Beyond user centered design: A web design approach based on information foraging theory

Nikolaos Tselios[1], Christos Katsanos[1] and Nikolaos Avouris[1]

[1] Human-Computer Interaction Group, University of Patras, Greece,
{nitse, ckatsanos}@ece.upatras.gr, avouris@upatras.gr

Abstract. In this paper we describe a novel, tool-based approach for designing and evaluating web based information environments. The approach is inspired by information foraging theory which led to realization of appropriate tools to evaluate hyperlink's semantic appropriateness and automate the process of information architecture construction with limited participation of end users. We argue that it is crucial to provide increased level of automation in the design and evaluation lifecycle and the presented model-based tools could offer such a characteristic. In addition, they reinforce established techniques which are often neglected due to limited resources.

Keywords: information foraging, hyperlink's semantic appropriateness, card sorting

1 Introduction

User-centered design (UCD) is a general term for methods and techniques involving users in the design of computer applications. Users according to UCD participate throughout the design process as partners in the design and evaluation, taking active part in requirements gathering, participatory design and usability testing. However such practices are often accused for lack of sound theoretical underpinnings, while on the other hand, involvement of users throughout the process is not always considered beneficial, slowing down the design and reducing innovation, in particular in today's complex environments.

In this paper we present some experience that has been gained by applying a model- based approach for key design decisions of web sites, based on Information Foraging theory. We argue that the proposed approach, in some aspects of a web site's design and evaluation lifecycle, is more efficient and effective than user based approaches. First a brief introduction to the theoretical framework of our approach in attempted in the following. The core idea of Information Foraging theory [1] is that information seeking and retrieval is analogous to the food foraging mechanisms of animals. The user adapts her behaviour (or even the structure of the system interface) in order to maximize her information gain per specific effort (unit cost). The theory refers to cognitive activities associated with assessing, seeking, and handling information sources, typical tasks in most web applications. While the users search for

information in the web, they are exposed to many "segments" of information. While navigating through them, they assess the appropriateness of following a particular path by considering a representation, usually a textual or graphic description, of the target content. This representation was defined by Pirolli and Card as information "scent" and elaborated as a user's "(imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues, such as web links" [1]. If the scent of information is sufficiently pungent, users will be generally more able to find, evaluate, and use materials relevant to their information goals.

CoLiDeS (Comprehension-based Linked model of Deliberate Search) is a similar comprehension-based cognitive model explaining expected users' information foraging in the web as a two phase process [2]. During the *attention* phase the user creates a mental representation of the page by dividing it into sub-areas and she focuses on a sub-area that is believed to be semantically closer to her goal. Next, during the *action-selection* phase, the user comprehends all the items in the focused sub-region and chooses to act on the one whose description is perceived to be closer to her target. It is of particular importance to model the expected user behaviour during the design of a web site and to organize information and information descriptions in a way that will facilitate the users' information seeking tasks.

The objective of the approach presented in this paper is to transform the theoretical notions of information foraging to an effective set of design methods and tools. Predictive models of human behaviour can be used as a way to understand and interpret the behaviour and searching strategies of users navigating a web site. As a result, we distinguish the mediating and aligning role of information foraging theories, as a tool to properly take into account information architecture issues and encapsulate usability research towards formulation of our design and evaluation approach. By building on this notion, we propose a tool-based approach that could inform the design and evaluation process.

We argue that two distinct types of impasses could be tackled. One type of problems can be detected at a *Web page level* due to inadequate scent emitted from the proximal cues in Web pages. Another type of impasse can be traced at a *Web site level*, and is caused by flaws in the information architecture of Web-based systems.

At a Web page level the categories of possible problems are:

- Weak Scent Cues, when a correct link is not semantically similar to the user's goal and there are no other correct links that have moderate or strong similarity. The possibility the users encounter inadequate scent on Web pages that use short and/or ambiguous link labels is quite common [3].
- Competing Cues, when a page contains one or more links that are semantically similar to the user's goal but do not lead to the targeted information, thus resulting to frustration and in a loss of orientation.
- Unfamiliar Cues, when the users do not comprehend a link due to varying knowledge and cultural backgrounds – a common phenomenon in web sites

As far as Web site level impasses are concerned, they are mainly observed due to too abstract starting nodes, misclassified nodes in the web site architecture and too long information paths (usually having the intrinsic problem of too abstract starting nodes and insufficiently specific-end points). We should stress that lack of adequate information structure cannot be tackled using a powerful search engine inside a web

site. When clear labels and prominent navigation options are established, users tend to browse rather than search. Searching is no faster than browsing in this context [4].

In order to validate design decisions both at the web page level and web site level, we have developed two tools to automate the process of evaluating the quality of the web page's hyperlinks and the content architecture. Since the traditional user-centered design process is largely empirical even nowadays, we argue that the presented here tools can greatly improve the design and evaluation lifecycle. Two tools are presented next: the *InfoScent Evaluator* and *AutoCardSorter*. The first automatically evaluates semantic appropriateness of the hyperlinks' descriptions and the second proposes a structure for the content of a complex information hyperspace. The utility of the tools is twofold. To explore alternative designs and solutions ("what-if" scenarios) and to support formative and summative usability evaluations, reducing the need to involve actual users, thus reducing costs and accelerating the design process.

2. Evaluating Hyperlink Semantic Appropriateness: The ISEtool

Despite the plethora of guidelines concerning the appropriate ways to provide efficient and effective link labels, it is crucial to offer an increased level of automation in the design and evaluation process of web applications and provide an objective measure of the appropriateness of the hyperlink descriptions. One such tool that we have implemented, is one that automatically evaluates the semantic appropriateness of the hyperlinks' descriptions presented in a web site [5]. This tool, named *InfoScent Evaluator (ISEtool)*, is capable of simulating users' activity and interaction with a web site according to Information Foraging theory. The basic underlying assumption in our tool is that users have some information goal and their surfing patterns are guided by information scent. The tool attempts to quantify the concept of information scent, using Latent Semantic Analysis (LSA, [6]).

LSA was developed to mimic human ability to detect deeper semantic associations among words, phrases or whole sentences. LSA builds a semantic space representing a given user population's understanding of words sentences, and whole texts from documents that these users are likely to have read. The meaning of a word, sentence or any text is represented as a vector in a high dimensional space, typically with about 300 dimensions. The degree of semantic relatedness or similarity between any pair of texts, such as the description of a user's goal and a link label on a webpage, 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.

LSA also served as a computational model of information scent in ACWW, a conceptual artefact based on Cognitive Walkthrough method and the CoLiDES theory [2]. However, despite the fact that the results presented seem to be very promising, lack of integration of useful functions - such as automatic grabbing of links, storage of results, automatic prediction and walkthrough of a user's path - inspired us to create a more complete and automated tool described briefly in the following.

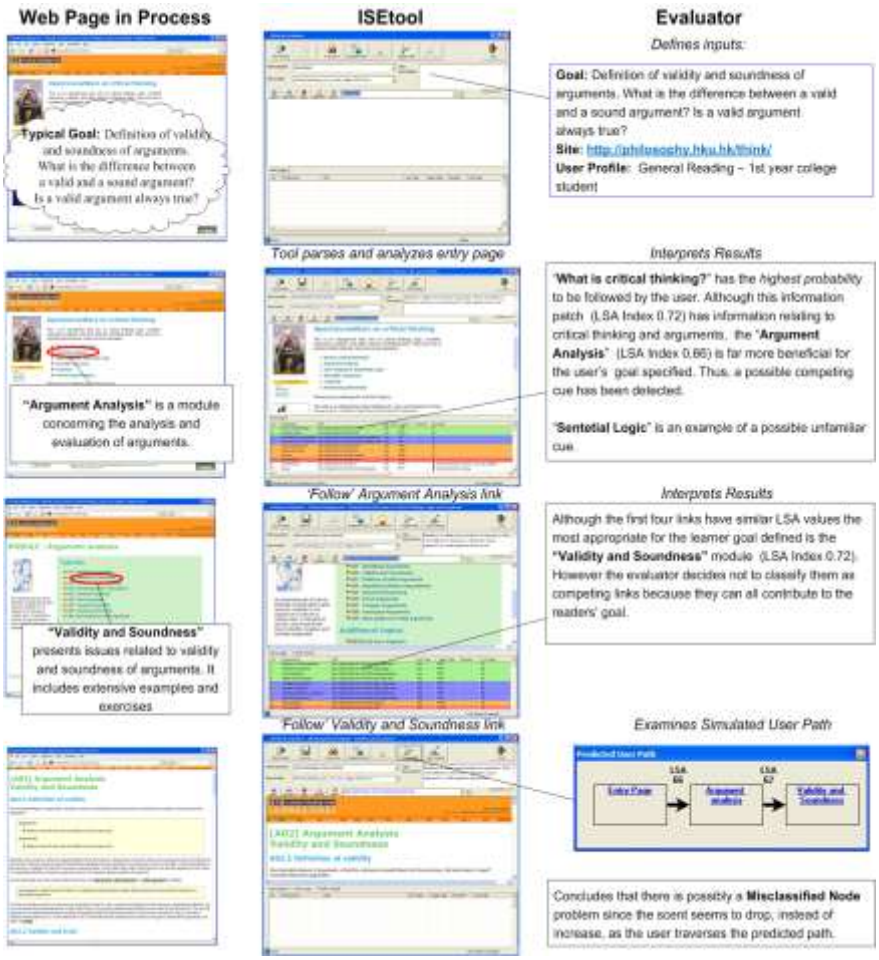

Fig. 1. A representative example of using *ISEtool* to analyze a web page.

The tool offers a number of options to the designer and can be used during web design and evaluation to identify some of the aforementioned web page and web site problems that are caused due to inadequate scent emitted from the proximal cues of hyperlinks. A typical usage scenario of ISEtool is the following (Figure 1): First, the designer defines the profile of the users by choosing one of the available LSA semantic spaces and describes a typical user's goal. Next, the ISEtool renders and parses the page that the designer has defined as the user's entry point in its embedded browser. The tool also discriminates automatically the external and internal links. The calculation of information scent for all links is achieved by running a one-to-many analysis of the LSA algorithm in a transparent and automated way using the free service located at http://lsa.colorado.edu/. LSA computes and returns the semantic similarity (LSA index) of the user's goal against all the proximal cues of the links. The calculated information scent for each proximal cue can be also depicted in the internal browser of the tool next to each hyperlink.

Moreover, the tool produces warnings of unfamiliar cues based on the LSA term vector length of each proximal cue, a measure that is correlated with word frequency. This is a good indication that the users modelled by the semantic space selected will perceive them to be relatively meaningless and thus they will be incapable of relating them semantically to their goal. At a site level analysis, the tool provides at any step of the simulation a graphical visualization of the predicted user's path and the scent trail followed. Thus, the designer can identify and investigate further cases of steps in the trail that have low scent (e.g. below a threshold that she defines) or lead the user to the 'wrong' path.

3. Automated Information Structuring Technique

A second tool that we realized is *AutoCardSorter* that automates structuring of information using a technique by calculating semantic similarity of different web pages and clustering accordingly the information space. This tool addresses the problem of reasonable content structuring and helps avoiding the web site level problems during the design of a site. Although, card-sorting study results can be stable with 20 or even fewer participants [7], an automated approach could substantially accelerate the design and evaluation lifecycle.

A typical scenario of the tool is the following: First the designer inputs all the titles of the pages that the site will contain (Figure 2, leftmost part of the user's screen). The tool then runs an automated analysis, using the LSA algorithm and machine learning techniques. LSA is used to calculate the semantic similarity among all the pages, while clustering algorithms are applied to group together semantically similar pages. Currently three complementary hierarchical agglomerative clustering algorithms have been realised, average linkage, complete linkage and single linkage to ensure the quality of the obtained results.

The output of the tool is an interactive dendrogram presenting the recommended clustering of the pages comprising the navigational model of the web site. Specifically, the tool clusters the described information space, suggests how the pages should be distributed and which pages should have links to each other, according to their semantic similarity. A possible variation to the previous scenario could include the definition of the desired number and/or the labels of the sections to be created. This variation implements an automated process of a closed card sorting technique where the designer specifies in advance the number and/or the names of the sections to be created and the tool places each page to a section according to their semantic similarity. In addition the designer has the ability to differentiate the number of the desired groups in a visual way, simply by dragging the line depicting the similarity strength among the clusters' items (Figure 2) and the tool reorganizes the results, thus showing the most effective item clustering in real time.

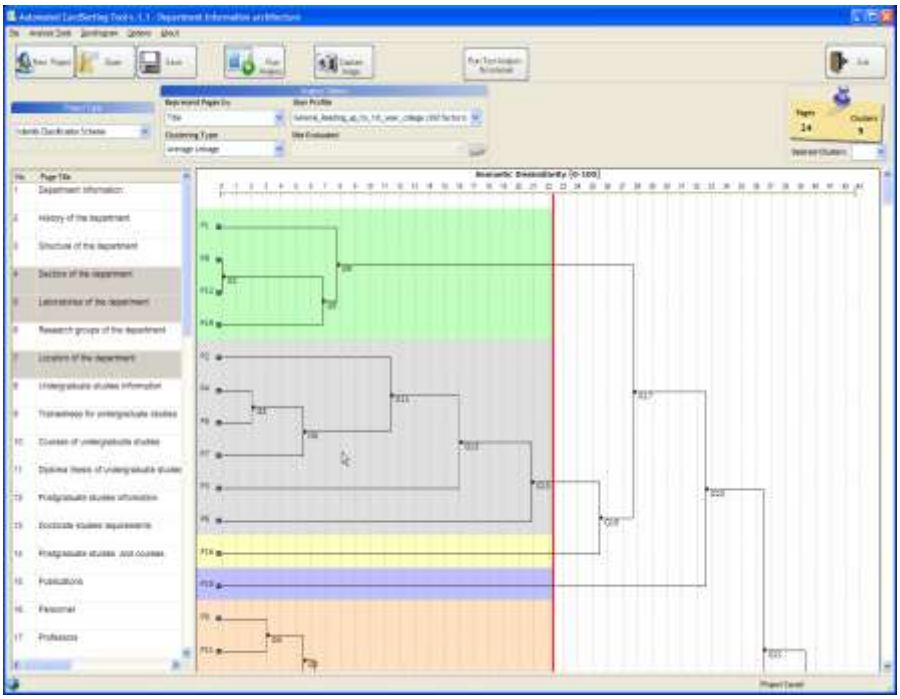

Fig. 2. Using *AutoCardSorter* to obtain the information architecture for a University Department's web site.

4. Discussion

In this paper we presented tools based on information foraging theory that could support web design and evaluation process. The presented tools predict expected users' behaviour, and support relevant design decisions of the information system environment. This knowledge can then be extended and used to develop methodologies and tools that lead both the design and the evaluation process of web sites. In addition, the presented tools may be used for validation of user techniques and for testing alternative design scenarios.

Initial application of the tool showed promising results, thus further validating our argument that researchers can make use of this knowledge in assessing interaction design for the Web. Despite the encouraging results obtained using ISEtool, there are still some issues to tackle in order to automate fully the process. For instance, initially when a web page is analyzed, the alternative text (e.g. the ALT Tag) of each graphical-link is assumed to be the proximal cue of this link. Currently, the adopted solution is to notify the user of the tool for such instances and ask for manual input of descriptions. Future work concerns full scale validation of the approach by conducting studies with real users, in order to determine the accuracy of the developed tools. In addition, we aim at studying phenomena relating with goal reformulation while users explore a

web information space towards finding desired patches of information and to examine possible extensions to the developed models such as data describing users' profiles or specific particularities of the site's context.

Even affective aspects of the user interface such as aesthetics [8], could influence significantly the user's motivation and engagement. Therefore, further research is needed on the effects of other aspects of user's experience design, such as credibility and aesthetics and their influence in user's motivation and engagement. Finally, it should be stressed that, we do not underestimate the value of established user based techniques and use of existing knowledge, like guidelines. Instead, tools that permit automatic validation of semantic appropriateness of web page hyperlinks and information organization, confirm the usefulness of traditional user methods which can be applied in a more focused way with significantly less effort and in a more formalised than empirical manner.

Acknowledgements
We thank European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II) and particularly the Program PYTHAGORAS, for funding the above work.

References

1. Pirolli, P., Card, S.: Information Foraging. Psychological Review 106(4), 643--675 (1999)
2. Blackmon, M., Kitajima, M., Polson, P.: Tool for Accurately Predicting Website Navigation Problems, Non-Problems, Problem Severity, and Effectiveness of Repairs. In Human Factors in Computing Systems CHI '05 Conference Proceedings, pp. 51--58. ACM Press, Portland, Oregon (2005)
3. Resnick, M., Sanchez, J.: Effects of organizational scheme and labeling on task performance in product-centered and user-centered retail websites. Human Factors, 46 (2004)
4. Katz, M., Byrne, M.: Effects of Scent and Breadth on Use of Site Specific Search on E-commerce Websites. ACM Transactions on Computer-Human Interaction. 10(3), 198--220 (2003)
5. Katsanos, C., Tselios, N., Avouris, N.: InfoScent Evaluator: A semi-automated tool to evaluate hyperlinks' semantic appropriateness of a web site. In J. Kjeldskov & J. Paay (Eds.) Proceedings of OZCHI 2006, pp. 373--376. ACM Press, Sydney (2006)
6. Landauer, T., Dumais, S.: A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104, 211--240 (1997)
7. Tullis, T., Wood, L.: How Many Users Are Enough for a Card-Sorting Study? In Proceedings of UPA'2004 Proceedings UPA'2004. Minneapolis, MN (2004)
8. Papachristos, E., Tselios, N., Avouris, N.: Inferring relations between color and emotional dimensions of a web site using Bayesian Networks. Interact 2005, pp. 1075--1078. Springer, Rome: (2005)