Automated web site evaluation: Researchers' and practitioners' perspectives* (Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003).
2. A. Chevalier and M. Y. Ivory, Web site designs: influences of designer's expertise and design constraints, *Int. J. of Human–Computer Studies* **58**(1) (2003) 57–87.
3. P. Baldi, P. Frasconi and P. Smyth, *Modeling the internet and the web: Probabilistic methods and algorithms* (John Wiley and Sons, 2003).
4. B. Huberman, P. Pirolli, J. Pitkow, and R. Lukose, Strong regularities in world wide web surfing, *Science* **280**(5360) (1998) 95–97.
5. A. Cockburn and B. McKenzie, What do web users do: An empirical analysis of web use, *Int. J. Human–Computer Studies* **54**(6) (2001) 903–922.
6. S.K. Card, T.P. Moran and A. Newell, *The psychology of human–computer interaction,* 1st edn. (Lawrence Erlbaum Associates, 1983).
7. T. Brinck and E. Hofer, Automatically evaluating the usability of web sites, in *Workshops of CHI'02* (2002).
8. E. Chi, P. Pirolli and J. Pitkow, The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a website, in *Proc. CHI 2000* (ACM Press, The Hague, The Netherlands, 2000), pp. 161–168.
9. L. Rosenfeld and P. Morville, *Information architecture for the world wide web* (O'REILLY publications, Sebastopol, California, 1999).
10. P. Morville, *Ambient findability: What we find changes who we become* (O'REILLY publications, Sebastopol, California, 2005).
11. M. Y. Ivory and M. A. Hearst, The state of the art in automating usability evaluation of user interfaces, *ACM Computing Surveys* **33**(4) (2001) 470–516.
12. M. L. Resnick, and J. Sanchez, Effects of organizational scheme and labeling on task performance in product-centered and user-centered retail websites, *Human Factors*, **46** (2004).

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

14. M. Katz and M. Byrne, Effects of scent and breadth on use of site specific search on e-commerce websites, *ACM Trans. on Computer–Human Interact.* **10**(3) (2003) 198–220.

15. B. Stone and S. Dennis, Using LSA semantic fields to predict eye movement on web pages, in *Proc. of the 29th annual conference of the cognitive society*, pp. 665–670.

16. P. Pirolli and S. Card, Information Foraging, *Psychological Review* **106**(4) (1999) 643–675.

17. G. Furnas, Effective view navigation, in *Proc. of CHI'97* (Atlanta, USA, 1997), pp. 367–374.

18. P. Pirolli, Computational models of information scentfollowing in a very large browsable text collection, in *Proc. of CHI'97* (Atlanda, Georgia, USA, 1997), pp. 3–10.

19. J. Withrow, Do your links stink?, *American Soc. for Inform. Sci. Bulletin* (June 1) (2002).

20. 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 *Proc. of CHI'05* (ACM Press, Portland, Oregon 2005), pp. 31–40.

21. S. K. 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 *Proc. of CHI'01* (ACM Press, USA, 2001), pp. 498–505.

22. 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 *Proc. of CHI'03* (ACM Press, Ft. Lauderdale, Florida, USA, 2003), pp. 505–512.

23. P. Pirolli, *Information foraging theory: Adaptive interaction with information* (Oxford University Press, USA, 2007).

24. N. Tselios, C. Katsanos and N. Avouris, Investigating the effect of hyperlink information scent on users' interaction with a web site, in *Proc. of Interact 2009*, LNCS 5727 Vol.2 (Springer Verlag, Uppsala, Sweden, 2009), pp. 138–142.

25. Y. Habuchi, M. Kitajima and H. Takeuchi, Comparison of eye movements in searching for easy–to–find and hard–to–find information in a hierarchically organized information structure, in *Proc. of ETRA 2008* (ACM Press, Savannah, Georgia, USA, 2008), pp. 131–134.

26. M. Kitajima and P. G. Polson, A comprehension-based model of exploration, *Human–Computer Interaction* **12**(4) (1997) 345–389.

27. M. Kitajima, M. H. Blackmon and P. G. Polson, Cognitive architecture for website design and usability evaluation: Comprehension and information scent in performing by exploration, in *Proc. of HCI International* (2005).

28. 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!* (Springer–Verlag), pp. 357–373.

29. H. V. Oostendorp and I. Juvina, Using a cognitive model to generate web navigation support, *In Int. J. of Human.–Computer Studies* **65**(10) (2007) 887–897.

30. W. Fu and P. Pirolli, SNIF–ACT: a cognitive model of user navigation on the world wide web, *Human–Computer Interaction* **22**(4) (2007) 355–412.

31. C. C. Kuhlthau, *Seeking meaning: A process approach to library and information services*, 2nd edn. (Libraries Unlimited, Westport, CT., 2004).

32. 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) (1997) 211–240.

33. P. D. Turney, Mining the web for synonyms: PMI–IR versus LSA on TOEFL, in *Proc. of Machine Learning: ECML 2001* (Springer Berlin/Heidelberg, 2001), pp. 491–502.

34. I. Kaur and A. J. Hornof, A comparison of LSA, WordNet and PMI–IR for predicting user click behavior, in *Proc. of CHI'05* (Portland, Oregon, USA, 2005), pp. 51–60.

35. C. Katsanos, N. Tselios, and N. Avouris, InfoScent evaluator: a semi-automated tool to evaluate semantic appropriateness of hyperlinks in a web site, in *Proc. of OZCHI'06* (ACM Press, Sydney, Australia, 2006), pp. 373–376.

36. J. R. Anderson and P. Pirolli, Spread of activation, *J. of Exp. Psychology: Learning, Memory and Cognition* **10**(1984) 791–798.

37. J. Nielsen, Did poor usability kill e-commerce?, *Nielsen's Alertbox* (August 19, 2001), http://www.useit.com/alertbox/20010819.html.

38. M. H. Blackmon, P. G. Polson, M. Kitajima and C. Lewis, Cognitive walkthrough for the web, in *Proc. of CHI'02* (ACM Press, Minneapolis, Minnesota, USA, 2002), pp. 463–470.

39. M. H. Blackmon, M. Kitajima, and P. G. Polson, Repairing usability problems identified by the cognitive walkthrough for the web, in *Proc. of CHI'03* (ACM Press, Ft. Lauderdale, Florida, USA, 2003), pp. 497–504.

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

41. 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 *Proc 8th RIAO* (2007).

42. C. Katsanos, N. Tselios, and N. Avouris, Are ten participants enough for evaluating information scent of web page hyperlinks?, in *Proc. of Interact 2009*, LNCS 5726, Vol.1 (Springer Verlag, Uppsala, Sweden, 2009), pp. 419–422.

43. L. Teo, B. E. John and P. Pirolli, Towards a tool for predicting user exploration, in *Extended Abstracts of CHI'07* (ACM Press, San Jose, CA, USA, 2007).

44. L. Teo and B. E. John, CogTool–Explorer: Towards a tool for predicting user interaction, in *Extended Abstracts of CHI 2008* (ACM Press, Florence, Italy, 2008).

45. B. E. John, K. Prevas, D. D. Salvucci and K. Koedinger, Predictive human performance modeling made easy, in *Proc. of CHI '04* (ACM Press, Vienna, Austria, 2002), pp. 455–462.

46. T. Halverson and A. J. Hornof, A minimal model for predicting visual search in human–computer interaction, in *Proc. of CHI'07* (ACM Press, San Jose, CA, 2007), pp. 431–434.

47. C. Katsanos, N. Tselios and N. Avouris, Automated semantic elaboration of web site information architecture, *Interacting with computers* **20**(6) (2008) 535–544.

48. D. Mauer and T. Warfel, Card sorting: a definitive guide, *Boxes and Arrows* (April 7) (2004), http://www.boxesandarrows.com/view/card_sorting_a_definitive_guide.

49. R. B. Cattell, The scree test for the number of factors. *Multivariate Behavioral Research*, **1**(1966) 245–276.

50. T. Tullis and L. Wood, How many users are enough for a card–sorting study?, in *Proc. of UPA'2004* (Minneapolis, MN, 2004).

51. R. Sinha and B. Boutelle, Rapid information architecture prototyping, in *Posters of DIS04* (ACM Press, 2004), pp. 349–352.

52. M. Maragoudakis, N. Tselios, N. Fakotakis and N. Avouris, Improving SMS usability using Bayesian networks, in *Methods and Applications of Artificial Intelligence*, eds. I. P. Vlahavas and C. D. Spyropoulos (Springer–Verlag, Berlin,2003), pp. 179–190.

53. V. D. Veksler and W. D. Gray, Desiderata for comparing measures of semantic distance, in *Posters of CogSci2006* (Cognitive Science Society, Vancouver, BC, Canada, 2006).

54. R. Tadeusiewicz and M. R. Ogiela, Artificial intelligence techniques in retrieval of visual data semantic information, in *Proc. of AWIC2003* (Springer–Verlag, Madrid, Spain, 2003), pp. 18–27.

55. E. Papachristos, N. Tselios and N. Avouris, Inferring relations between color and emotional dimensions of a web site using Bayesian networks, in *Proc. of Interact 2005* (Springer Verlag, Rome, Italy, 2005), pp. 1075–1078.