# Automated Semantic Elaboration of Web Site Information Architecture

**Christos Katsanos[a], Nikolaos Tselios[a], Nikolaos Avouris[a]**

[a]*HCI Group, Dept. of Electrical and Computer Engineering, University of Patras, 26500 Rio, Patras, Greece*

**ABSTRACT**

Structuring of the content is an important step in web site design, affecting greatly navigability and the overall user experience. Automated support of this task is the object of this paper. AutoCardSorter, a computational tool that supports clustering of the web pages of a site, is introduced. The proposed tool-based methodology uses semantic similarity measures, such as Latent Semantic Analysis, and hierarchical clustering algorithms, in order to suggest suitable information navigation schemes. In the paper, after introducing AutoCardSorter design and functionality, three independent studies are discussed. The studies, that were conducted in order to validate the proposal, compared the proposed method with the established card-sorting approach, in different domains. It was found that substantial gain in effectiveness was achieved without expense in the quality of results, therefore, reducing the required time and human resources.

**Keywords**

Information architecture, automated tool, semantic similarity, card-sorting, latent semantic analysis, cluster analysis.

## 1. INTRODUCTION

The exponential growth of web content has created an abundance of easily accessible information and a poverty of human attention (Simon, 1996). In addition, the design of a web site is a complex process characterized by an inherent dualism. The web should be treated simultaneously both as a user interface as well as a hypertext system (Garrett, 2002). As a result, one of the biggest challenges in web site design is creating the information architecture, which is "the structural design of an information space to facilitate task completion and intuitive access to content" (Rosenfeld and Morville, 2006). According to Kalbach (2007), information architecture represents the underlying structures that give shape and meaning to the content and functionality of the web site. Users come to a web site with some information needs and expectations about where to look for information. User-centered information architecture requires understanding of how users tacitly group, sort and label content in order to increase findability of information (Morville, 2005) and, eventually, usability during information interaction.

However, there is no generally accepted methodology for the design of user-centered information architecture, backed up by sound theoretical foundations (Jones, et al., 2006). This is mainly due to the fact that the complexity and the unstructured nature of information interaction are not expressed well in typical models of human-computer interaction (Toms, 2002). A variety of methods, such as card-sorting, contextual inquiry and ethnographic interviews are used, but it remains difficult to go from user research to the design itself (Sinha and Boutelle, 2004). Furthermore, such techniques are often overlooked, due to required resources and increased complexity to carry out the analysis. In addition, organizations tend to adopt navigation schemes, reflecting their own structure, instead of trying to identify and compromise with users' mental models and expectations.

Lack of appropriate information structure can cause various usability problems and deteriorate the overall interaction experience. The result of a vague information structure is cognitive overload and lostness which have long been recognized as major barriers experienced by users in hypermedia navigation (Conklin, 1987). Users often find typical web navigation tasks to be very difficult and have low success rates, even when they are first

***Please cite this document as***: Katsanos C., Tselios N., Avouris N. (2008). Automated semantic elaboration of web site information architecture. Interacting with computers, 20 (6), p.p. 535-544, doi:10.1016/j.intcom.2008.08.002

taken to a particular web site containing the information sought (Spool et al., 1999). Disoriented searchers seem to have difficulty forming a cognitive model of the information structure (Otter and Johnson, 2000) and can become lost because of the non-linear nature of hypertext systems (Chen and Macredie, 2002).The so-called ‘hyperspace lostness’ is quite common in cases where users lose track of the context when following a sequence of links and are unsure how to proceed in terms of satisfying their original goal (Gwizdka and Spence, 2007; Otter and Johnson, 2000).

Advances in search engines’ technologies should not be treated as an excuse for inappropriate information structuring. Although some users may arrive at a web site with well-formulated goals and adequate relevant knowledge, others do not (Marchionini, 1997; Maurer, 2006). Wu and Miller (2007) are in line with the above point, arguing that “search through navigation remains an indispensable method for locating unfamiliar information goals”. Teevan et al. (2004) report study results where the participants used keyword search in only 39% of their searches, despite knowing their information need up front. They suggested that ‘orienting’ is preferred to ‘teleporting’ because it has less cognitive overload, helps in maintaining sense of location and feel in control, and contributes in understanding the context of the information and get a sense of its trustworthiness.

Card-sorting (Rosenfeld and Morville, 2006) is one of the main methods used to improve the degree to which a web site supports navigation and information retrieval and can lead the design of a web site’s information architecture. A number of reporting studies confirm the validity of the method in the field of Human Computer Interaction (Capra, 2005; Nielsen and Sano, 1995; Sinha and Boutelle, 2004; Tullis and Wood, 2004). The technique elicits conceptual structures from participants by asking them to sort a group of cards with concepts written on them. In the context of web information architecture “it can provide insight into users’ mental models, illuminating the way that they often tacitly group, sort and label tasks and content within their own heads” (Rosenfeld and Morville, 2006).

While applying the method, representative users are given a stack of index cards, each containing one word or phrase representing the information or services provided on web pages. They are asked to group the cards in piles that make sense to them and subsequently name the resulting groups. There are two primary alternatives, *open* and *closed* card-sorting. The difference lies in the existence or not of a pre-established initial set of groups. Open card-sorting (with no pre-established groups) is used primarily in new web sites, while the closed variation is used for adding content to an existing structure or validating the results of an open card-sorting. The main quantitative data from a card-sorting study is a set of similarity scores that measures users’ view on the similarity of the various items. For instance, if all users sorted two cards into the same pile, then the two items represented by the cards would have 100% similarity. If only half the users placed two cards together those two items would have a 50% similarity score.

However, the method is demanding in terms of time and human resources. Card-sorting study results can be stable with 20 participants (Tullis and Wood, 2004), 15 (Nielsen, 2004) or even fewer, e.g. 7-10 representative users of each identified user group (Mauer and Warfel, 2004). However, the process of finding representative users early in the lifecycle can be daunting, time consuming and costly. Nielsen and Sano (1995) report that for the Sun web site each user typically completed the entire process involving 51 cards in about 30 minutes, though some took about 40 minutes. In addition, despite the fact that the method can provide valuable insight, thorough data pre-processing and statistical analysis is required, which diminishes the possibility of wider adoption. As derives from the experience of various usability practitioners, the typical effort to conduct and analyze a card-sorting session ranges from 3.5 to 7 person-days (Intranet Leadership Forum, 2006).

To tackle the latter issue, there is a range of software tools that can support the gathering of card-sorting data and/or assist with their analysis, such as *USort/EZCalc* (Dong et al., 2001), *CardZort*[1] and *OptimalSort*[2]. These tools automate aspects of the data collection and/or analysis and in general have three components: an administration tool for setting up the experiment, a tool for participants to conduct the sorting and an analysis tool. Among these tools, *MindCanvas*[3] is a commercial remote research tool that provides an interesting game-like elicitation approach to gather card-sorting data and elaborate them to rich visualizations. However, even with these tools available, the effort for gathering and analyzing the data remains substantial. As a result, the method is often neglected by practitioners, who opt for empirical and superficial approaches.

In this paper, an innovative tool-based approach for the design and evaluation of a web site's information architecture is presented. This method is offered as an automated alternative to the tedious, but empirically-proven useful, card-sorting method. It employs a novel approach combining semantic similarity measures, in specific Latent Semantic Analysis (LSA- Landauer and Dumais, 1997), clustering algorithms (Witten and Frank, 2005) and mathematical heuristics, such as the eigenvalue-one criterion (Hatcher, 1994), to address the problem of content structuring. The method is aimed at providing the necessary flexibility and efficiency to practitioners. Three independent studies, in different domains, have been conducted in order to validate the proposed tool-based approach against users' card-sorting groupings.

In the rest of the paper, the proposed tool-based methodology is presented first, followed by a brief description of the semantic similarity measure system it is based on. Next, a typical scenario of applying the approach is described. Subsequently, three independent studies carried out in different domains are presented. The studies validate the quality of the results and the efficiency of the proposed methodology against the widely-used card-sorting method. Finally, we discuss the implications, conclusions and future directions of the presented research.

## 2. A TOOL-BASED METHOD TO DESIGN AND EVAVUATE INFORMATION ARCHITECTURES

The proposed tool-based methodology addresses the problem of content structuring (structural navigation) and helps in creating semantic relationships between related pieces of content across levels of a hierarchy (associative navigation - Kalbach, 2007). The proposed approach is aimed at providing the necessary flexibility and efficiency to the practitioners, and can be used for both the initial design and redesign of information-rich web sites. The methodology is expressed in the form of a computational tool, the AutoCardSorter (Katsanos et al., 2008). Currently, the tool is made freely available upon request at http://hci.ece.upatras.gr/autocardsorter.

### 2.1. Semantic Similarity Measurement

Various approaches to estimate semantic similarity between words, phrases and passages have been proposed in the literature (Landauer and Dumais, 1997; Lund and Burgess, 1996; Manning and Schutze, 1999; Miller, 1995; Rhode et al., 2004; Turney, 2001). The proposed methodology uses such techniques to derive a quantitative estimation of the semantic similarity among text descriptions of the web pages of the designed or evaluated web site. Given that there is much ongoing research targeted at understanding which of the semantic similarity measures performs better, AutoCardSorter was built on a software framework that allows the easy integration of alternative algorithms.

Kaur and Hornof (2005) classify semantic similarity measurement systems as: a) taxonomical, b) statistical and c) hybrid. *Taxonomical* approaches calculate measures, like path-length between two node-words, relying on

[1] http://www.cardzort.com
[2] http://www.optimalsort.com
[3] http//themindcanvas.com

manually-created lexical databases, such as WordNet (Miller, 1995), to derive a quantitative value of similarity between terms. In *statistical* techniques, semantic relationships between terms are captured from the probability of their co-occurrence in a text corpus, which is a large collection of documents. *Hybrid* methods attempt to combine taxonomies of concepts with statistical properties of a text corpus.

The main advantage of taxonomical approaches is that they rely on human-encoded taxonomies. Thus, they ensure a certain quality of the results and make it possible to model multiple word senses by encoding synonyms. However, Kaur and Hornof (2005) argue that this knowledge acquisition is tedious, subject to the vagueness of human judgment, and not easily scalable to new terms, domains and languages. They compared various semantic similarity measures for predicting human judgments in web navigation tasks and found that the taxonomical approaches perform the worst. In addition, their results suggest that most hybrid approaches, such as res (Resnik, 1999), do not improve significantly the results.

On the other hand, statistical methods are quite flexible as they rely on unsupervised machine learning techniques and reflect human performance quite accurately (Landauer and Dumais, 1997; Wolfe and Goldman, 2003). These systems are also attractive from a cognitive modeling standpoint because they bear an obvious similarity to patterns of mean activation over collections of neurons (Rhode et al., 2004). Most algorithms that fall into this category are variants of Latent Semantic Analysis (LSA - Landauer and Dumais, 1997), Point-wise Mutual Information (PMI - Manning and Schutze, 1999) and Hyperspace Analogue to Language (HAL - Lund and Burgess, 1996).

The underlying idea of LSA is that "the totality of information about all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and set of words to each other" (Landauer et al., 1998). First, LSA parses suitable, large text corpora that represent a given user population's understanding of words and produces a term-document matrix of each word's frequency of occurrence. Next, each cell is weighted by a function that expresses both the word's importance in the particular document and the degree to which the word type carries information in the domain of discourse in general. Alternatively, the term-document matrix can be viewed as a huge, multidimensional space where each term is considered as a separate dimension and each document is a vector in this term-space. Subsequently, LSA applies Singular Value Decomposition (SVD), the mathematical generalization of factor analysis, to project this large, multidimensional space down into the least-squares best fit number of dimensions. This is a key step in the LSA method since the SVD algorithm reveals similarities that are latent in the document collection. LSA is, actually, exploiting the property of natural language that words with similar meaning tend to occur together. In this way, the initial term-space is transformed into a semantic space, where the degree of semantic similarity between any pair of texts, such as the descriptions of two web pages, 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.

The adequacy of LSA's reflection of human meaning-based judgements has been established in a variety of cases, such as modeling discourse comprehension (McCarthy et al., 2006), judging essay quality (Landauer et al., 2003), identifying navigability problems (Blackmon et al., 2005; Katsanos et al., 2006) and providing real-time navigation support (van Oostendorp and Juvina, 2007). Moreover, unlike most of the other approaches that calculate similarity between words, LSA provides a built-in method of computing the similarity between passages, such as descriptions of web pages, based on the underlying vector space model. Through empirical analysis (Kaur and Hornof, 2005), this approach has been found to be superior to the indirect approaches of computing passages similarity from word similarity, such as Sutcliffe et al. (1995), used by the other statistical techniques. The proposed tool-based methodology, currently, employs the LSA statistical method (available at http://lsa.colorado.edu) to calculate the semantic similarity among text descriptions of under-design or existing web pages, and structure the information space. The approach is presented in the following.

### 2.2. AutoCardSorter – A typical Usage Scenario

Automating aspects of the design and evaluation process is critical given the quantity, frequency of updates and sheer size of sites being produced (Brinck and Hofer, 2002). Such an automated approach could, substantially, accelerate the design and evaluation lifecycle by providing the necessary efficiency and flexibility. This increased efficiency offered by AutoCardSorter is expected to be even more important when designing or evaluating large sites, as the established card-sorting method becomes too time-consuming and complicated for more than 100 cards (Mauer and Warfel, 2004). In addition, the proposed methodology increases the possibility to explore alternative designs and, therefore, can lead to better solutions for web sites' structures.

AutoCardSorter employs a novel algorithm that uses LSA and clustering algorithms to support structuring of information simulating an open card-sorting experiment. This is achieved by creating a matrix of semantic similarities among descriptions of content items using LSA as a similarity metric. Subsequently, clustering algorithms are used to construct the information space. Furthermore, AutoCardSorter implements additional automated analyses to support the designer towards creating a suitable navigation scheme: (a) two complementary ways to determine the optimal number of categories in terms of variance explained and (b) a method to identify associative links in the adopted classification.

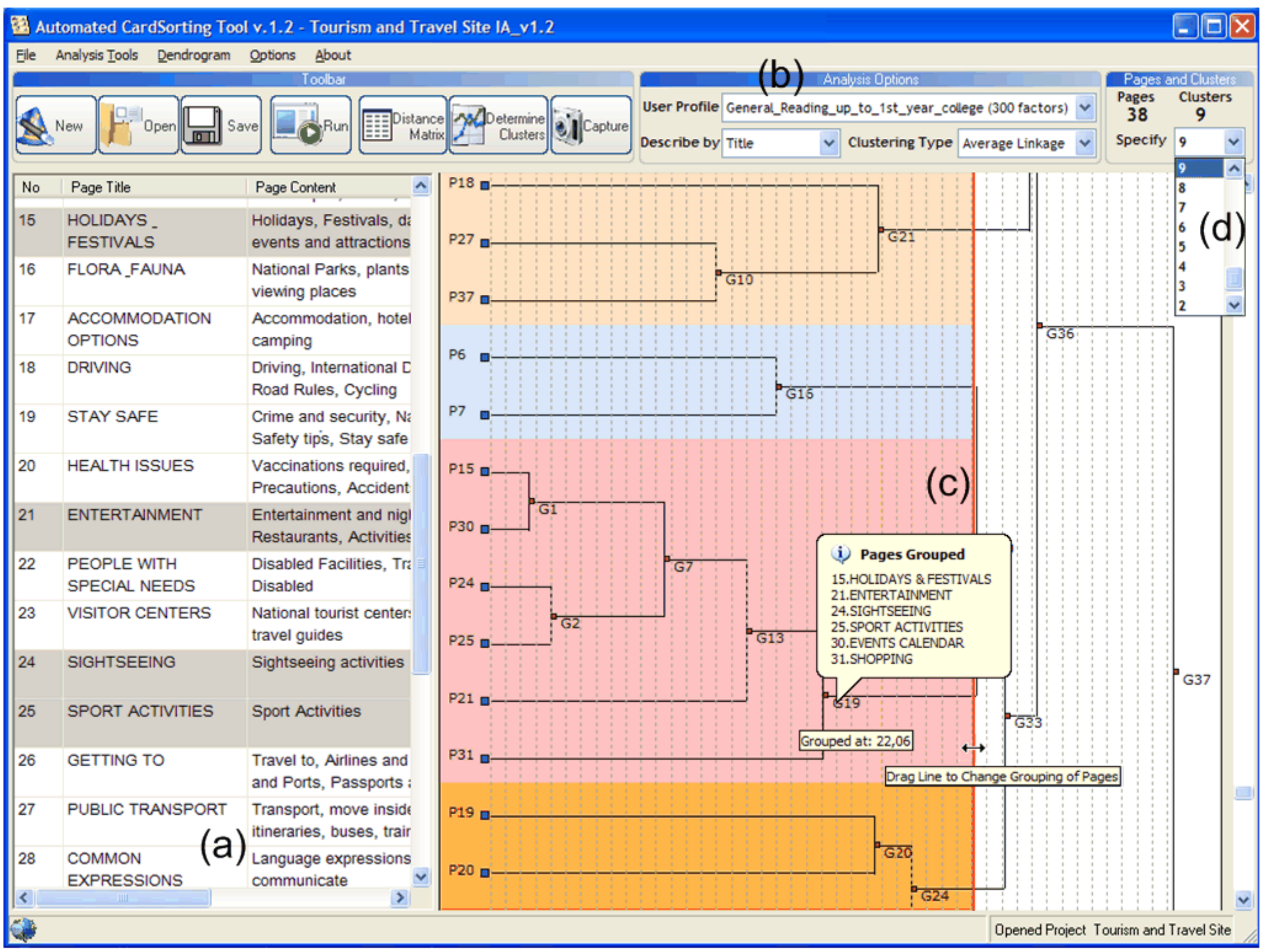


**Figure 1.** Using AutoCardSorter to identify the information architecture of a web site dealing with traveling and tourism issues.

A typical scenario of using the tool is the following. First, the designer provides descriptions of the pages that the web site will contain (Figure 1-a), chooses an appropriate LSA semantic space to represent the typical users of the site and selects the desired type of clustering algorithm (Figure 1-b).Then, the tool runs an automated analysis, combining the LSA method with the selected clustering algorithm. In the context of AutoCardSorter, the Euclidian distance of the LSA index among the web pages is used as the similarity metric for the clustering algorithms employed. Ideally, items that belong to the same cluster share a higher degree of semantic similarity compared to items of other clusters.

While there is a great variety of clustering techniques, hierarchical clustering algorithms have been, extensively, used for information structuring problems because of their intrinsic attribute to create a hierarchy of groupings (Dong et al., 2001; Sinha and Boutelle, 2004; Tullis and Wood, 2004). Therefore, hierarchical clustering algorithms have been, also, adopted for AutoCardSorter. Currently, three, popular, complementary, hierarchical agglomerative clustering algorithms have been implemented to ensure the quality of the obtained results: (a) average linkage, (b) complete linkage and (c) single linkage (Witten and Frank, 2005). Table 1 sketches the main steps of the algorithm employed by AutoCardSorter.

A possible variation could include the definition of the desired number and labels of the sections to be created. This variation implements an automated process of a closed card-sorting technique, where the tool places each page to a section, according to their semantic similarity.

1) Begin with the text descriptions of the N web pages.
2) Create a similarity matrix S using a semantic similarity measure (e.g. LSA) to calculate the semantic similarity for each pair (i), (j) of pages.

   S [(i), (j)] = LSA (i, j), where S is a symmetric matrix of NxN dimensions.
3) Convert the similarity matrix S to a normalized dissimilarity (or distance) matrix D:

   D [(i), (j)] = d { S [(i), (j)] }, with d defined as:

   d (i, j) = 50 (1 – LSA (i, j)), where d (i, j) ∈ [0,100]
4) Consider each web page as a separate cluster.
5) Find the less dissimilar pair of clusters in the current clustering step, say pair (r), (s), according to:

   D [(r),(s)] = min { D [(i),(j)] }
6) Merge clusters (r) and (s) into a single cluster.
7) Update the distance matrix, D, by deleting the rows and columns corresponding to clusters (r) and (s) and adding a row and column corresponding to the newly formed cluster. The distance d between the new cluster, denoted (r,s) and old cluster (k) is defined based on the selected type of hierarchical clustering algorithm:
   a. If Single-Linkage selected then d[(k), (r,s)] = min { D[(k),(r)], D[(k),(s)] }.
   b. If Complete-Linkage selected then d[(k), (r,s)] = max { D[(k),(r)], D[(k),(s) }.
   c. If Average-Linkage selected then d[(k), (r,s)] = (D[(k),(r)] + D[(k),(s)])/2.
8) If all web pages are in one cluster, stop. Else, go to step 5.

**Table 1.**The main steps of the algorithm employed by AutoCardSorter.

The output of the tool is an interactive tree diagram (known as dendrogram, see Figure 1c), presenting the recommended web site’s navigation scheme. In specific, the tool clusters the described information space, suggests how the web pages should be distributed and which pages should have links to each other according to their semantic similarity. In addition, the designer is offered with the option to differentiate the number of the desired groups in a visual way. This can be achieved either by dragging the line depicting the similarity strength among the grouped items (Figure 1-c) or by specifying explicitly the desired number of top-level categories (Figure 1-d). In both cases, the tool reorganizes the results, showing the most effective item-clustering in real time.

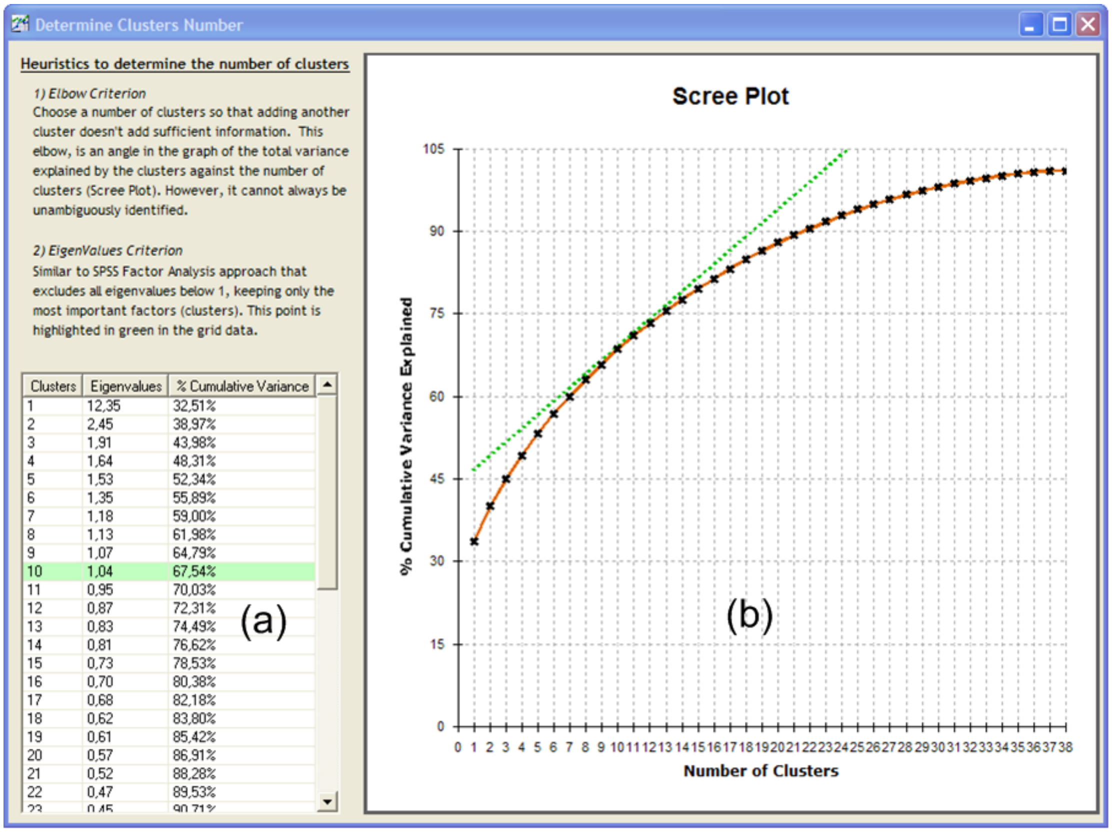


**Figure 2.** Determining the optimal number of a web site's categories using eigenvalue-one criterion (a) and scree-plot analysis (b).

Furthermore, the tool provides a suggestion about the number of clusters that should be chosen based on the widely-used eigenvalue-one criterion (Hatcher, 1994). The rationale behind this criterion is that interpretation of proportions of variance smaller than the variance contribution of a single variable are of dubious value. AutoCardSorter identifies the optimal number of clusters, in terms of variance explained, by implementing an eigenvalue analysis of the pages' similarity matrix and by keeping only the eigenvalues that are greater than 1 (Figure 2-a). However, the eigenvalue-one criterion has been criticized to overestimate the number of factors to retain (Lance et al., 2006) leading to results that can be justified on the mathematical level, but with no interpretable meaning in the conceptual level. For this reason, AutoCardSorter offers a complementary way of determining the number of categories known as Scree Plot Analysis (Cattell, 1966) or Elbow Criterion. This is achieved by plotting the percentage of variance explained against the number of clusters (Figure 2-b). The first clusters always explain a lot of variance, and subsequently the plot converges asymptotically to 100%. Nevertheless, this elbow cannot always be unambiguously identified. Thus, this rule introduces a bias possibility due to the subjectivity involved in selecting the elbow (Lance et al., 2006). AutoCardSorter provides both complementary ways of determining the number of categories to better support the designer.

In addition, AutoCardSorter supports the identification of associative links across levels of the adopted categorization. This is achieved by running an automated analysis of the similarity-matrix which takes into account the current classification selected by the designer and proposes associative links for pages that are semantically close but have not been, currently, grouped together. Such links create alternative paths to pages of the web site and increase findability of information (Morville, 2005).

## 3. VALIDATION STUDIES

Three studies have been designed and conducted in order to investigate the quality of results and the efficiency of the proposed approach. Our goal was to compare the widely-used card-sorting method and AutoCardSorter in the design or redesign of the information architecture of web sites for various domains and sizes. The first study involved the design of a small project web site about nutrition and health issues, the second the redesign of a larger educational portal and the third the design of an even larger traveling and tourism web site.

In the next sections, first, the common elements of the three studies are described, and then, each study is presented separately, followed by a discussion, the conclusions and future directions of the presented research.

### 3.1. Methodology and Procedures

In all three studies, the proposed tool-based approach was used to define the information architecture and the results were compared with those of an open card-sorting study. For both methods, the same descriptions for the web pages were used. However, it should be noted that as a first step, the descriptions of the pages were enriched with the contextual information that they were referring to. This is an essential step to accommodate the differences in the ways humans and computers perceive the meaning of words. Therefore, all the pronouns were replaced with their implicitly related nouns or noun-phrases and the same label for all instances of semantically related terms was consistently used. At a next step, the descriptions of the web pages were provided as input to AutoCardSorter and an appropriate LSA semantic space was selected to reflect the typical users of each web site.

An appropriate number of representative users were recruited for each open card-sorting study. Research (Nielsen, 2004; Tullis and Wood, 2004) has shown that for card-sorting studies 15 to 20 users are sufficient for stable results, so in our studies 18 to 34 users participated. The sessions were split into three sections. First participants were given information regarding the general nature of the study. Second, the overall scope of the web site was communicated to them. Finally, the content items to be sorted were provided to them. A facilitator kept notes of participants' comments and tried to help them in their inquiries without leading them. The average-linkage hierarchical cluster analysis was adopted for both approaches as it, usually, produces balanced groupings that are easier to interpret (Witten and Frank, 2005). The total time required to prepare, conduct and analyze the results of the card-sorting exercise as well as using AutoCardSorter was measured.

In order to investigate the quality of results of the proposed tool-based approach, three different types of comparisons with the results of the open card-sorting studies were performed: a) similarity-matrices correlation analysis, b) base-clusters comparison and c) elbow-based navigation schemes comparison.

In the context of our analysis, a similarity-matrix contains a measurement of the semantic similarity for each pair of page descriptions. In the proposed tool-based approach this metric is the LSA index, while for the card-sorting studies the normalized frequency of card-pairs appearing in the same pile was used. As a first step in the validation of the proposed approach, a correlation analysis of the similarity-matrices produced by the two methods was conducted.

To further investigate the accuracy of our technique, the approach proposed by Tullis and Wood (2004) to compare two dendrograms objectively was, also, applied. They defined the amount of separation as "the degree of separation of two cards in each of the original base-clusters, the basic two-card clusters that are formed during a cluster analysis, based on cards that are most similar". They measured this amount of separation by counting the number of nodes (i.e. intersections in the dendrogram) between the two cards of each base-cluster. Following their rationale, the dendrogram produced by the analysis of the card-sorting data was defined as the 'original dendrogram'. Then, the amount of base-clusters separation between the original dendrogram and the one produced by AutoCardSorter was calculated. This amount of separation was normalized based on its maximum possible value, which corresponds to having to traverse all nodes in the dendrogram to connect the two cards that form the base-cluster.

Finally, a comparison of the optimal, in terms of variance explained, navigation scheme produced by each approach was conducted. The optimal navigation scheme proposed by each approach was derived by applying the Elbow criterion to determine the number of clusters for each case. First, the elbow-based navigation scheme produced by the card-sorting data was defined as the 'original navigation scheme'. Next, the percentage of agreement between the original navigation scheme and the one produced by AutoCardSorter was calculated. This was achieved by calculating the percentage of pages that AutoCardSorter grouped together in the same category that the original navigation scheme did. However, as AutoCardSorter tended to produce a slightly larger number of categories for the optimal navigation scheme, there were cases where AutoCardSorter created two or more sub-categories which corresponded to one category in the original navigation scheme. In such cases, the items in the larger sub-category were classified as 'in agreement' and all the other items as 'in disagreement".

Although the three aforementioned types of comparisons are not orthogonal, they provide useful insight from different perspectives. The *similarity-matrices correlation analysis* is the strictest approach of all, as it compares the measurements of the semantic similarity for all pairs of page descriptions produced by the two methods. In addition, this comparison analysis is more general, as it does not presuppose that cluster analysis is applied to analyze the card-sorting data. *The base-clusters comparison* is an objective way to compare the most similar page-pairs produced by applying cluster analysis on the similarity-matrices of each method. Finally, even if perfect base-clusters agreement is not achieved, the resulting elbow-based navigation schemes can still be identical as the cutting line in a dendrogram is never drawn on the base-clusters level. The elbow-based navigation schemes comparison attempts to capture such cases. Other methods proposed to analyze card-sorting results, such as "eye-balling the data" (Mauer and Warfel, 2004) or factor analysis (Capra, 2005), are not applicable in our case.

### 3.2. Study 1 – Designing a Project Site for Health and Nutrition

The goal of the first web site was to enable partners' collaboration during the project lifetime and communicate to the general public the project's results on potential positive and negative effects of food consumption. Following the aforementioned methodology, AutoCardSorter was used to structure the 16 pages of the web site's information space. The whole process required approximately 1 hour.

18 participants, 10 male, 8 female, aged 22-55 with a mean of 28, took part in the open card-sorting session. The participants shared a high level of education and reported high internet experience. A pile of index cards representing the web site's pages was provided to the participants for the card-sorting exercise. Each session lasted approximately 1 hour. EZCalc (Dong et al., 2001) was used for the analysis of the data. The whole process to prepare the card-sorting study, collect and analyze the data required approximately 27 hours.

The correlation analysis indicated a very high degree of correlation ($r=0.80$, $p<0.01$) between the similarity-matrices obtained by the two methods. In addition, the average amount of base-clusters separation was found to be 0 and the elbow-based navigation schemes were identical (see Figure 3). Furthermore, AutoCardSorter proved approximately 27 times faster when compared to the traditional card-sorting technique.

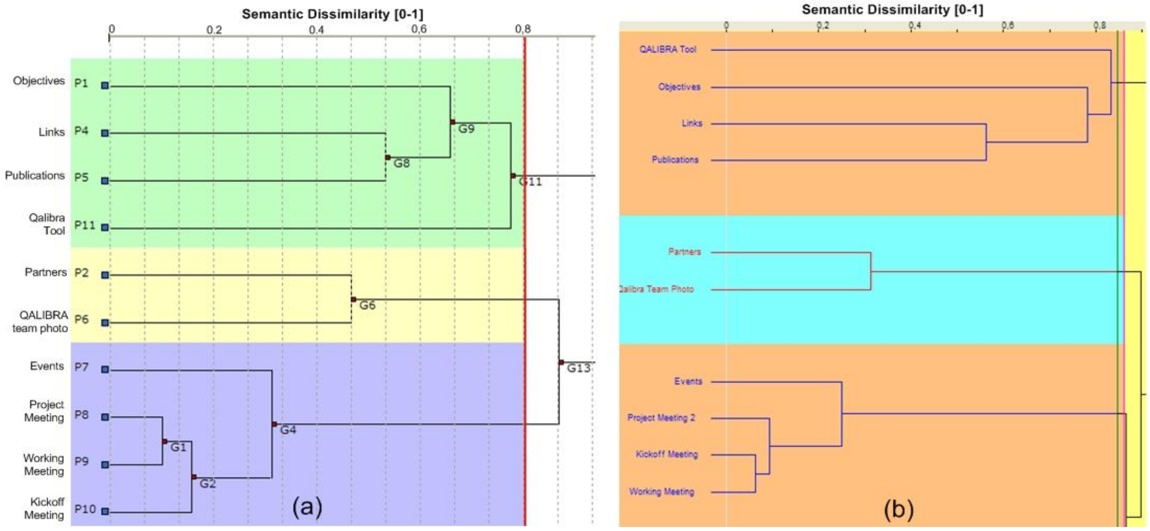


**Figure 3.** Part of the average-linkage, elbow-based dendrograms produced for the design of a health and nutrition project site by using AutoCardSorter (a) and by analyzing card-sorting data (b).

### 3.3. Study 2 – Redesigning an Educational Portal

The second validation study dealt with the redesign of an education portal. The portal provides various electronic services to the students of a University's Department, including personalized diaries and course schedules, electronic submission of thesis and specialized faculty-search facilities. Despite the undoubted usefulness of such services, many students had been complaining that experienced problems navigating the portal and finding the desired information or service.

A redesign of the portal was decided and AutoCardSorter was used as an integral part of the redevelopment process. Descriptions of all the 27 content items and services offered by the portal were provided as input to the tool and a semantic space appropriate for the reading and understanding skills of college students was selected. Alternative designs and navigation schemes were explored, by taking advantage of the wide range of the aforementioned functionalities offered by the tool. The whole process took approximately 2 hours.

Study participants were 26 students of the Department (17 male, 9 female, aged 22-25 with a mean of 23). Participants were encouraged to ignore the current structuring of information, which they were familiar with, and propose a structure that would feel natural to them. A card-sorting tool, USort (Dong et al., 2001), was used to facilitate the card-sorting exercise. The whole process to conduct the card-sorting study and analyze the users' groupings required approximately 22 hours.

Comparison of the results indicated a high degree of correlation ($r=0.52$, $p<0.01$) between the similarity-matrices produced by the users and AutoCardSorter. The average amount of base-clusters separation was found to be 7.5% and the comparison of the elbow-based navigation schemes indicated a 93% agreement. Parts of the dendrograms produced by each approach are depicted in Figure 4. In addition, the proposed approach proved approximately 11 times faster.

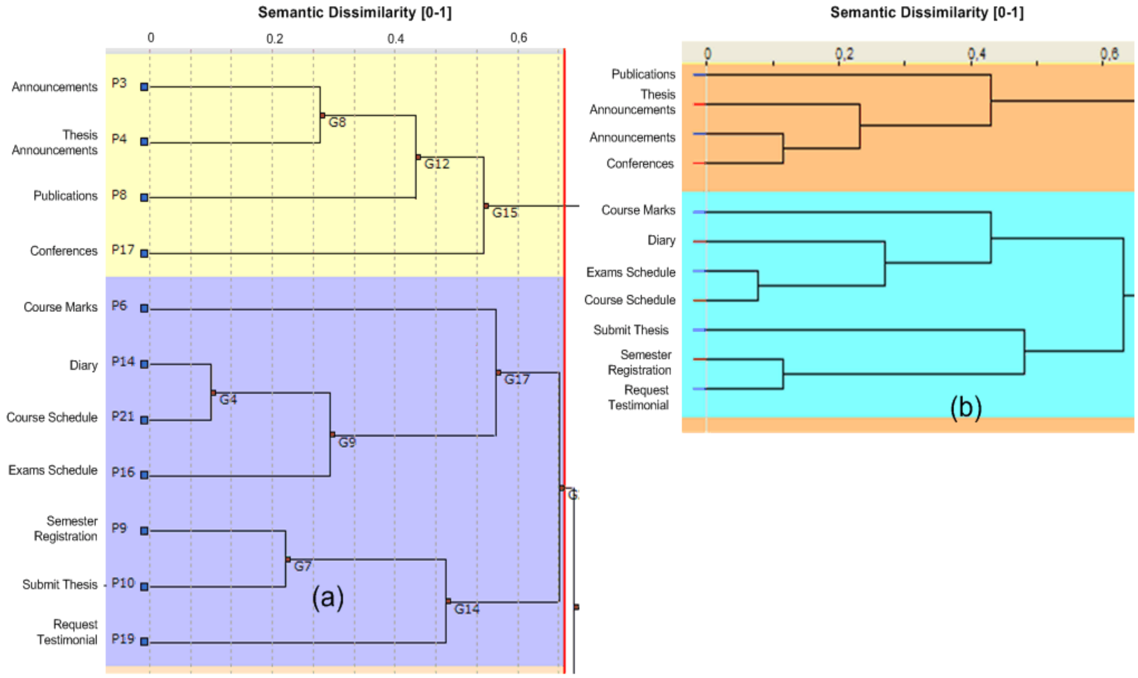


**Figure 4.** Part of the average-linkage, elbow-based dendrograms produced for the redesign of an educational portal by using AutoCardSorter (a) and by analyzing card-sorting data (b).

### 3.4. Study 3 – Designing a Web Site for Traveling and Tourism

The third validation study concerned the design of the information architecture for a traveling and tourism web site. A total of 38 content items were used in the study, many of which represented information offered by typical sites of this domain, such as "Sightseeing", "Travel Planner", "Accommodation Options" and "Weather Forecasts". Some cards represented general services, such as "Send e-card", or addressed specialized information needs, such as "For people with special needs". The content items were selected from highly-rated web sites evaluated for the 2007 Webby Awards[4].

As discussed previously, AutoCardSorter was used to design the information space of the web site. Descriptions of the content items to be organized were provided as input to AutoCardSorter and a general semantic space was selected as representative user profile. The whole process required approximately 2 hours.

34 participants (25 male, 9 female, aged 21-26 with a mean of 22) were recruited. All of them shared a high level of education and reported high internet experience. The time required to carry out the card-sorting study was approximately 28 hours.

Following the same methodology, a high degree of correlation ($r=0.59$, $p<0.01$) was observed for the similarity-matrices produced by the users and AutoCardSorter. The average amount of separation for the base-clusters was very low (2.5%). Moreover, the elbow-based navigation schemes of the two approaches (Figure 5) shared a high degree of agreement (87%). Finally, AutoCardSorter proved 14 times faster than the card-sorting technique, in this case.

[4] http://www.webbyawards.com

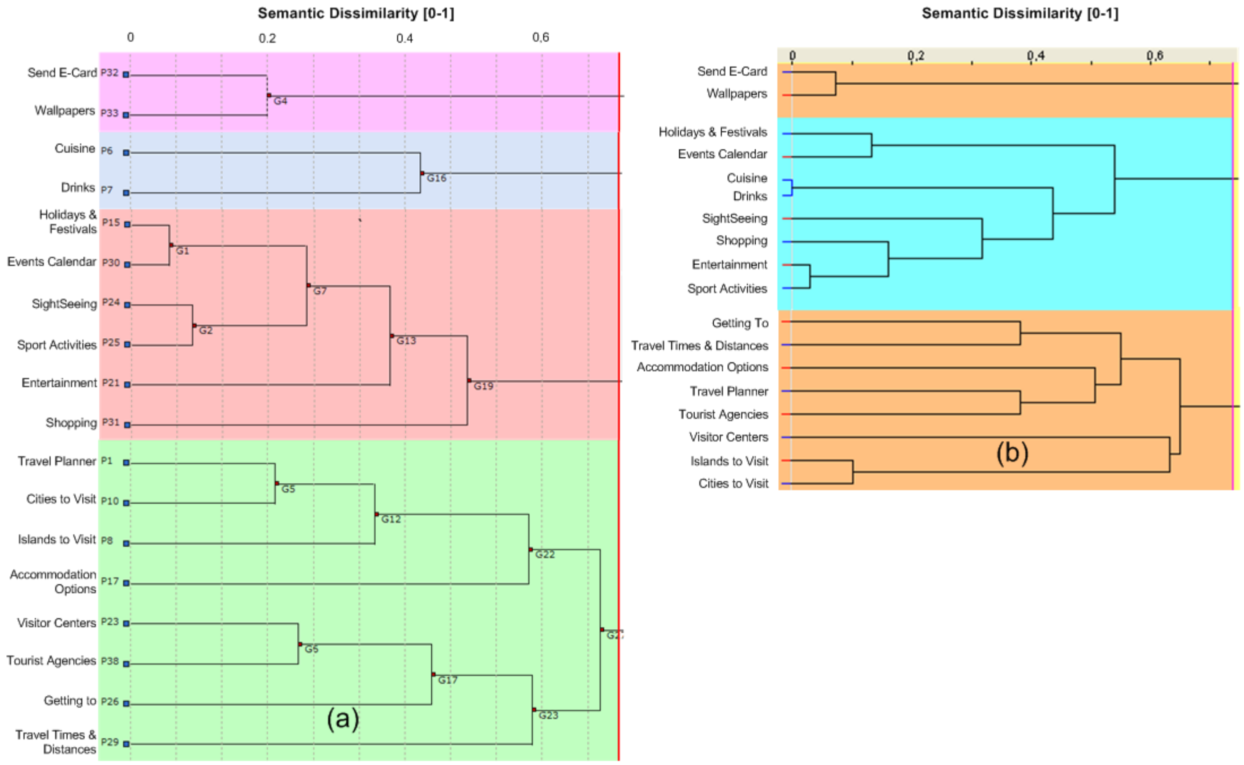


**Figure 5.** Part of the average-linkage, elbow-based dendrograms produced for the design of a traveling and tourism web site by using AutoCardSorter (a) and by analyzing card-sorting data (b).

## 4. DISCUSSION

Table 2 summarizes the results of the three studies that were conducted in order to investigate the validity and the efficiency of the proposed tool-based approach. The studies compared the results of open card-sorting exercises involving representative users with the outcome derived by using AutoCardSorter to design or redesign the information architecture of web sites for various domains and sizes

As Table 2 shows, AutoCardSorter provided semantically similar groupings and overall information structures to the ones derived by representative users involved in the card-sorting sessions. First, large correlation, ranging from 0.52 to 0.80 ($p<0.01$), between the similarity-matrices obtained by the users and AutoCardSorter was observed for all three studies. Moreover, the base-clusters comparison proposed by Tullis and Wood (2004) further proved that the results produced in all studies were highly similar. In addition, the average agreement of the elbow-based navigation schemes produced by each approach was 93% (lowest 87%, highest 100%).

Contrasting the three studies, it seems that the correlation of the results obtained by the two approaches is much higher in the first study than in the other two. A possible explanation of this could be related to the lower number of card-sorting participants and content items to be sorted in the first study. As a result, card-sorting participants reached a higher level of agreement about the perceived semantic similarity of content-items and thus, the variance of the card-sorting results was lower and the correlation to the tool's outcome was higher. In addition, differences in the domain expertise of participants could have, also, influenced the observed correlations. However, given that all these factors varied simultaneously, it is not safe to draw any conclusions yet. Instead, more validation studies are required in which one of these factors is experimentally controlled while the others are held constant.

Furthermore, AutoCardSorter seems to produce a slightly larger number of categories for the elbow-based navigation scheme with less but tightly-connected items. Such differences in the number of categories could explain the observed lowest similarity-matrices correlation and highest average amount of base-clusters separation for the second study although it seems to have slightly higher elbow-based navigation schemes agreement than the third study.

| | | | **Study 1**<br>**Health Project Site** | **Study 2**<br>**Educational Portal** | **Study 3**<br>**Travel & Tourism Site** |
|---|---|---|---|---|---|
| **Study Details** | Number of Content Items | | 16 | 27 | 38 |
| | Number of Participants in Card Sorting Experiments | | 18 | 26 | 34 |
| **Correlation Analysis** | Similarity-Matrices Correlation | | 0.80 (p<0.01) | 0.52 (p<0.01) | 0.59 (p<0.01) |
| **Base-Clusters Comparison** | Number of Base-Clusters | *Card Sorting* | 4 | 9 | 12 |
| | | *AutoCardSorter* | 4 | 8 | 12 |
| | % Average Amount of Separation | | 0% | 7.5% | 2.5% |
| **Elbow-Based Navigation Schemes Comparison** | Number of Categories | *Card Sorting* | 4 | 6 | 7 |
| | | *AutoCardSorter* | 4 | 7 | 9 |
| | Average Items per Category | *Card* Sorting | 4 | 4.5 | 5.4 |
| | | *AutoCardSorter* | 4 | 3.9 | 4.2 |
| | % Navigation schemes Agreement | | 100% | 93% | 87% |
| **Efficiency** | Total Time Required | *Card Sorting* | 27 hours | 22 hours | 28 hours |
| | | *AutoCardSorter* | 1 hour | 2 hours | 2 hours |
| | AutoCardSorter against Card Sorting | | 27 times faster | 11 times faster | 14 times faster |

**Table 2.** Summary of the three independent studies that compared in different contexts the validity and efficiency of the proposed tool-based method against manual card-sorting with representative users.

As Table 2 shows, AutoCardSorter produced results of at least comparative quality, while it was more efficient than the established card-sorting technique, as it was 11 to 27 times faster. The highest gain in the efficiency is observed for the first study where the card-sorting experiment was conducted with traditional index cards. It is worth mentioning that approximately 70% of the time required to design the information architecture of a web site based on the card-sorting technique was spent on analysis of data. Therefore, the proposed-tool-based method remains highly efficient (on average 12 times faster) even when card-sorting tools are used, as in the case of the other two studies.

In addition, the efficiency of the proposed approach is influenced to a lesser extent by the number of content items to sort. AutoCardSorter remains an efficient solution when designing or evaluating large sites, in contrast to card-sorting that becomes too complicated and time-consuming for more than 100 cards (Mauer and Warfel, 2004).

## 5. CONCLUSIONS AND FUTURE WORK

The main strength of hypertext-based systems is their flexible structure giving users freedom to browse and interact with the embedded information. However, lack of appropriate information clustering can cause various usability problems and deteriorate the overall interaction experience (Otter and Johnson, 2000). In this paper, an innovative methodology expressed in a form of a computational tool, the AutoCardSorter, that supports design and evaluation of information-rich applications, such as web sites, has been described.

The proposed method automates the process of identifying information navigation schemes by employing a novel algorithm that combines semantic similarity measures, in specific LSA (Landauer and Dumais, 1997), clustering algorithms (Witten and Frank, 2005) and mathematical heuristics, such as the eigenvalue-one criterion (Hatcher, 1994). Furthermore, additional automated analyses based on mathematical modeling are provided in order to support the designer in creating a suitable navigation scheme: (a) two complementary ways to determine the number of categories based on total variance explained and (b) a method to identify associative links in the adopted categorization. Practitioners using the tool should place emphasis in selecting a representative semantic space of their typical users and producing page descriptions enriched with the contextual information that they are referring to. The proposed approach can be used for both the initial design and redesign of the information architecture of hypertext-based systems, providing the necessary flexibility and efficiency, as demonstrated in the three studies discussed.

The three independent studies that were conducted in different contexts and related to different domains depicted the validity and efficiency of the presented approach. The tool-based approach proved approximately 17 times faster compared to a typical card-sorting study, providing at the same time highly similar results, as demonstrated by three different types of analyses. The increased efficiency offered by AutoCardSorter is expected to be even more important when designing or evaluating large sites, where card sorting studies are not able to tackle the complexity of the information spaces.

However, the presented tool-based approach has, also, some limitations. Unlike card-sorting, it lacks the qualitative feedback obtained from representative users and cannot provide insight into the labels that should be chosen for the produced categories. Future work includes investigating ways to address this issue, such as automatically identifying the most frequent words in the categories created and providing a set of ‘near-neighbor’ words (i.e. semantically close) from which the designer could choose to form a valid label.

Furthermore, Capra (2005) argues that hierarchical cluster analysis, currently used by AutoCardSorter, is more suitable for highly structured settings, like software menus, and suggests alternatives, such as factor analysis, for architecting web sites. In addition, LSA is often criticized for poor performance when the words are relatively rare in the corpora it has been trained on. Moreover, much ongoing research aims at understanding which of the semantic similarity measures produces better similarity scores (Budiu et al., 2007; Falconer et al., 2006; Kaur and Hornof, 2005). A promising future direction to address these issues is the inclusion of more dynamic approaches for training of corpora (i.e. web search results) to induce semantic similarities between passages, such as LSA-IR (Falconer et al., 2006) and PMI-IR (Turney, 2001).

Last but not least, future work includes the conduction of additional similar validation studies, in which the tool’s results are compared with the ones derived by card-sorting user studies in different contexts. In addition, we plan to further investigate the validity of the proposed approach for a web site's redesign by contrasting measures of its perceived and actual usability in two conditions: (a) the initial version of a web site and (b) an AutoCardSorter-revised version of the same web site.

As a footnote, it should be stressed that despite the advantages of the presented automated approach, the value of established user-based techniques should not be neglected. Instead, the proposed tool-based approach coupled with similar methods, such as an automated tool that evaluates semantic appropriateness of hyperlink's descriptions (Katsanos et al., 2006), could be used as a complementary part of an iterative design process in conjunction with user-based methods, allowing deeper exploration of alternative solutions.


## ACKNOLEDGEMENTS

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. We, also, acknowledge the support of the EU-funded project FOOD-CT-2006-022957: QALIBRA (Quality of Life – Integrated Benefit and Risk Analysis; Web-based tool for assessing food safety and health benefits). In addition, we would like to thank Dimitra Ioannou from the Psychology Department of the University of York for her comments and valuable help with the statistical analysis. Last but not least, we thank the anonymous reviewers for their constructive comments and very useful suggestions that contributed to the improvement of this paper.


## REFERENCES


Blackmon, M., Kitajima, M., Polson, P. (2005). Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 51-58). Portland, Oregon: ACM.

Brinck, T., Hofer, E. (2002). Automatically evaluating the usability of web sites. *CHI '02 extended abstracts on Human factors in computing systems* (pp. 906-907). Minneapolis, Minnesota: ACM.

Budiu, R., Royer, C., Pirolli, P. (2007). Modeling information scent: a comparison of LSA, PMI and GLSA similarity measures on common tests and corpora. *Proceedings of RIAO'07.* Pittsburgh.

Capra, M. (2005). Factor analysis of card sort data: An alternative to hierarchical cluster analysis. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 691-695). Santa Monica, CA: HFES.

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

Chen, S. Y., Macredie, R. D. (2002). Cognitive style and hypermedia navigation: development of a learning model. *Journal of the American Society for Information Science and Technology , 53*, 3-15.

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

Dong, J. M., Shirley, M., Waldo, P. (2001). A user input and analysis tool for information architecture. *CHI '01 extended abstracts on Human factors in computing systems* (pp. 23-24). Seattle, Washington: ACM.

Falconer, S. M., Maslov, D., Storey, M-A. (2006). Combining web-based searching with latent semantic analysis to discover similarity between phrases. In *On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE* (pp. 1075-1091). Montpellier, France: Springer Berlin /Heidelberg.

Garrett, J. (2002). *The elements of user experience: User-centered design for the web.* New Riders Press.

Gwizdka, J., Spence, I. (2007). Implicit measures of lostness and success in web navigation. *Interacting with Computers , 19* (3), 357-369.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling.* Cary, NC: SAS Publishing.

Intranet Leadership Forum. (2006, September). *How-to guide: Card sorting*. Retrieved February 2, 2008, from Professional community in Australia and New Zealand: http://www.intranetleadership.com.au/doc/ILF-HowtoCardSorting.pdf

Jones, W., Pirolli, P., Card, S. K., Fidel, R., Gershon, N., Morville, P., et al. (2006). "It's about the information stupid!": why we need a separate field of human-information interaction. *CHI '06 extended abstracts on Human factors in computing systems* (pp. 65-68). Montreal, Quebec, Canada: ACM.

Kalbach, J. (2007). *Designing web navigation: optimizing the user experience.* O'Reilly Media, Inc.

Katsanos, C., Tselios, N., Avouris, N. (2008). AutoCardSorter: Designing the information architecture of a web site using latent semantic analysis. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (article in press). Florence, Italy: ACM.

Katsanos, C., Tselios, N., Avouris, N. (2006). InfoScent Evaluator: A semi-automated tool to evaluate semantic appropriateness of hyperlinks in a web site. *Proceedings of the 20th conference of the computer-human interaction special interest group (CHISIG) of Australia on Computer-human interaction* (pp. 373-376). Sydney, Australia: ACM.

Kaur, I., Hornof, A. J. (2005). A comparison of LSA, WordNet and PMI-IR for predicting user click behavior. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 51-60). Portland, Oregon: ACM.

Lance, E., Butts, M., Michels, L. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods , 9* (2), 202-220.

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

Landauer, T. K., Foltz, P. W., Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes , 25*, 259-284.

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

Lund, K., Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers , 28*, 203-208.

Manning, C., Schutze, H. (1999). *Foundations of statistical natural language processing.* The MIT Press.

Marchionini, G. (1997). *Information seeking in electronic environments* (New ed.). Cambridge University Press.

Mauer, D., Warfel, T. (2004, April 7). *Card sorting: a definitive guide*. Retrieved February 1, 2008, from BoxesandArrows: http://www.boxesandarrows.com/view/card_sorting_a_definitive_guide

Maurer, D. (2006, March 14). *Four Modes of Seeking Information and How to Design for Them*. Retrieved February 2, 2008, from BoxesandArrows:
http://www.boxesandarrows.com/view/four_modes_of_seeking_information_and_how_to_design_for_them

McCarthy, P. M., Lewis, G. A., Dufty, D. F., McNamara, D. S. (2006). Analyzing writing styles with Coh-Metrix. In G. Sutcliffe, and R. Goebel (Ed.), *Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS)* (pp. 764-769). Melbourne, Florida: Melbourne, Florida.

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

Morville, P. (2005). *Ambient findability: what we find changes who we become.* O'Reilly Media, Inc.

Nielsen, J. (2004, July 19). *Card sorting: How many users to test?*. Retrieved February 1, 2008, from Useit, AlertBox: http://www.useit.com/alertbox/20040719.html

Nielsen, J., Sano, D. (1995). Sun Web: user interface design for Sun Microsystem's internal web. *Computer Networks and ISDN Systems , 28* (1-2), 179-188.

Otter, M., Johnson, H. (2000). Lost in hyperspace: metrics and mental models. *Interacting with Computers , 13* (1), 1-40.

Resnik, P. (1999). Semantic similarity in a taxonomy: An information based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research , 11*, 95-130.

Rhode, D. L., Gonnerman, L. M., Plaut, D. C. (2004, March 22). *An Introduction to COALS.* Retrieved February 4, 2008, from COALS (the correlated occurrence analogue to lexical semantics): http://dlt4.mit.edu/~dr/COALS/Coals.pdf

Rosenfeld, L., Morville, P. (2006). *Information architecture for the world wide web* (3rd ed.). O'Reilly Media, Inc.

Simon, H. A. (1996) The Sciences of the Artificial, Third Ed., The MIT Press, Cambridge, MA.

Sinha, R., Boutelle, B. (2004). Rapid information architecture prototyping. *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 349-352). Cambridge, MA: ACM.

Spool, J., Scanlon, T., Snyder, C., DeAngelo, T. (1999). *Web site usability: A designer's guide.* Morgan-Kaufmann.

Sutcliffe, R., Boersma, P., Bon, A., Donker, T., Ferris, M. C., Hellwig, P., et al. (1995). Beyond keywords: Accurate retrieval from full text documents. *Proceedings of the 2nd Language Engineering Convention.*

Teevan, J., Alvarado, C., Ackerman, M., Karger, D. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 415-422). Vienna, Austria: ACM.

Toms, E. G. (2002). Information interaction: Providing a framework for information architecture. *Journal of the American Society for Information Science and Technology , 53* (10), 855-862.

Tullis, T. S., Wood, L. (2004). How many users are enough for a card-sorting study? *Usability Professionals Association (UPA) 2004 Conference.* Minneapolis, MN.

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

Van Oostendorp, H., Juvina, I. (2007). Using a cognitive model to generate web navigation support. *International Journal of Man-Machine Studies , 65* (10), 887-897.

Witten, I. H., Frank, E., (2005). *Data mining: practical machine learning tools and techniques* (2nd ed.). Morgan Kaufmann.

Wolfe, M. B., Goldman, S. R. (2003). Use of latent semantic analysis for predicting psychological phenomena: Two issues and proposed solutions. *Behavior Research Methods, Instruments, and Computers , 35*, 22-31.

Wu, S.-C., Miller, C. S. (2007). Preliminary evidence for top-down and bottom-up processes in web search navigation. *CHI '07 extended abstracts on Human factors in computing systems* (pp. 2765-2770). San Jose, CA: ACM.