The Role of Human Factors in Stereotyping Behavior and Perception of Digital Library Users: A Robust Clustering Approach

ENRIQUE FRIAS-MARTINEZ, SHERRY Y. CHEN, ROBERT D. MACREDIE, XIAOHUI LIU

Brunel University

To deliver effective personalization for digital library users, it is necessary to identify which human factors are most relevant in determining the behavior and perception of these users. This paper examines three key human factors: cognitive styles, levels of expertise and gender differences, and utilizes three individual clustering techniques: k-means, hierarchical clustering and fuzzy clustering to understand user behavior and perception. Moreover, robust clustering, capable of correcting the bias of individual clustering techniques, is used to obtain a deeper understanding. The robust clustering approach produced results that highlighted the relevance of cognitive style for user behavior, i.e., cognitive style dominates and justifies each of the robust clusters created. We also found that perception was mainly determined by the level of expertise of a user. We conclude that robust clustering is an effective technique to analyze user behavior and perception.

Keywords: Digital Libraries, Human Factors, Stereotypes, Robust Clustering, Perception, Behavior

1. INTRODUCTION

The World Wide Web (Web) enables wide dissemination of information and services (Ivory and Megraw 2005), digital libraries (DLs) being one of the most common mechanisms for this purpose. DLs can be defined as collections of information that have associated services delivered to user communities using a variety of technologies (Callan et al. 2003). In general, DLs are the logical extensions and augmentations of physical libraries in the electronic information society. Such extensions and augmentations offer new levels of access to broader audiences of users (Marchionini et al. 2003). In other words, users with DLs may have diverse background, knowledge and/or skills. From this perspective, it is necessary to provide personalization in order to match with the needs of different users. Personalization is defined as the ways in which information and services can be tailored to match the unique and specific needs of an individual or a community (Callan et al. 2003). The key element of developing personalized environments is user modeling in which unobservable information about a user is inferred from observable information from that user in order to construct a user model (Zukerman et al. 1999). User models can be constructed with a number of approaches, one of which is data mining.

There are two main types of data mining techniques, supervised learning and unsupervised learning (Witten and Frank, 1999). The former is also known as classification and the latter is also known as clustering (Sander et al. 2005). Classification requires the training data to be pre-classified. This means that each training item is assigned a unique label, indicating the class to which the item belongs. Given these data, the classification builds a characteristic description for each class, covering the examples of this class. The important feature of this approach is that the class descriptions are built conditional to the pre-classification of the examples in the training set. In contrast, clustering does not require pre-classification of the training examples and forms clusters of examples, which share common characteristics.

Our work differs from previous works in several aspects. Firstly, previous works generally used supervised learning techniques, i.e., classification, for user modeling. Decision tree is one of the most popular supervised learning techniques. Generally speaking, the decision tree is used to construct user models based on particular characteristics. For example, Cho, Kim and Kim (2002) classified customers according to their age, jobs, etc. Paliouras et al. (1999) divided users based on their department types, industry types, and market types, etc. Alternatively, they would divide the users by a particular human factor and then find statistical differences among them. For instance, Chen and Macredie (2004) identified users' preferences to web-based instruction on the basis of cognitive styles. These approaches assume that a set of users belonged to a same human factor would have the same behavior/perception, but this assumption is subject to a lot of noise because of the fuzziness of user behavioral/perceptional data. The other problem is that users are classified based on human factors, instead of behavioral/perceptional data themselves. If the human factors were not properly selected, the accuracy of the user models might be influenced. To address these problems, our work uses unsupervised learning techniques, i.e., clustering, to identify clusters of users that share similar behavior or perception, which are usually called stereotypes (Kobsa 2001). Clusters are created only using behavioral and perceptional data and then we examine the statistical significance to identify which human factors are

responsible for a given stereotype. This approach is similar to the approach presented by Shapira et al. (1997) for information filtering, but while in their approach the authors, after clustering, identified the common demographic characteristics of the users, we, on the other hand, are not interested in the common demographics, but on the relevance of a set of human factors.

Secondly, most of the aforementioned approaches make a link between a human factor and a single feature. For example, Chen and Macredie (2004) showed that a specific navigation tool (e.g., hierarchical map or alphabetical index) is favored by a particular type of cognitive style. On the other hand, our approach produces a direct relationship between a human factor and an integrated stereotype. By doing so, the human factor can be incorporated into the stereotype such that the characteristics of users can also be considered for personalization. The third aspect is that existing works mainly use a clustering algorithm on its own. For instance, Doux et al. (1997) built user profiles with the k-means clustering algorithm. Conversely, we use a robust clustering technique, which integrates the results of three clustering algorithms: k-means; hierarchical clustering; and fuzzy clustering. The stereotypes are created only when these three clustering algorithms agree. This approach can help us reduce the biases of each single clustering algorithm (Swift et al. 2004).

In brief, our study aims to investigate which human factors are responsible for the behavior and the perceptions stereotypes of DL users so that these human factors can be justified to be considered for personalization. To achieve this aim, we have studied if there is a statistical significance between the stereotypes created by robust clustering and each human factor, including cognitive styles, levels of expertise, and gender differences. The rest of the paper is organized as follows. In Section 2, we discuss the role of human factors, (especially emphasizing gender differences, levels of expertise, and cognitive styles), in determining web user behavior and perceptions. Subsequently, Section 3 highlights work on the use of clustering techniques for user stereotyping, and we focus on k-means clustering, hierarchical clustering and fuzzy clustering. Section 4 describes the experiment to collect the necessary behavioral and perceptional data for our study. Section 5 presents the results of applying clustering techniques to the behavioral data obtained from the experiment and the relationships between the human factors and the clusters. Section 6 presents the concept of robust clustering and its application for analyzing behavioral data, detailing the relationships of the clusters with each human factor. Section 7 presents the same analysis, but using perceptional data, and the relationships between the cluster obtained and each human factor. Section 8 presents a personalized interface designed with the results found in this study and investigates whether the adaptive interface can increase user satisfaction to show the usefulness of the results. The paper ends by drawing conclusions and presenting future work.

2. ROLE OF HUMAN FACTORS IN DETERMINING WEB USER BEHAVIOR AND PERCEPTION

DLs are widespread in the Web and previous research indicates that human factors have significant effects on users' interaction with web-based applications in general and the DLs in particular. Among all human factors, gender differences (Ford and Miller 1996), levels of expertise (Mitchell et al. 2005) and cognitive styles (Chen and Macredie 2004) have been recognized as especially relevant to users' interaction with the Web. In terms of gender difference, a number of studies have found that it is an important factor to determine users' navigation styles. Large et al. (2002) investigated gender differences in collaborative web searching and their results revealed that males spent less time viewing pages than females. In addition, they found that the male group was more actively engaged in browsing than the female group, and that the male group explored more hypertext links per minute. The other study by Roy and Chi (2003) examined gender differences in searching the web for information by analyzing students' navigation styles. Their findings are in agreement with the results of Large et al. (2002), indicating that males and females possess different navigation styles while searching information on the Web. Males tended to navigate in a broader way than females. They also found that males tended to perform more page jumps per minute, which indicates that they navigate the information space in a nonlinear way.

Level of expertise is another very interesting variable because it can highlight how level of satisfaction and problems of a user evolve over time (Mitchell et al., 2005). Some studies have highlighted the relevance of level of expertise for web interaction, e.g., Lazander et al. (2000). In general, individuals with higher levels of expertise require less time to search information, and have more positive perceptions towards the web (Yi and Hwang, 2003). Torkzadeh and Van Dyke (2002) examined the change of users' Internet self-efficacy, in terms of belief in their own ability to succeed, before and after computer training, with the results of their study indicating that computer training significantly improved Internet self-efficacy. In other words, when the users develop from novice to experienced, their Internet self-efficacy increases. In addition, Venkatesh (2000) showed that the effect of enjoyment on ease of use became stronger as users gained more direct experience with the system.

The other relevant factor, cognitive style, can be defined as an individual's preferred and habitual approach to organizing and representing information (Riding and Rainer 1998). There are a variety of dimensions of cognitive style, but among these dimensions, Field Dependence versus Field Independence has significant impacts on users' information processing, because it reflects how well an individual is able to restructure information based on the use of salient cues and field arrangement (Weller et al. 1994). Their different characteristics are summarized in Table 1. Recent studies have found that users' cognitive styles significantly influence their information seeking on the web. Ford, Miller and Moss (2005) conducted an experiment with AltaVista search engine. The results showed that users with a Field Dependent tendency had a higher level of using *Boolean searching* while those with a Field Independency had a higher level of using *Best Matching*. Furthermore, Palmquist and Kim (2000) found that Field Independent users tend to use search engines, the find option and URLs more frequently to reach the desired Web sites, whereas Field Dependent users tend to follow links prescribed by the Web page, especially when they are novices.

Results from the aforementioned studies suggest that gender differences, levels of expertise, and cognitive styles have significant effects on users' behavior on the web and their perceptions towards the use of the web. Nevertheless, current work does not reveal any links between a stereotype of perception/behavior and any of these human factors, but only links between different features and human factors. There is a need to study the links between user stereotypes and human factors so that these links can be incorporated into the development of personalized web-based applications to accommodate the needs of different users.

Table 1. Differences between Field Dependence and Field Independence (Witkin, et al., 1977).

Field Dependent Users	**Field Independent Users**
▪ They are externally directed and are easily influenced by salient features;	▪ They are internally directed and process information with their own structure;
▪ they experience surroundings in a relatively global fashion and struggle with individual elements;	▪ they experience surroundings analytically and are good with problems that require taking elements out of their whole context;
▪ they are more passive to accept ideas as presented.	▪ they are more active to accept ideas strengthened through analysis.

3. CLUSTERING FOR USER STEREOTYPING: RECENT APPROACHES

A number of clustering techniques can be used for the identification of stereotypes. These techniques can be classified as hard clustering and fuzzy clustering. Table 2 summarizes studies and applications of hard clustering, presenting their application and the input data that they used. Table 3 summarizes studies and applications of fuzzy clustering.

Table 2. Examples of applications using hard clustering.

Authors	Application	Input Data
Doux et al. (1997)	K-means clustering algorithm for user profiling in order to derive prototypical behavior from each user.	Data collected from a dedicated set of experiments where users are asked about their preferences.
Paliouras et al. (1999)	News-filtering system based on stereotypes of users. Clustering allows recommending interesting news to a user.	When registering a user specifies his/her interests.
Fu et al. (1999)	Grouping of users (stereotypes) with similar behavior in a web server taking into account access patterns.	Data collected from UMR web server log (www.umr.edu) containing 2.5 million records.
Hay et al. (2001)	Clustering methods that capture the inherent sequentiality of web visits. A metric, Sequence Alignment Method, is introduced to be used instead of Euclidean distance for clustering purposes.	Log files of a Belgian telecom provider collected over a one-week period.
Mobasher et al. (2001)	Stereotyping for collaborative filtering	12,000 sessions collected from the Association for Consumer Research web site.

Table 3. Examples of applications using Fuzzy Clustering.

Authors	Application	Input Data
Nasraoui et al. (1999)	A new algorithm (CARD) to mine user stereotypes from access logs is proposed.	12 days log data at the University of Missouri.
Joshi et al. (2000)	Two algorithms to mine user stereotypes: Fuzzy c-Medoid Algorithm (FCMdd) or the Fuzzy c-Trimered Medoids Algorithm (FCTMdd).	CSEE logs of University of Maryland
Krishnapuram et al. (2001)	Web access log analysis for user profiling using Robust Fuzzy c-Medoids (RFCMdd).	Five days of CSEE web server activity of the Univ. of Maryland.
Lampinen and Koivisto (2002)	Obtain application stereotypes from network traffic data to manage network resources.	274000 samples of different applications from an edge router of a LAN network.

In hard clustering, data are divided into crisp clusters, where each data point belongs to exactly one cluster. Hard clustering techniques can be subdivided into two categories: hierarchical and non-hierarchical (Jain and Dubes 1999). A hierarchical clustering procedure involves the construction of a hierarchy or tree-like structure, which is basically a nested sequence of partitions, while non-hierarchical or partitional procedures end up with a particular number of clusters at a single step. In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the instances are membership grades which indicate the degree to which they belong to the different clusters. In other words, a user can be at the same time in more than one cluster with different degrees of truth. This allows to better capture the inherent uncertainty that modeling user behavior typically has. Both hard clustering and fuzzy clustering are commonly used for building stereotypes. In particular, k-means clustering (hard and non-hierarchical), hierarchical clustering (hard and hierarchical) and fuzzy clustering are very popular clustering techniques and their details are presented in the following subsections.

3.1 K-means Clustering

The k-means clustering technique (MacQueen 1967) is given as an input the number of clusters k. The algorithm then picks k items, called seeds, from the training set in an arbitrary way. Then, in each iteration, each input item is assigned to the most similar seed, and the seed of each cluster is recalculated to be the centroid of all items assigned to that seed. This process is repeated until the seed coordinates stabilize. This algorithm aims at minimizing an objective function, J, typically a squared error function:

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n} d_{ij} = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\| x_i^{(j)} - c_j \right\|^2 \tag{1}$$

where d_{ij} is the distance measure between a data point x_i and the cluster centre c_j. J is an indicator of the distance of the n data points from their respective cluster centres and it represents the compactness of the clusters created.

The k-means algorithm is popular because it is easy to understand and easy to implement. Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the optimal configuration, corresponding to the global objective function minimum. Another problem is that it is sensitive to the initial partition and may converge to a *local* minimum of the criterion function value if the initial partition is not properly chosen. A possible remedy is to run the algorithm with a number of different initial partitions. If they all lead to the same final partition, this implies that the global minimum of the square error has been achieved. However, this can be time-consuming and may not always work.

3.2 Hierarchical Clustering

The main problem of non-hierarchical clustering approaches is that when working with high dimensional problems, in general, there will not be enough items to populate the vector space, which will imply that most dimensions will be unreliable for similarity computations. In order to solve this problem, hierarchical clustering techniques were developed. There are two types of hierarchical clustering: agglomerative and divisive. Both share a common characteristic: they create a hierarchy of clusters. While the agglomerative approach creates a bottom-up hierarchy, the divisive approach creates a top-down one. Generally speaking, divisive algorithms are computationally less efficient because of the complexity of deciding which cluster to divide and how to do it (Savaresi et al. 2002).

A typical hierarchical agglomerative clustering algorithm is outlined below:

1) Place each pattern in a separate cluster
2) Compute the proximity matrix of all the inter-pattern distances for all distinct pairs of patterns
3) Find the most similar pair of clusters using the matrix. Merge these two clusters into one, decrement number of clusters by one and update the proximity matrix to reflect this merge operation.
4) If all patterns are in one cluster, stop. Otherwise, go to the above step 2.

The output of such algorithm is a nested hierarchy of trees that can be cut at a desired dissimilarity level forming a partition. Hierarchical agglomerative clustering algorithms differ primarily in the way they measure the distance or similarity of two clusters where a cluster may consist of only a single element at a time. The most commonly used inter-cluster measures are:

$$d_{AB} = \min_{\substack{i \in A \\ j \in B}}(d_{ij}),$$

(2)

$$d_{AB} = \max_{\substack{i \in A \\ j \in B}}(d_{ij}),$$

(3)

$$d_{AB} = \frac{1}{n_A n_B} \sum_{i \in A} \sum_{j \in B} d_{ij},$$

(4)

where d_{AB} is the dissimilarity between two clusters A and B, d_{ij} is the dissimilarity between two individual patterns i and j, n_A and n_B are the number of individuals in clusters A and B respectively. These three inter-cluster dissimilarity measures are the basis of three of the most popular hierarchical clustering algorithms. The *single-linkage* algorithm uses Equation (2), the *minimum* of the distances between all pairs of patterns drawn from the two clusters (one pattern from each cluster). The *complete-linkage* algorithm uses Equation (3), the *maximum* of all pairwise distances between patterns in the two clusters. The *group-average* algorithm uses Equation (5), the average of the distances between all pairs of individuals that are made up of one individual from each cluster.

A challenging issue with hierarchical clustering is how to decide the optimal partition from the hierarchy. One approach is to select a partition that best fits the data in some sense, and there are many methods that have been suggested in the literature (Everitt 1993). It has also been found that the single-linkage algorithm tends to exhibit the so-called *chaining* effect: it has a tendency to cluster together at a relatively low level of elements linked by chains of intermediates. As such, the method is appropriate if one is looking for "optimally" connected clusters rather than for homogeneous spherical clusters. The complete-linkage algorithm, on the other hand, tends to produce clusters that are tightly-bound or compact, and has been found to produce more useful hierarchies in many applications than the single-link algorithm (Jain and Dubes 1999). The group-average algorithm is also widely-used. Detailed discussion and practical examples of how these algorithms work can be found in Jain and Dubes (1999) and Webb (1999).

3.3 Fuzzy Clustering

One of the most widely used fuzzy clustering algorithms is the Fuzzy C-Means (FCM) Algorithm (Bezdek 1981). The FCM algorithm attempts to partition a finite collection of elements $X=\{x_1,...,x_n\}$ into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c cluster centres $C=\{c_1,...,c_c\}$ and a partition matrix $U=u_{i,j} \in (0,1)$, $i=1,...n, j=1,...,c$, where each element tells the degree to which element x_i belongs to cluster c_j. Like the k-means algorithm, the fuzzy c-means aims to minimize an objective function. The standard function is:

$$J = \sum_{j=1}^{c} \sum_{i=1}^{n} (u_{i,j})^m \left\| x_i^{(j)} - c_j \right\|^2$$

(5)

which differs from the k-means objective function by the addition of the membership values u_{ij} and the fuzzifier m. The fuzzifier m determines the level of cluster fuzziness. A large m results in smaller memberships u_{ij} and hence, fuzzier clusters. In the limit $m=1$, the memberships u_{ij} converge to 0 or 1, which implies a crisp partitioning. In the absence of experimentation or domain knowledge, m is commonly set to 2. The basic Fuzzy C-Means Algorithm,

given n data points $(x_1,…,x_n)$ to be clustered, a number of c clusters with $(c_1,…,c_c)$ the center of the clusters, and m the level of cluster fuzziness with,

$$m \in (1,\infty),$$ (6)

first initializes the membership matrix U to random values, verifying that:

$$u_{ij} \in [0,1], \sum_{j=1}^{c} u_{ij} = 1$$ (7)

After the initialization, the algorithm obtains the center of the clusters $c_j, j=1,…,c$:

$$c_j = \frac{\sum_{i=1}^{n}(u_{ij})^m x_i}{\sum_{i=1}^{n}(u_{ij})^m}$$ (8)

And obtains the distance between all points $i=1,…,n$ and all cluster centers $j=1,…,c$

$$d_{ij} = \left\| x_i^{(j)} - c_j \right\|$$ (9)

Updating matrix U according to the new distances,

$$d_{ij} = 0 \Rightarrow u_{ij} = 1$$

$$u_{ij} = \left[\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}} \right]^{-1}$$ (10)

This process is repeated until the set of cluster centers is stabilized. There are other algorithms, which are optimizations of the original FCM, like Fuzzy c-Medoid Algorithm (FCMdd) or the Fuzzy c-Trimered Medoids Algorithm (FCTMdd) (Krishnapuram et al. 2001). One of the main problems of Fuzzy Clustering is that the number of clusters is a parameter of the algorithm. In order to identify the optimum number of clusters, some algorithms, such as Substractive Clustering (Chiu 1994), have been proposed.

One of the major challenges for cluster analysis is that there is a lack of inter-method consistency. Each clustering algorithm imposes a specific structure on the data due to different distance or correlation measurements, different search strategies and different objective functions used. The choice of a specific clustering algorithm for a given application is still very much an open issue. To address this issue, another algorithm named robust clustering has been developed. This algorithm basically creates clusters based on the information of other clustering techniques, producing clusters only if all the clustering techniques agree with each other. As a collateral effect, elements for which the techniques do not agree are filtered out from the final classification (Swift et al. 2004). This method would be ideal for user modeling because: (1) it eliminates the bias of the techniques, due to the fact that clusters are created only if all techniques agree; and (2) it filters out users that do not have a well-defined behavior, because for these users there is no agreement among clustering techniques to which cluster the user should belong. Given these arguments, we used robust clustering to analyze users' behavior and perception in an empirical study conducted in a DL, described in the next section.

4. EXPERIMENT DESIGN

This section describes the different characteristics of the experiment that were designed to capture users' behavior and perceptional data. Such data will be used as input to clustering techniques in order to identify groups of users that share similar behavior or perception. The following subsections present the characteristics of the participants, the

research instruments used, including the DL on which this study focuses, and the tasks designed and data collection techniques used.

4.1 Participants

A total of 48 individuals participated in this study and such size of participants is in line with other studies in user modeling, e.g., Goren-Bar, et al. (2006) and Tarpin-Bernard and Habieb-Mammar (2005). Participants were students at Brunel University in the United Kingdom and they volunteered to take part in the study. A request was issued to students in lectures, and further by email, making clear the nature of the studies and their participation. All participants had the basic computing and Internet skills necessary to operate the Brunel digital library catalog.

4.2 Research Instruments

The research instruments used include: (1) Cognitive Style Analysis (Riding 1991) to measure participants' cognitive styles, (2) the Brunel digital library catalogue, which is the focus of the study, (3) a tool for capturing user interaction and storing a user questionnaire, WebQuilt, and (4) a set of questionnaires for capturing the perception of the users.
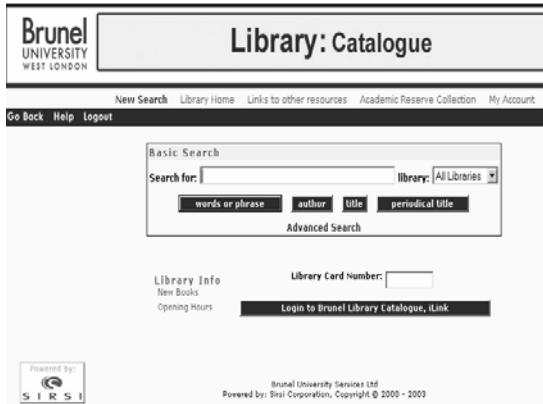
4.2.1 Cognitive Style Analysis

A number of techniques have been developed to measure Field Dependence/Field Independence, from which we chose the Cognitive Styles Analysis (CSA) by Riding (1991). The CSA test includes two sub-tests: (1) the first presents items containing pairs of complex geometrical figures that the individual is required to judge as either the same or different and (2) the second sub-test presents several items each comprising a simple geometrical shape, such as a square or a triangle, and a complex geometrical figure and the individual is asked to indicate whether or not the simple shape is contained in a complex one by pressing one of the two marked response keys (Riding and Grimley 1999). These two sub-tests have different purposes. The first sub-test is a task requiring *Field Dependent* capacity, while the second sub-test requires the disembedding capacity associated with *Field Independence.* This provides an important advantage over other methods that only measure one of the factors.
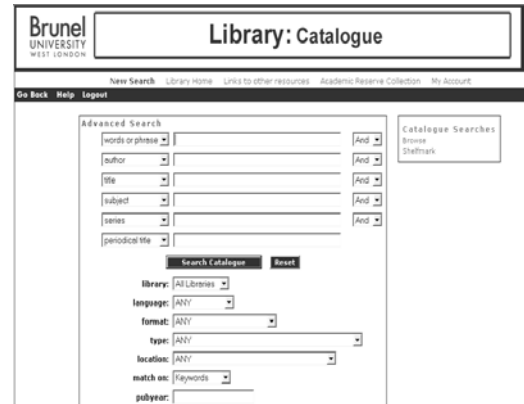
The CSA measures what the authors refer to as a *Wholist/Analytic* (WA) dimension, noting that this is equivalent to Field Dependence (Riding and Rayner 1998). The WA dimension is a real number between the values of 0.6 and 3.0 that indicates the degree of Field Dependence. Riding's (1991) recommendations are that scores below 1.03 denote Field Dependent individuals; scores of 1.36 and above denote Field Independent individuals; and scores between 1.03 and 1.35 are classified as Intermediate. In this study, the categorizations were based on these recommendations.

4.2.2 Brunel Library Catalogue

Brunel Library Catalogue (BLC) is a typical digital library and is used to access the bibliographical resources of Brunel University. BLC has two main mechanisms that provide different strategies for finding information: (1) Basic Search (Figure 1a), which is the one presented by default by the system, and (2) Advanced Search (Figure 1b), which is accessed through the corresponding link presented in Figure 1a. Basic Search provides users with major search options, including "*word or phrase*", "*author*" "*title*" or "*periodical title*". Moreover, the users can choose in which library he/she wants to search for that information. The help link briefly describes what each link is supposed to do. Advanced Search, as presented in Figure 1b, presents the users with a much broader way of searching for information, including author, title, subject, periodical title, series, etc and combine these items using and/or Boolean operators. The Advanced Search also allows users to select other information, like the library, the language, the publication year, etc.
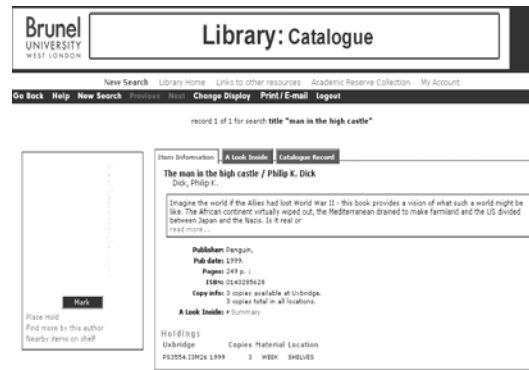
(a)

(b)

(c)

(d)

Fig. 1a). Basic Search Interface of BLC, (b) Advanced Search Interface of BLC, (c) Multiple Results Interface of BLC, and (d) Single Result Interface of BLC.

Once a user submits a query to the system using the Basic Search or the Advanced Search, the system responds with the items found in the database. An example of the interface presented is given in Figure 1c. The system presents a set of buttons in the top part: "*Go Back*", "*Limit Search*", "*New Search*", "*Backward*", "*Forward*", "*Prefs*" and "*Exit*". The "*Limit Search*" option is a link to the bottom of the page where the search mechanism used (Basic Search or Advanced Search) is presented with the terms used and a set of options for Search Limits (language, publication year, etc.). The limit search is obtained by adding more words to the set of terms already introduced. The "*New Search*" option presents again the interface of Figure 1a. The "*Backward/Forward*" button allows the user to move up and down the items found. Once a user selects one item, the information and interface given is presented as in Figure 1d.

4.2.3 WebQuilt

The WebQuilt Proxy Server (Hong et al. 2001) is a proxy system that unobtrusively gathers click stream data as users complete specified tasks. It is designed to conduct remote usability testing on a variety of Internet-enabled devices and provides a way to identify potential usability problems. WebQuilt utilizes Java Servlet and JSP technology to track users' interactions and then store that data by (1) creating a log file of each user's web use and (2) additionally caching the pages a user accesses for later viewing. Once the proxy server is running, each user connects to every web page through the web server. The proxy server stores any interaction between the user and any web pages and a snapshot of each page visited by each user. These snapshots are given a number that is the same one used to describe the sequence of pages visited by the user. Figure 2 presents an example of the records collected by Webquilt and Table 4 details the information captured.

The Web proxy server has the possibility of adding a task box that can be used to indicate when a task has been finished. Once a user finishes each task and uses the task box links to finish it, Web Quilt provides each user with a set of questions regarding the task. All the information captured is stored in the proxy server and there is an identification number for each user. This allows centralizing all the information in the same place and at the same

time being able to access the information of each user independently. The use of a proxy server architecture allows us to easily capture all the interactions between users and the BLC, which would otherwise be far more difficult as significant software changes would need to be implemented in the BLC.

```
0       0 1    -1      200 -1   -1      GET   http://library.brunel.ac.uk:8080/uhtbin/webcat
1688    1 2    -1      200 -1   -1      POST http://library.brunel.ac.uk:8080/uhtbin/cgisirsi/0/57/49      user_id=WEBSERVER
19629   2 3    -1      200 -1   -1      POSThttp://library.brunel.ac.uk:8080/uhtbin/cgisirsi/88 search_type=search&
        searchdata1=the+man+in+the+high+castle&library=ALL&sort_by=ANY&relevance=off&S_ICON5ETITLES%5ETitle+Processing
163881  3 4    -1      200 -1   4       GET   http://library.brunel.ac.uk:8080/uhtbin/cgisirsi/49
```

Fig. 2. Example of the records collected by Webquilt.

Table 4. Information stored by Webquilt for each request.

Field	Description
Time	The amount of time, in milliseconds, since the start of the user's session.
From	The transaction ID of the previous page the user came from.
To	The current transaction ID.
Parent	The transaction ID of the current page's frame parent, or -1 if none.
Code	The HTTP response code. 200 means OK, 404 means page not found.
Frame	The frame number of the current page . -1 if the page is not a frame.
Link	The link the user clicked to get to this page. This counts both <A> and <AREA> tags. This value is -1 if the page was not reached through a link.
Method	The HTTP method used to retrieve the page (e.g. GET or POST).
URL	The current URL.
Query	The query data sent along with the page request, if any.

4.2.4 Perception Questionnaires

In order to capture users' perceptions when using BLC, three standardized questionnaires were used in this study: the Questionnaire for User Interface Satisfaction (QUIS) (Chin et al. 1988), Computer Usability Questionnaire (CSUQ) (Lewis 1995), and After-Scenario Questionnaire (ASQ) (Lewis 1995). QUIS is a tool designed to assess users' subjective satisfaction with specific aspects of the human-computer interface. Although QUIS is a very complete questionnaire, for the purpose of this study we used a summarized QUIS test available on-line (http://www.acm.orf/~perlman). In this version, the questionnaire is divided into five sections (Overall reaction to the software, Screen, Terminology and System Information, Learning and System Capabilities) with a total of 27 questions. Each area measures the users' overall satisfaction with that facet of the interface, as well as the factors that make up that facet, using a (0-9) scale. CSUQ, developed by IBM, was designed to evaluate the usability of a computer program, not necessarily a web service. It contains 19 questions, each being a statement that the user has to rate on a (1-7) scale ranging from "strongly disagree" to "strongly agree". ASQ is a CSUQ complementary test, and is designed to be done once the user has finished all the tasks. It contains three questions that ask for the user's overall reactions to the system. As in CSUQ, the answers are in the range (1-7).

4.3 Task Design

The purpose of this study is to identify the relevant human factors for the behavior and the perception stereotypes of DL users. The perception was measured using the questionnaires described in Section 4.2.4. Two types of behavior were considered: browsing and searching (Bryan-Kinns et al. 2000). In this context, browsing is defined as the search of ill-defined information while searching is defined as the localization of specific, well-defined information.

In order to identify these two types of behavior, participants were asked to perform a set of seven practical tasks. The set of tasks was designed to involve all the functionalities that BLC provides to each user and the different behaviors (searching and browsing) that a user can show (Table 5). The first question captures a searching behavior, as it has a clear well-defined answer contained in the library catalogue. It is also designed to capture if the user uses the "*Word or Phrase*", "*Author*" or "*Title*" options (which are different ways of approaching the problem) or if an Advanced Search is used. When the Advanced Search is used, it will be interesting to capture which elements are used (title, author, year), and if any search limit is introduced. The second task is a browsing question designed to test if the user takes the "*Subject*" option or prefers an approach using "*Title*" or "*Word or Phrase*". The third question is a combination of searching-browsing task designed especially to see if users take the option of selecting a specific campus or do it manually. It is also interesting to see how the user approaches the problem (searching by author or by

general search). The rest of the tasks are designed to replicate some of the functionalities and/or behaviors in order to have more relevant data to work with.

Table 5. Set of tasks designed for the experiment and their type.

No.	Task	Type
1	Find the Call Number of the book "The Man in the High Castle" by Philip Kendred Dick.	Search
2	Find the title of any book related to applications of fuzzy logic.	Browse
3	Find the number of books written by Aldous Huxley that are part of the TWICKENHAM Library	Search-Browse
4	Find a book about how to implement data mining with Java.	Browse
5	Find a Java book written by Hugh Vincent.	Search
6	Find a book about 20th century American Drama in TWICKENHAM Library.	Browse
7	Find an IEEE journal on consumer electronics.	Search

4.4 Experimental Procedure

The experiment was conducted using the Brunel Library Catalogue (BLC) and comprised five different steps:

1) Participants were given a task sheet, which described the task activities that they needed to complete with the BLC. One participant carried out the experiment at a time.
2) The CSA was used to classify participants' cognitive styles into Field Independent, Intermediate or Field Dependent.
3) Participants were observed while they were carrying out the tasks presented in Section 4.3, and clarifications were given when requested. All interactions between the participants and the BLC were stored by Webquilt.
4) Each participant answered the QUIS, CSUQ and ASQ questionnaires.
5) Participants ended by answering the following questions: (a) gender, (b) level of expertise in BLC (Never used the system, Novice, Medium or Expert), where the assignment of a user to a group is done by the user according to his/her own perception of his/her expertise.

4.5 Data Collection and Summarization

The interaction data collected from each user was centrally stored on the proxy server. For each user, the interaction data captured for solving the seven given tasks was summarized into six dependant variables, variables 1 to 6 in Table 6. Each variable was then normalized "to one task" by dividing each value by seven. After normalization, dimensions 1 to 6 captured the way in which each user interacts with the BLC to solve one generic task. Each vector also contained the 49 answers given to the questionnaires and also four independent variables: users' cognitive style, WA ratio, level of expertise, and gender. In the end, we have for each user a 60-dimensional vector containing information regarding behavior and perception.

Table 6. Dimensions of a BLC user vector.

No.	Variable	Information
1	BS	Number of times that the user used the Basic Search functionality to solve the seven tasks.
2	AS	Number of times that the user used the Advance Search functionality to solve the seven tasks.
3	SE	Number of times that the user used the "Word or Phrase" option from the Basic Search Interface to solve the seven tasks.
4	ATS	Number of times that the user used the "author", "title" and "periodical" options from the Basic Search interface to solve the seven tasks.
5	NS	Number of times that the user pressed the New Search functionality to solve the seven tasks.
6	GB	Number of times that the user pressed the Go Back button to solve the seven tasks.
7-55		Answers to QUIS, CSUQ and ASQ.
56	CS	User cognitive style obtained using CSA test (Field Dependent, Intermediate of Field Independent).
57	WA	WA ratio of the user provided by the CSA test.
58	LE	Level of expertise indicated by the User (Expert, Medium, Novice or Do not Usually use the system).
59	P	Position within the university.
60	G	Gender.

5. HUMAN FACTORS AND USER BEHAVIOR

In this section, k-means, hierarchical clustering and fuzzy clustering are used to cluster users in stereotypes according to their similar behavior. Once users have been clustered, we study the relevance of cognitive style, gender and level of expertise to determine different user behaviors. Matlab's Statistics Toolbox 5.1 was used for k-means and hierarchical clustering, while Fuzzy Logic Toolbox 2.2 was used for Fuzzy Clustering. The information used to represent the behavior of each user consists of all cases of a vector containing dimensions 1 through 6 of the elements presented in Table 6. The information applied to cluster users does not contain any human factors.

5.1 Stereotyping with K-means

In k-means algorithm, the main parameter is the number k of clusters (stereotypes) used to partition the original data. In order to determine the optimal number of clusters, we ran the algorithm for $k=2,...,9$ without giving any initial value for the cluster centers and using Euclidean distance. To avoid the solution given for a given k being a local minima, because of the randomness of the original centers, k-means was run for each value of k 100 times and the solution used was the one that minimized the objective function J presented in Equation (1). In order to determine how good the partition obtained was, for each user i we obtained an indication ϕ_i representing how similar the behavior of that user is with users of the same cluster compared with the behavior of users of all other clusters, formally:

$$\phi_i = \frac{\min(b_{i,m}, m=1,...,k) - d_i}{\max(d_i, \min(b_{i,m}, m=1,...,k))} ,$$ (11)

where ϕ_i is a value ranging in $(-1, +1)$, d_i is the average distance of user i to all the users of its own cluster, $b_{i,k}$ is the average distance of user i to all the users of cluster k, and m is the number of user stereotypes. A value of $+1$ indicates that the user is very distant to the rest of the clusters, a value of 0 or near 0 indicates that the user is not distinctive of that cluster, and a negative value that indicates that the user has probably been assigned to the wrong cluster. The quality of a partition, q_k, ($k=2,...,9$), is obtained as the mean value of all the ϕ_i values of the system,

$$q_k = \frac{\sum_{i=1}^{N} \phi_i}{N} ,$$ (12)

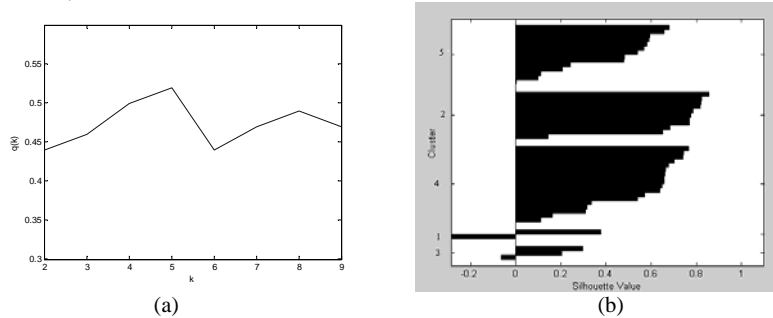with N the number of users, 48 in our case.

Fig. 3. (a) Evolution of the quality of the clusters for different values of k, and (b) representation of the optimum five cluster partition found, using behavioral data.

Figure 3(a) presents the evolution of the quality of the partitions obtained for the values of k tested. As can be seen the optimum partition, as we have measured the quality of a partition, is obtained with a value of $k=5$. Figure 3(b) presents a representation of the five clusters produced and ϕ_i for each user within that clusters. We can observe that, from the five clusters, three are easily distinguishable with a high number of users and a high ϕ_i value for its elements, indicating well defined behavior of its users, while two clusters have a low number of users included and lower ϕ_i values.

Table 7. Cluster centers obtained by k-means.

	Users	BS	AS	SE	ATS	NS	GB
Cluster 1	2	2.42	0.71	1.21	1.21	0.92	0.78
Cluster 2	11	1.61	0.02	0.09	1.44	0.09	0.23
Cluster 3	3	1.09	1.57	0.28	0.5	0	0.5
Cluster 4	18	1.6	0.04	0.89	0.4	0.07	0.1
Cluster 5	14	0.94	0.36	0.3	0.54	0.17	0.03

Table 7 presents the center of each cluster, indicating the value of each dimension and also the total number of users included in the cluster. Those centers of clusters can be translated into the behavior of the users. We do not detail Cluster 1 and Cluster 3 because they do not contain a sufficient number of users, while the behavior of Cluster 2, Cluster 4, and Cluster 5 are described below.

- Cluster 2: Users use exclusively the "Basic Search" option instead of the "Advanced Search" and search by "Author/Title" (ATS) almost exclusively, without using the "Word or Phrase" option.
- Cluster 4: Users take exclusively the "Basic Search" option instead of the "Advanced Search" and use the option "Word or Phrase" twice as much as the "Author/Title"(ATS).
- Cluster 5: Users use the "Basic Search" and "Advanced Search" options (the first three times as often as the second) and that use the "Word or Phrase" and the "Author/Title" options in the same proportion.

In order to identify the role of human factors on determining the behavior of a user, ANalysis Of VAriance (ANOVA) was used to obtain the significance values (*p-value*) of gender, level of expertise and cognitive style in forming the clusters found because it is powerful in analyzing the results of empirical studies and has been used in relevant research, e.g., Downing, Moore, and Brown (2005) and Ling and Schaik (2006). In general, a significance level of $p < 0.05$ was adopted (Stephen and Hornby 1997). In other words, only *p-values* that were less than 0.05 are considered as an indication of relevance. The results indicated that levels of expertise and gender differences did not have any relevance in determining behavior clusters, with *p=0.593* and *p=0.238* respectively. However, we found the significance of cognitive style with the clusters obtained by k-means, having a value of *p=0.006*. This shows that cognitive style plays a very important role in determining users' behavior. 38.9% of Field Independent users appear in Cluster 2, 50% of Intermediate users in Cluster 5, and 62.5% of Field Dependent users in Cluster 4.

5.2 Stereotyping with Fuzzy Clustering

In Fuzzy Clustering, the main parameter is the number of clusters used to divide the data. In order to estimate the number of clusters, we used the subtractive clustering method (Chiu 1994). Subtractive clustering is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data. Basically the algorithm assumes that each data point is a potential cluster center and calculates a measure of the likelihood of that point of being cluster center based on the density of surrounding data points.

In Subtractive Clustering, the main parameter is the *radii* of influence, which is used to determine the size of the possible clusters. Good values for the *radii* are around *(0.2, 0.5)*. Smaller values have clusters with very small influence area, which causes a high number of clusters, and higher values create a very small number of clusters because the influence of the cluster would be more than half of the input space. Different values of *radii* can be applied to each dimension. However, for simplicity, we applied the same value to all dimensions of the input data. The algorithm has the following steps:

1) For each data point, the algorithm obtains the density of surrounding points using the *radii*.
2) The set of data points with the highest potential to be a cluster (with the higher density) is selected.
3) The data points within the vicinity of the cluster determined by the *radii* are removed.
4) The process is repeated until all data is included in one of the clusters.
5) The algorithm outputs the optimum number of clusters and candidates for the centers of those clusters.
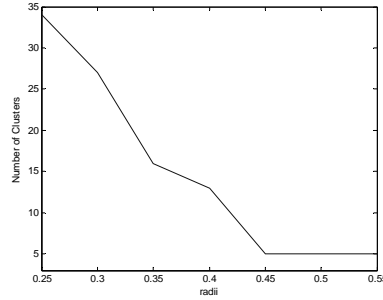
Fig. 4. Evolution of the number of cluster depending on the *radii* value.

We used the behavior data to run the subtractive algorithm for values of *radii* from 0.25 to 0.55 to determine the optimum number of clusters. Figure 4 presents the evolution of the number of clusters for each one of the *radii* values. As can be seen, values between *(0.25, 0.45)* produce a high number of clusters for the total number of 48 users. Nevertheless, in the range *(0.45, 0.55)* the number of clusters stabilizes to five. Considering this result, and also that five was the optimum number found for k-means, we decided to divide the original data using fuzzy clustering into five clusters.

Table 8. Cluster centers obtained by Fuzzy Clustering.

	Users	BS	AS	SE	ATS	NS	GB
Cluster 1	8	1.6	0.06	1.12	0.31	0.08	0.11
Cluster 2	10	1.59	0.06	0.07	1.4	0.09	0.21
Cluster 3	5	1.64	0.08	0.7	0.65	0.13	0.11
Cluster 4	14	1.26	0.23	0.44	0.64	0.19	0.12
Cluster 5	11	0.74	0.62	0.3	0.25	0.07	0.036

Fuzzy clustering assigns to each user a degree of inclusion in each cluster. We considered each user to be included in the cluster for which there was the highest degree of truth. Table 8 presents the center of each cluster, indicating the value of each dimension. Those centers of clusters can be translated into the behavior of the users included in them:

- Cluster 1: Users who exclusively use "Basic Search" in combination with mainly "Word or Phrase" and "Author/Title" occasionally.
- Cluster 2: Users who exclusively use "Basic Search" in combination with "Author/Title".
- Cluster 3: Users who exclusively use "Basic Search" in combination with "Word or Phrase" or "Author/Title" in the equal proportion.
- Cluster 4: Users who use mainly "Basic Search", and "Advanced Search" occasionally, and who use "Word or Phrase" or "Author/Title" in the equal proportion.
- Cluster 5: Users who use "Basic Search" and "Advanced Search" in the equal proportion in combination with "Search Word or Phrase" or "Author/Title".

As we did with k-means, ANOVA was used to identify the influential human factors. We found that levels of expertise and gender differences did not have any relevance in determining behavior clusters, with *p=0.852* and *p=0.177* respectively. Again, we found the significance of cognitive style with the clusters obtained by Fuzzy Clustering, having a value of *p=0.010*. This shows that cognitive style plays an important role in determining behavior groups. 33.3% of Field Independent users appear in Cluster 2, 45.5% of Intermediate users in Cluster 4, and 50% of Field Dependent users in Cluster 3.

5.3 Stereotyping with Hierarchical Clustering

We then used hierarchical agglomerative clustering to identify users that share similar behavior, using a Euclidean distance to construct clusters. Figure 4 presents the hierarchical clustering tree obtained, where the X-axis presents the users (some of them have been grouped to present a clear representation) and the Y axis illustrates the distance between the two elements being connected. The hierarchical tree has been cut at a height of 1.1, which creates five different clusters. This has been motivated by: (1) graphically from Figure 5, it can be inferred that the systems identifies five different groups and (2) the two previous techniques, for the same dataset, have established that the optimum number of clusters is five.
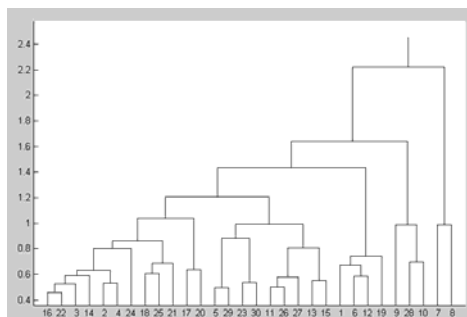
Fig. 5. Schematic representation of the hierarchical tree constructed using behavioral data

Table 9. Cluster centers generated by Hierarchical Clustering

	Users	BS	AS	SE	ATS	NS	GB
Cluster 1	22	1.59	0.02	0.7	0.3	0.17	0.5
Cluster 2	13	1.6	0.06	0.07	1.44	0.09	0.03
Cluster 3	8	0.86	0.45	0.55	0.89	0.08	0.09
Cluster 4	3	1.26	0.04	0.32	1.16	0	0.1
Cluster 5	2	2.1	0.9	1.12	1.12	0.03	0.2

Table 9 presents the number of users included in each cluster and the center of each cluster. In this case, each dimension of the cluster center has been obtained as the mean of all the users included in that cluster. We do not detail Cluster 4 and Cluster 5 because they do not contain a sufficient number of users while the behavior of Cluster 1, Cluster 2, and Cluster 3 are described below.

- Cluster 1: Users who exclusively use "Basic Search" in combination with "Word or Phrase" or "Author/Title".
- Cluster 2: Users who exclusively use "Basic Search" in combination with "Author/Title".
- Cluster 3: Users who use "Basic Search" and "Advanced Search" in the equal proportion in combination with "Word or Phrase" or "Author/Title".

With regard to relevant human factors, the ANOVA showed that levels of expertise, with $p=0.366$, and gender differences, with $p=0.645$, did not play an influential role again. We found the significance of cognitive style with the clusters obtained by Hierarchical Clustering, having a value of $p=0.005$. Therefore, cognitive Style was an influential factor in determining the clusters. 75% of users in Cluster 3 are Field Dependent and 69% of users in Cluster 2 are Intermediate. As for the rest of the clusters, there is no clear relationship with cognitive style, and they appear to present a set of users with a mix of cognitive styles.

5.4 Comparative Analysis

The aforementioned techniques roughly identify the same behaviors for the pool of users. Nevertheless, although they agree in the number of clusters created, they do not necessarily agree in the classification within a cluster of a given user. These differences within clustering techniques are typical, and they have motivated the development of techniques that compare the level of agreement between two classifiers. These techniques can also be presented as a way of assessing the consistency of a partition. A simple method for comparing two data partitions is the *Kappa* metric (Altman 1997; Uebersax 1987; Valiquette 1994). This metric rates the agreement between the classification decisions made by two observers. The metric has a value in the range (-1, +1), where -1 indicates that there is no concordance between the observers, and +1 indicates that there is complete concordance. From a clustering perspective, a high kappa value would indicate that the two arrangements are similar, while a low value would indicate that they are dissimilar.

Table 10. Kappa values for each technique comparison when using behavioral data

	K-means Clustering	Hierarchical Clustering	Fuzzy Clustering
K-means Clustering	1	0.764	0.509
Hierarchical Clustering	-	1	0.429
Fuzzy Clustering	-	-	1

Table 10 presents the *Kappa* value of each pair of clustering techniques used. As shown in this table, k-means and hierarchical clustering have good agreement strength of 0.764, while k-means-fuzzy clustering and hierarchical-fuzzy clustering only have moderate agreement strength. These results show that the partitions created are not very consistent. This may be due to the fact that each technique has a bias that deeply affects its classification. For example, k-means uses the concept of distance selected and the randomness of the process of selecting the initial centers of the cluster; fuzzy clustering utilizes the concept of *radii* and again the randomness of initializing the cluster centers; and hierarchical clustering employs the distance used to aggregate users. Also, in our context, the data we are handling are very noisy because of the inherent fuzziness of capturing human behavior. This implies that, because the techniques do not filter out any users, and because users can show behavior that actually is not relevant in modeling user behavior, the behavior captured by each cluster is blurred by the addition of these 'noisy' users. In order to identify a reliable human factor in determining user behavior, we need to use a technique that counteracts the bias of the techniques and the noise of the data. Robust clustering, which is a potential technique to address these problems, will be discussed in the next section.

6. ROBUST CLUSTERING FOR USER STEREOTYPING

6.1 Robust Clustering

Robust clustering is based on compiling the results of different clustering methods and on reporting only the elements that are co-clustered together by all the different algorithms. For two elements, all clustering methods must have allocated them to the same cluster in order for them to be assigned to a robust cluster. This gives a higher level of confidence to the correct assignment of elements appearing within the same cluster. Robust clustering is a very valuable technique for a rapid drilling-down of datasets into clusters whose pattern is identified in a manner independent of the cluster method, thus eliminating the bias of each technique.

Input : Agreement Matrix $A(n \ x \ n)$, C number of clustering techniques
Output: $RC = \{RC_1, ..., RC_m\}$ Set of robust clusters

$AgreementList$=Pairs (x, y) of A, with $A(x, y) = C$
$RC = \{\}$
$RC_1 = AgreementList_1$
For $i = 1$ **to** $Size(AgreementList)$
    $Found = False$
    For $j = 1$ **to** $Size(RC)$
        If $(AgreementList_{i,1}$ **or** $AgreementList_{i,2}) \in RC_j$
          $Found = True$
          If $AgreementList_{i,1} \notin RC_j$
              $RC_j = RC_j \cup AgreementList_{i,1}$
          end_if
          If $AgreementList_{i,2} \notin RC_j$
              $RC_j = RC_j \cup AgreementList_{i,2}$
          end_if
          $j = Size(RC)$
        end_if
    end_for
    if *NOT found*
        $RC = RC \cup \{AgreementList_i\}$
    end_if
end_for

Fig. 6. Robust Clustering algorithm.

Robust clustering (Swift et al. 2004) is based on an agreement matrix. The agreement matrix, of size n x n, with n the number of elements to be clustered, is an upper triangular matrix that indicates for each combination of elements the

number of agreements among the methods for clustering together the two elements, represented by the row and the column indices. Robust clustering uses the agreement matrix to generate an agreement list that contains all the pairs of elements of the matrix where the value is equal to the number of clustering methods used, *C*. Then, starting with an empty set of robust clusters, the first element is created containing the elements of the first pair of the list of agreement. Then, the algorithm iterates for the rest of the elements of the agreement list, where, if one element of the current pair is found in a robust cluster and the other is not, that element is added to the robust cluster, otherwise a new robust cluster is created. After the algorithm has iterated for each pair of elements in the agreement list, it outputs the set of robust clusters found. Figure 6 presents a description of the algorithm that implements robust clustering.

In general, the set of robust clusters obtained will not contain all the original elements, because those elements for which the clustering techniques used do not agree will not be included in the final set of robust clusters. This filtering property of robust clustering is very useful in our case because it will eliminate users that cannot be robustly grouped with other users. This helps to remove users who would introduce fuzziness in the definition of the behavior of a cluster.

6.2 User Stereotyping

We used the results obtained from using k-means, hierarchical and fuzzy clustering to apply robust clustering. Since we have 48 users in the dataset and used results from three different clustering algorithms, the Agreement Matrix was of dimension *48*x*48*, with *C=3*. After applying the algorithm, we obtained eight clusters and a total of 11 users were filtered by robust clustering. As for Clusters 2, 5 and 8, they only group two users, which is actually just one element of the Agreement List, so we considered that they did not represent relevant behavior. Table 11 presents the center of the remaining clusters where each dimension of the cluster center has been obtained as the mean of all the users included in that cluster. From here we can derive the behavior of each one of these clusters:

- Cluster 1: Users that exclusively use "Basic Search" in combination with "Word or Phrase";
- Cluster 3: Users that use "Basic Search" and "Advanced Search", in the almost same proportion;
- Cluster 4: Users that use "Basic Search" four times as much as they use "Advanced Search" and "Word or Phrase";
- Cluster 6: Users that exclusively use "Basic Search" and "Word or Phrase" three times as often as "Author/Title".
- Cluster 7: Users that exclusively use "Basic Search" in combination with "Author/Title".

Table 11. Cluster centers when using Robust Clustering.

	Users	BS	AS	SE	ATS	NS	GB
Cluster 1	7	1.74	0	0.8	0.5	0.08	0
Cluster 3	7	0.65	0.54	0.31	0.17	0.08	0
Cluster 4	7	1.12	0.24	0.3	0.7	0.2	0.06
Cluster 6	5	1.57	0.02	1.19	0.35	0.07	0.19
Cluster 7	5	1.53	0.01	0.02	1.46	1.1	1.19

With regard to identifying relevant human factors in determining the clusters, level of expertise had a significance of *p=0.656* and gender of *p=0.317*. However, we found the significance of cognitive style with the clusters, having a value of p*=0.000*. 50% of Field Independent users are in Cluster 1 and 80% of Field Dependent users are included in Cluster 7. These results appear consistent with the WA Score of each cluster. Cluster 1 has the highest WA score (Mean=1.79; SD=0.19) while Cluster 7 has the lowest WA score (Mean=0.88; SD=0.06). In other words, the users of Cluster 1 have a Field Independent tendency whereas the users of Cluster 7 have a Field Dependent tendency. The main difference between Cluster 1 and Cluster 7 is that the users of the former prefer to use "Word or Phrase" and those of the latter favor to use "Author/Title". Field Independent users tend to actively explore the systems by themselves (Liu and Reed 1994) and jump from one point to another in hyperspace (Chen 2002), which explains that they tend to use "Word or Phrase", providing information without a predefined structure. Conversely, Field Dependent users tend to take a passive approach by relying on salient cues (Anastasi, 1988) and have more difficulties in separating the individual parts within a whole (Witkin et al. 1977). The "Author/Title" gives additional guidance to help Field Dependent users to search for information. Another interesting finding is that Intermediate users are almost evenly distributed in Cluster 3 (25%), Cluster 4 (25%), and Cluster 6 (20%), suggesting they are equally comfortable using each information seeking strategy. One possible interpretation is that individuals possessing an Intermediate cognitive style combine the characteristics of both Field Independence and Field Dependence and employ a more versatile repertoire of information seeking strategies. Versatile users, who have

acquired the skill to move back and forth between different information seeking strategies, are more capable of adapting themselves to suit the system (Ford, 2000). These results suggest that cognitive styles play a key role in determining the user behavior.

However, these results are only for users with strong definitions of cognitive styles because we found that 90% of the users that were filtered by robust clustering were within a 10% margin of the cognitive styles borders. In other words, they are included within the ranges (0.93-1.13) which define the Field Dependent-Intermediate border and (1.26-1.46) which define the Intermediate-Field Independent border. As noted in our previous discussion, one of the benefits of using robust clustering is to filter noisy data. A user whose WA score is near the border of a cognitive style may have mixed properties and may not show clear behavior related to his/her cognitive style. Such users add noise to the behavior characteristics so it is better to filter them when we study user behavior in digital libraries or other web-based applications.

Table 12: CSUQ questions used for clustering

No.	Question
2	It was simple to use this system: 1 (Strongly disagree) – 7 (Strongly agree)
3	I can effectively complete my work using this system: 1 (Strongly disagree) – 7 (Strongly agree)
6	I feel comfortable using this system: 1 (Strongly disagree) – 7 (Strongly agree)
11	The information provided with this system is clear: 1 (Strongly disagree) – 7 (Strongly agree)
13	The information provided for the system is easy to understand: 1 (Strongly disagree) – 7 (Strongly agree)
16	The interface of this system is pleasant: 1 (Strongly disagree) – 7 (Strongly agree)
18	This system has all the functions I expect it to have: 1 (Strongly disagree) – 7 (Strongly agree)
19	Overall, I am satisfied with this system: : 1 (Strongly disagree) – 7 (Strongly agree)

Table 13. QUIS questions used for clustering.

No.	Question
1	The interface is: terrible (0) – wonderful (9)
2	The interface is: Difficult (0) – Easy (9)
4	The interface has: Inadequate Power (0) – Adequate Power (9)
6	The system is: Rigid (0) – Flexible (9)
9	Organization of Information: Confusing(0) – Very Clear (9)
10	Sequence of Screens: Confusing (0) – Very Clear (9)
17	Learning to operate the system is: Difficult (0) – Easy (9)
18	Exploring new features by trial an error is: Difficult (0) – Easy (9)
27	The system is designed for all level of users: Never (0) – Always (9)

7. HUMAN FACTORS AND USER PERCEPTION

In addition to user behavior, we also examined the relationships between human factors and user perception. A similar approach to that used in the previous section was applied for this case. The data obtained from the satisfaction questionnaires were analyzed by using clustering techniques and the significances of the corresponding human factors for each cluster were also examined. The questionnaires provided for each user a vector containing the answer to the corresponding question in each dimension, making a total of 49 dimensions. In order to study the significance of each dimension, and due to the lack of statistical properties of clustering methods, typically, ANalysis Of VAriance (ANOVA) models are implemented as a precursor to clustering. This approach has been typically used in bioinformatics (Wolfinger et al. 2001; Park et al. 2003; Liu et al. 2005) to identify the genes that are statistically more meaningful. This pre-processing of data usually drastically reduces the dimension of the original data and improves performance of the clustering method (Liu et al. 2005). In our case, we worked with the 49 dimensions describing the perception of a user, in order to identify which questions were more significant, and these questions were then used for further clustering. From the original 49 dimensions, 27 were identified to be significant with $p<0.0 1$ (Liu et al. 2005). Analyzing the semantics of each question, in order to avoid similar questions and/or concepts, we reduced the set to 17 questions. These questions have been presented in Table 12, which shows the CSUQ questions that have been used, and Table 13, which shows the QUIS questions used. The three ASQ questions were filtered by ANOVA. In the end, each user's perceptional information is represented by a 17-dimensional vector containing integer numbers in the range (0-9) or (1-7), representing the answer to the corresponding question.

7.1 K-means, Hierarchical Clustering and Fuzzy Clustering for Identification of Perceptions

We started by applying k-means to the perception data. In order to determine the optimal number of clusters, we ran the algorithm for $k=2,...,9$ without giving any initial value for the cluster centers and using Euclidean distance. To avoid the solution given for a given k being a local minima, because of the randomness of the original centers, k-

means was run for each value of k 100 times and the solution used was the one that minimized the objective function J presented in (1). In order to determine how good the partition obtained was, for each user i we obtained ϕ_i using Equation (11). The quality of a partition, q_k, $(k=2,...,9)$, was obtained as the mean value of all the ϕ_i values of the system. Considering this approach, the optimum number of clusters was two, as shown in Figure 7(a). Figure 7(b) presents the optimum partition and the value of ϕ_i for each user included in that partition. Subsequently, we obtained the significance of cognitive styles, levels of expertise and gender differences for these partitions, with values $p=0.890$, $p=0.321$ and $p=0.114$ respectively. Levels of expertise and gender differences may play a more important role than cognitive style in determining perception clusters, but the relevance is not high.
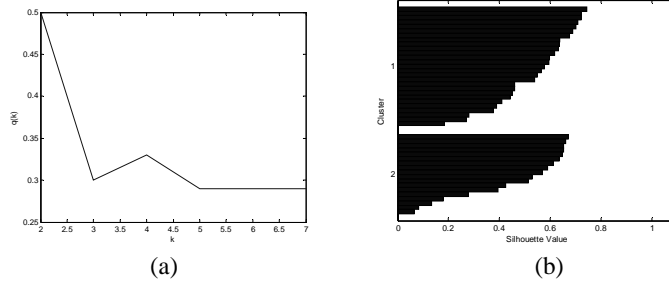

(a)   (b)

Fig. 7. (a) Evolution of the quality of the clusters for different values of k, and (b) representation of the optimum two cluster partition found, using perceptional data.
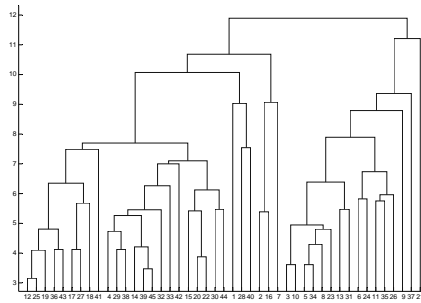

Fig. 8. Schematic representation of the hierarchical tree constructed using perceptional data.

Furthermore, hierarchical agglomerative clustering was used to analyze the same data and a Euclidean distance was applied to construct clusters. Figure 8 presents the hierarchical clustering tree obtained, where the X-axis presents the users (some of them have been grouped to present a clear representation) and the Y axis the distance between the two elements being connected. The hierarchical tree was cut at a height of 11.5, which creates two different clusters. This was motivated by: (1) graphically from Figure 6, it can be inferred that the system identifies two different groups and (2) k-means, for the same dataset, has established that the optimum number of clusters is two. Based on this result, the significance of cognitive styles, levels of expertise and gender differences were $p=0.897$, $p=0.217$ and $p=0.169$ respectively. Again, it is clear that cognitive styles do not play any role in determining the perception of a user, and there is a slight indication that levels of expertise and gender differences may be relevant. The two clusters obtained when applying Fuzzy Clustering produce very similar results, with significance values of $p=0.719$, $p=0.310$ and $p=0.451$ for cognitive styles, levels of expertise and gender differences.

Table 14. Kappa values for each technique comparison when using perceptional data.

	K-means Clustering	Hierarchical Clustering	Fuzzy Clustering
K-means Clustering	1	0.906	0.509
Hierarchical Clustering	-	1	0.717
Fuzzy Clustering	-	-	1

The previous results clearly showed that there is no relationship between the cognitive style of a user and his/her perception, while there is a possible correlation between the perception and gender and level of expertise of that user. As indicated in Section 6.2, robust clustering is useful to filter 'noisy' users and to reduce the bias of each technique, so we re-examined the relevance of each human factor determining the perception of a user by applying robust clustering. The *kappa* values, as presented in Table 14, indicate that there is a reasonably good agreement between all

18

the techniques, which implies that the partitions created are consistent (Altman 1997).

Table 15. Cluster centers obtained by Robust Clustering when using perceptional data.

	QUIS1	QUIS2	QUIS4	QUIS6	QUIS9	QUIS10	QUIS17	QUIS18	QUIS27
Cluster 2	6.29	7.25	6.25	6.18	6.55	5.59	7.25	7.00	6.44
Cluster 3	2.94	5.22	3.05	2.5	3.27	4.33	5.5	4.94	4.61

	CSUQ2	CSUQ3	CSUQ6	CSUQ11	CSUQ13	CSUQ16	CSUQ18	CSUQ19
Cluster 2	5.40	5.29	5.44	4.51	5.37	4.81	4.4	4.96
Cluster 3	4.00	3.77	3.83	2.88	3.66	2.44	2.88	3.22

7.2 Robust Clustering for Identification of Perceptions

We used the results obtained from using k-means, hierarchical and fuzzy clustering with perception data to apply robust clustering. The Agreement Matrix was of dimension *48*x*48*, with *C=3*. After applying the algorithm, we obtained five clusters, with only three users filtered. Clusters 1, 4 and 5 were not considered because they only grouped two users. Table 15 presents the mean value of each dimension that characterizes the perception of a user for Cluster 2 and Cluster 3. The first table presents the values for the QUIS questionnaire (which has a range of (0-9)) and the second for the CSUQ questionnaire (which has a range of (1-7)). According to the semantics of the questions, we can obtain the following perceptions:

- Users of Cluster 2 think that the system is acceptable (6.29 in QUIS1) easy to use (7.25 in QUIS2), that offers adequate power (6.25 QUIS4) and is flexible enough (6.18 QUIS6). In general these users are moderately satisfied with the BLC, and think that they have the tools they need to efficiently accomplish a task.
- Users of Cluster 3 think the system is not up to what they would expect (2.94 in QUIS1), have a neutral opinion regarding if it is difficult or easy (5.22 QUIS2) to use, think the BLC offers inadequate power (3.05 in QUIS4) and that the system is rigid (2.5 in QUIS6). In general, the users of this cluster are dissatisfied with all the characteristics of the system, and would expect the system to improve.

To identify human factors in determining the robust clusters, we obtained the significance *p=0.730* for cognitive styles and *p=0.009* for levels of expertise and *p=0.239* for gender differences. This shows that the level of expertise is the main factor that determines the perception of a user, while at the same time confirming that cognitive style does not play any role and gender differences may play a minor role. By studying the level of expertise of each cluster, we found:

- Cluster 2 groups users with a Medium expertise level. In this cluster 70% of users are of this type, representing 66% percent of all the users of that expertise level.
- Cluster 3 groups users with an Advanced expertise level. In this cluster 43% of users are of this level, grouping 70% of all the expert users of the pool.

These results suggest that expert users are more dissatisfied with the BLC than users with medium expertise. One of the main areas of dissatisfaction is the flexibility of the system (QUIS6), which expert users consider to be very poor. This perception, to some extent counterintuitive, can be explained by the fact that the expert users know the BLC better and are probably more aware of functionalities provided by other library catalogues. The aforementioned results are only for users with medium or advanced levels of expertise. In other words, the perception of non-experienced users or novice users cannot be formed as clusters. This suggests that users do not have a well defined perception until they reach some degree of expertise (e.g. medium or advanced levels). These results highlight that there is a close relationship between the perception of a user and the levels of expertise of that user when using digital libraries.

8. DESIGN OF A PERSONALIZED INTERFACE FOR BLC

The goal of this section is to prove the effectiveness and feasibility of implementing a personalized interface for BLC using the relevant human factors studied in the previous sections, which have highlighted the connection between cognitive style/behavior and level of expertise/perception. Although there are relationships between level of expertise and user perception, novice users and users who do not use the system do not have a well defined perception. Thus, level of expertise may not be a suitable human factor for building the personalized interface. On the other hand, the results presented in Section 6 show that different cognitive style groups have different preferences to look for information when interacting with BLC. These preferences can be used as guidelines for designing a personalized interface for BLC:

- Field Dependent users only need the basic search interface in combination with searching by author/title. Other searching functionalities, including the advance search interface, are not needed as part of the interface. This simplified interface that uses basic search with author/title will help Field Dependent users identify salient cues and avoid the "feeling lost" problem (Wang et al. 2000).
- Intermediate users are almost evenly distributed in the clusters. In other words, they do not have a clear preference for a particular interface feature. It suggests that the combination of various interface features, i.e. the original interface, is a suitable personalized interface for them. Therefore, there is no need to provide another interface.
- Field Independent users only need the basic search interface in combination with word or phrase. Other searching functionalities, including the advance search interface, are not needed as part of the interface. The simplified interface that uses basic search with word or phase will provide Field Independent users with suitable navigation support to explore the system by themselves (Liu and Reed 1994).

Figures 9, 10 and 11 present the three personalized interfaces designed for each cognitive style. Each interface was automatically generated according to each cognitive style using Privoxy (Privoxy, 2006), a programmable proxy server that filtered the elements of the interface that were irrelevant for each cognitive style. Figure 12 presents the basic connection between a user and BLC through the proxy server used to adapt the interface
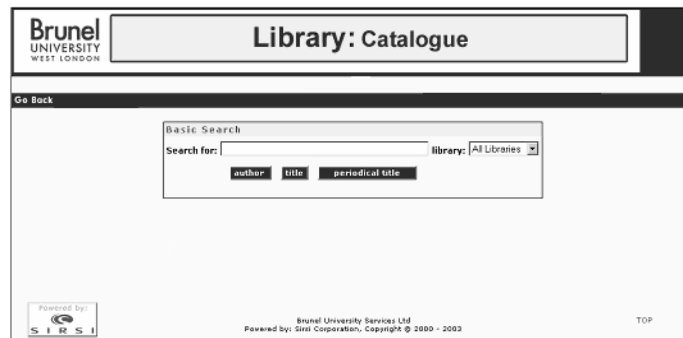

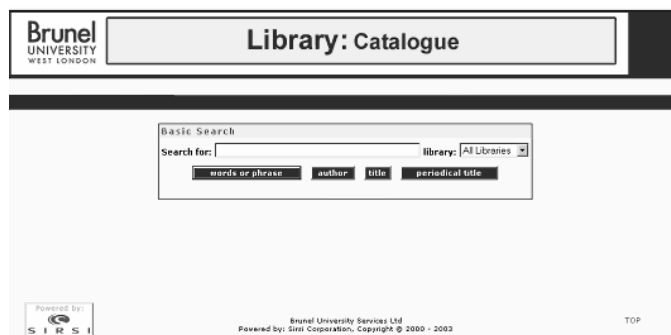Fig. 9. Personalized interface for Field Dependent users.

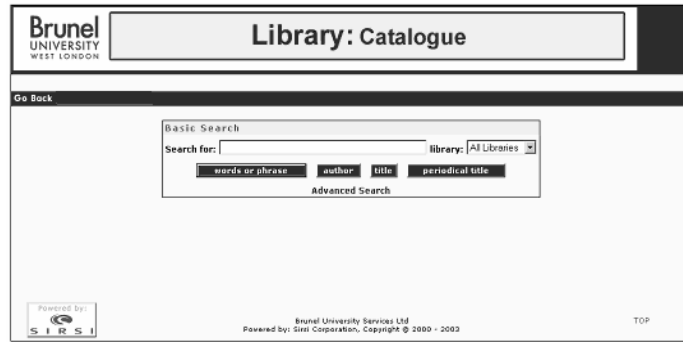Fig. 10. Personalized interface for Field Independent users.

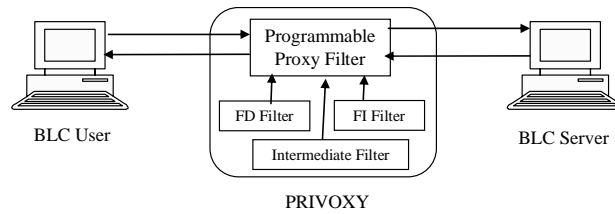Fig. 11. Personalized interface for Intermediate users.

Fig. 12. Architecture for the implementation of adaptive interfaces.

In order to measure the efficiency and effectiveness of our results for personalization, the same experiment presented in section 4 was conducted but using the personalized interfaces designed for BLC. The experiment includes 50 participants, who are different from the previous experiment. However, they were also students from Brunel University and volunteered to participate in the experiment. Riding's CSA was used to classify their cognitive styles as Field Independent, Intermediate and Field Dependent. In the experiment, each participant interacted with the personalized interface matching with his/her cognitive style and had to complete the tasks presented in Table 5. Finally, the participants needed to fill out the CSUQ and QUIS questionnaires to express their satisfaction with using the personalized interface.

To identify whether the personalized interface can increase user satisfaction, the participants' responses to the CSUQ and QUIS questionnaires in this experiment were compared with those in the previous experiment which used the original BLC interface. Instead of using all questions from these two questionnaires, we considered the semantic relevance to select relevant questions for this comparison. Table 16 and Table 17 show the means and the standard deviations (std) of users' responses to the selected QUIS and CSUQ questions respectively for the original BLC interface while Table 18 and Table 19 present the same values but considering the personalized interfaces. As showed in these tables, there is an increase in user satisfaction when going from the original BLC interface to the personalized interfaces. This increase is mainly stated by: (1) CSUQ 19, which indicates how satisfied the users are with the system, has gone from 4.33 to 4.98, a 9.2% increase in satisfaction over the scale of CSUQ and by (2) QUIS 1, which measures if the interface is terrible or wonderful, has gone from 5.23 to 5.84, a 6.5% increase in the scale of QUIS. In general, all questions have increased in the area of 5% to 10%.

Table 16. Global Mean and Standard deviation of selected QUIS questions considering the original interface.

	QUIS 1	QUIS 2	QUIS 4	QUIS 6	QUIS17	QUIS 18	QUIS 27
Mean	5.23	6.63	5.13	4.87	6.43	5.67	5.27
Std.	2.300	1.903	2.417	2.300	2.161	2.591	2.518

Table 17 Global Mean and Standard deviation for the selected CSUQ questions considering the original interface.

	CSUQ1	CSUQ3	CSUQ6	CSUQ7	CSUQ16	CSUQ17	CSUQ18	CSUQ19
Mean	4.74	4.78	4.93	5.33	3.96	3.67	3.98	4.33
Std.	1.443	1.525	1.439	1.274	1.636	1.554	1.548	1.467

Table 18. Users' Responses for QUIS questions considering the personalized interface.

	QUIS1	QUIS2	QUIS4	QUIS6	QUIS17	QUIS18	QUIS27
Mean	5.84	6.58	5.58	5.36	6.91	6.05	5.67
Std	1.511	1.562	1.829	2.058	1.616	2.011	2.032

Table 19. Users' Responses for CSUQ questions considering the personalized interface.

	CSUQ1	CSUQ3	CSUQ6	CSUQ7	CSUQ16	CSUQ17	CSUQ18	CSUQ19
Mean	5.44	5.07	5.60	5.60	4.49	4.42	4.51	4.98
Std.	1.385	1.549	1.158	1.294	1.454	1.607	1.454	1.504

After assessing the global impact of introducing personalization, it is necessary to examine how each cognitive style reacts to personalized interfaces. Table 20 and Table 21 present the mean values for the original interface for each cognitive style while Table 22 and Table 23 present the same mean values but for the personalized interface. It seems that all cognitive styles benefit from the personalized interface. In particular, Field Independent users enjoy more satisfaction than those with other cognitive styles.

Table 20. Different Cognitive Styles' Responses for QUIS questions considering the original interface.

FD/FI		QUIS 1	QUIS 2	QUIS 4	QUIS 6	QUIS 17	QUIS 18	QUIS 27
Field Independent	Mean	5.00	6.63	4.94	4.75	6.31	5.50	5.25
	Std. Deviation	2.852	2.125	2.620	2.745	2.549	2.582	3.044
Intermediate	Mean	5.56	7.00	5.89	5.22	6.22	5.22	5.44
	Std. Deviation	.882	1.323	1.764	1.641	1.394	2.728	1.667
Field Dependent	Mean	5.40	6.00	4.40	4.60	7.20	7.00	5.00
	Std. Deviation	2.408	2.236	2.881	2.074	2.168	2.449	2.345

Table 21. Different Cognitive Styles' Responses for CSUQ questions considering the original interface.

FD/FI		CSUQ 1	CSUQ 3	CSUQ 6	CSUQ 7	CSUQ 16	CSUQ 17	CSUQ 18	CSUQ 19
Field Independent	Mean	4.76	4.67	4.76	4.95	3.95	3.57	3.52	4.00
	Std	1.480	1.683	1.411	1.203	1.596	1.399	1.504	1.612
Intermediate	Mean	4.71	4.75	4.96	5.62	3.96	3.54	4.38	4.54
	Std	1.488	1.567	1.488	1.209	1.546	1.641	1.583	1.318
Field Dependent	Mean	4.78	5.11	5.22	5.44	4.00	4.22	4.00	4.56
	Std	1.394	1.054	1.481	1.509	2.121	1.716	1.414	1.509

Table 22. Different Cognitive Styles' Responses for QUIS questions considering the personalize interface.

FD/FI Dimension		QUIS1	QUIS2	QUIS4	QUIS6	QUIS17	QUIS18	QUIS27
Field Dependent	Mean	5.83	6.58	5.42	4.17	7.08	6.58	6.42
	Std	1.267	1.730	1.782	2.082	1.782	2.314	2.193
Intermediate	Mean	5.25	7.83	5.58	4.83	6.58	5.42	5.83
	Std	1.712	1.642	1.832	2.517	1.782	1.975	2.125
Field Independent	Mean	6.21	7.05	5.68	6.00	7.00	6.11	5.74
	Std	1.475	1.268	1.945	1.374	1.453	1.823	1.759

Table 23. Different Cognitive Styles' Responses for CSUQ questions considering the personalized interface.

FD/FI Dimension		CSUQ1	CSUQ3	CSUQ6	CSUQ7	CSUQ16	CSUQ17	CSUQ18	CSUQ19
Field Dependent	Mean	5.75	5.17	5.83	6.08	4.42	4.67	4.42	5.25
	Std	1.288	1.832	1.115	1.084	1.564	1.670	1.730	1.215
Intermediate	Mean	5.08	4.83	5.17	4.92	4.42	4.17	5.00	4.58
	Std	1.832	1.642	1.586	1.564	1.240	1.403	.603	1.676
Field Independent	Mean	5.47	5.32	5.74	5.74	4.58	4.42	4.26	5.05
	Std	1.124	1.336	.806	1.098	1.575	1.742	1.628	1.580

There is another interesting improvement of the personalized interfaces over the original BLC interface, which is related to the standard deviations. In the original interface, the average standard deviation for the QUIS and CSUQ questionnaires was 2.203 and 1.58 respectively, while in the personalized interface it has gone down to 1.8 and 1.4 respectively. These values imply that in the personalized interfaces users are not only more satisfied but also agree more on their perception. The effectiveness of the personalized interfaces is also showed in the reduction of the average time needed to complete a task. While the average amount of time needed to complete a task with original BLC interface was 63656 seconds, these values were reduced to 51436 seconds with the personalized interfaces. These results demonstrate the effectiveness of considering cognitive styles as a relevant human factor to build the personalized interfaces.

9. CONCLUSIONS

There is a lack of research to investigate the role of each human factor in determining the behavior and perception of web users. The main goal of our paper was to identify which human factors played the most relevant roles and to study their importance. In order to do so, we focused on a digital library environment due to the importance and relevance of such environments. We used the Brunel Library Catalogue (BLC) as an example of a digital library to collect user behavioral and perception data, which were analyzed using various clustering techniques. We started by studying behavioral data using k-means, hierarchical clustering and fuzzy clustering. These approaches revealed that although some human factors, mainly cognitive styles, had some significance in grouping users, the role was minor and not very relevant. Nevertheless, these approaches suffer from some problems, mainly because of the bias of each technique and the lack of any kind of filtering of noisy data. In order to avoid these problems, we proposed the use of robust clustering, which constructs clusters based on the agreement of a set of clustering techniques. Our results showed that the filtering property of robust clustering is very useful for user modeling because it can filter those that do not have a well-defined behavior, especially those who are near the borders of cognitive styles.

According the results from robust clustering, there are direct relationships between clusters and cognitive styles. This reinforces the idea that cognitive style plays a very relevant role in determining the behavior of a user. From a personalization perspective, this result also implies that one of the most important elements that a stereotype should have is the cognitive style of the user. However, this can be a complex and time consuming task, which usually involves the execution of a cognitive style test by each one of the users that is going to interact with that system. Another possible approach is to automatically identify the cognitive style of a user using previous interactions of the users with the system. In the other work (Frias-Martinez et. al., 2007), we detail how to automatically identify the cognitive style of a user using the previous transactions of that user with BLC based on a neuro-fuzzy classification system. Such an automatic approach of identifying users' cognitive styles makes it possible to consider the inclusion of cognitive styles in user models.

We used a similar approach to analyze user perception. The results, after using robust clustering, suggested a strong significance of the levels of expertise in determining the different perceptual clusters. The results indicated that expert users had more negative perceptions than medium users, probably because expert users are more aware of other search functionalities not presented in the BLC. The results also showed that novice users do not have a homogeneous perception of the system, which may explain why they do not form a cluster. The results are a strong indication that the level of expertise is the main factor that affects the perception of a user.

In summary, three main conclusions can be drawn from this study. The first is that robust clustering is an effective technique for user modeling. The advantages of robust clustering include the elimination of bias and filtering of noisy elements. Nevertheless, this study only selected k-means, hierarchical clustering and fuzzy clustering to build robust clusters. Given any dataset, there are often no strict rules that impose the use of a specific method over another in its analysis. Therefore, it will be necessary to conduct further work to build robust clusters using other clustering

techniques, and interesting to see whether similar results will be found by using these techniques. The second conclusion is that cognitive style is an influential factor in determining user behavior. The third is that level of expertise is an essential issue affecting user perception. However, it was only a small-scale study. Further studies need to be undertaken with a larger sample to provide additional evidence. There is also a need to conduct further research to investigate if there is any kind of relationship between user behavior stereotypes and user perception stereotypes. The findings of such studies could be integrated to build robust user models for personalization.

In terms of external validity, the experiments were only run for BLC, but the three main conclusions detailed in the previous paragraph should be valid for any library catalogues in general because of the standard sets of functionalities that these systems offer to users. For the same reasons, the design guidelines given for personalizing BLC should apply to any library catalogues. We also think that, to a large extent, the three main conclusions detailed in the previous paragraph are likely to be valid for search engines, too, although the personalization guidelines will not apply to these systems. In addition, the third conclusion regarding robust clustering may be applied for user modeling of web-based applications, and the first two conclusions could well be valid for this type of application but further experimental evidence need to be collected.

ACKNOWLEDGMENTS

REFERENCES

Altman, D.G. 1997. *Practical Statistics for medical Research*. Chapman and Hall, London.

Anastasi, A. 1988. *Psychological testing.* New York: Macmillan.

Bezdek, J.C. 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York.

Bryan-Kinns, N., Blandford, A., Thimbleby, H. 2000. Interaction Modelling for Digital Libraries. In *Workshop on Evaluation of Information Management Systems*.

Callan, J., Smeaton, A., Beaulieu, M., Borlund, P., Brusilovsky, P., Chalmers, M., Lynch, C., Riedl, J., Smyth, B., Straccia, U., Toms E. 2003. Personalization and Recommender Systems in *Digital Libraries, Joint NSF-EU DELOS Working Group Report* , URL http://www.ercim.org/publication/ws-proceedings/Delos-NSF/Personalisation.pdf.

Chen, S. Y. 2002. A Cognitive model for non-linear learning in hypermedia programmes. *British Journal of Educational Technology*, *33*, 453-464

Chen, S. Y., Macredie, R. D. 2004. Cognitive Modelling of Student Learning in Web-based Instructional Programmes. *International Journal of Human-Computer Interaction.* 17, 375-402.

Chen, S.Y., Macredie, R. 2002. Cognitive Styles and Hypermedia Navigation: Development of a Learning Model. *Journal of the American Society for Information Science and Technology*. 53, 3-15.

Chin, J.P., Diehl, V.A., Normal, K.L. 1988. *Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface*. ACM CHI'88 Proceedings, 213-218.

Chiu, S. L. 1994. Fuzzy Model Identification based on Cluster Estimation. *Journal of Intelligent & Fuzzy Systems. 2,* 267-278.

Cho, Y. H., Kim, J. K., Kim, S. H. 2002. A personalized recommender system based on web usage mining and decision tree. *Expert Systems with Applications*. 23, 329–342

Doux, A., Laurent, J., Nadal, J. 1997. Symbolic Data Analysis with the k-Means Algorithm for User Profiling, User Modeling. *Proceedings of the sixth International Conference*, UM97. pp. 359-361.

Downing, R. E., Moore, J., Brown, S. 2005. The effects and interaction of spatial visualization and domain expertise on information seeking. Computers in Human Behavior. 21, 195-209

Everitt, B.S. 1993. *Cluster Analysis*, Arnold. 3rd edition.

Ford, N. 2000. Cognitive styles and virtual environments. *Journal of the American Society for Information Science. 51*, 543-557

Ford, N., Miller, D. 1996. Gender differences in Internet perception and use. In *Electronic Library and Visual Information Research, Proceedings of the third ELVIRA Conference*, 87-202.

Ford, N., Miller, D., Moss, N. 2005. Web Search Strategies and Human Individual Differences. Cognitive and Demographic Factors, Internet Attitudes and Approaches. *Journal of the American Society for Information Science and Technology*. 56, 741-756.

Frias-Martinez, E., Chen, S., Liu, X. 2007. Automatic Cognitive Style Identification of Digital Library Users for Personalization. *Journal of the American Society for Information Science and Technology*. 58(2), 237-251

Fu, Y., Sandhu, K., Shih, M.Y. 2002. Clustering of Web users based on access patterns. In *Proceedings of the 1999 KDD Workshop on Web Mining*, July 2002.

Goren-Bar, D., Graziola, I., Pianesi, F., Zancanaro, M. 2006. The influence of personality factors on visitor attitudes towards adaptivity dimensions for mobile museum guides, *User Modeling and User-Adapted Interaction*, 16, 31 - 62

Hay, B., Wets, G And Vanhoof, K. 2001. Clustering navigation patterns on a website using a Sequence Alignment Method", in *IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization*, 2001.

Hong, J., Heer, J., Waterson, S., Landay, J.A. 2001. WebQuilt: A Proxy-based Approach to Remote Web Usability Testing. *ACM Transactions on Information Systems.* 19, 263-385.

Ivory, M. Y., Megraw, R. 2005. Evolution of web site design patterns. *ACM Transactions on Information System*. 23, 463-497.

Jain, A., Dubes, R.C. 1999. Data clustering. *ACM Computing Survey*, 31, 264-323.

Joshi, A. K.. Krishnapuram, R. 2000. On mining Web Access Logs. In *Proceedings of the ACM-SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 63-69.

Kobsa, A. 2001. Generic User Modeling Systems. *User Modeling and User-Adapted Interaction.* 11, 49-63.

Krishnapuram, R., Joshi, A., Nasraoui, O., Yi, L. 2001. Low-Complexity Fuzzy Relational Clustering Algorithms for Web Mining. *IEEE Transactions on Fuzzy Systems*. *9*, 595-608.

Lampinen, T., Koivisto, H. 2002. Profiling Network Applications with Fuzzy C-Means Clustering and Self-Organising Map. In *International Conference on Fuzzy Systems and Knowledge Discovery*, November 2002.

Large, A., Beheshti, J., Rahman, T. 2002. Design criteria for children's Web portals: The users speak out. *Journal of the American Society for Information Science and Technology. 53*, 79-94.

Lazander, A.W., Biemans, H.J., Wopereis, I.G. 2000. Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science. 51*, 76-581.

Lewis, J.R. 1995. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for User. *International Journal of Human-Computer Interaction*. *7*, 57-78.

Ling, J., Van Schaik, P. 2006. The influence of font type and line length on visual search and information retrieval in web pages. International Journal of Human-Computer Studies. 64, 395-404

Liu, H., Tarima, S., Borders, A.S., Getchell, T.V., Getchell, M.L., Stromberg, A.J. 2005. Quadratic regression for gene discovery and pattern recognition for non cyclic short time-course microarray experiments. *BMC Bioinformatics 6,* 106

Liu, M., Reed, W.M. 1994. The effect of hypermedia assisted instruction on second-language learning through a semantic-network-based approach. *Journal of Educational Computing Research. 12*, 159-175.

Macqueen, J.B. 1967. Some Methods for classification and Analysis of Multivariate Observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 281-297.

Marchionini, G., Plaisant, C., Komlodi, A. 2003. The people in digital libraries: multifaceted approaches to assessing needs and impact. In: Bishop, A., van House, N. A., Buttenfield, B.P. (Eds.) Digital Library Use Social Practice in Design and Evaluation. MIT Press, 119-160.

Mitchell, T. J. F., Chen, S. Y., Macredie, R. D. 2005. Hypermedia Learning and Prior Knowledge: Domain Expertise vs. System Expertise. *Journal of Computer Assisted Learning*. 21, 53-64.

Mobasher, B., Dai, H. Luo, T., Nakagawa, M. 2001. Effective personalization based on Association Rule discovery from web usage data. In *3rd. ACM Workshop on Web Information and Data Management*.

Nasraoui, O., Frigui, H. Joshi, A., Krishnapuram, R. 1999. Mining Web Access Logs Using Relational Competitive Fuzzy Clustering. In *Eight International Fuzzy Systems Association World Congress - IFSA 99*.

Paliouras, G., Karkaletsis, G. V., Papathedorou. C., Spyropoulos, C. 1999. Exploiting Learning Techniques for the Acquisition of User Stereotypes and Communities. In *Proceedings of the International Conference on User Modelling (UM '99)*.

Palmquist, R. A., Kim, K.-S. 2000. Cognitive style and on-line database search experience as predictors of Web search performance. *Journal of the American Society for Information Science*. 51, 558-66.

Park, T.I., Yi, S.G., Lee, S., Lee, S.Y., Yoo, D.H., Ahn, J.I., Lee, Y.S. 2003. Statistical tests for identifying differentially expressed genes in time-course microarray experiments. *Bioinformatics. 19*, 694-703.

Riding R. J., Grimley, M. 1999. Cognitive style, gender, and learning from multimedia materials in 11 year-old children. *British Journal of Education Technology*, 30(1), 43-56.

Riding, R. J. 1991. *Cognitive Styles Analysis*. Learning and Training Technology, Birmingham.

Riding, R., Rayner, S. G. 1998. *Cognitive Styles and Learning Strategies*. David Fulton Publisher, London.

Roy M., Chi M. T. C. 2003. Gender differences in patterns of searching the web. *Journal of Educational Computing Research*. *29*, 335-348.

Sander, J., Ng, R. T., Sleumer, M. C., Yuen, M. S., Jones, S. J. 2005. A methodology for analyzing SAGE libraries for cancer profiling. *ACM Transactions on Information Systems*. *23*, 35-60.

Savaresi, S.M., Gazzaniga, G., Boley, D.L., Bittani, S. 2002. Cluster Selection in Divisive Clustering algorithms. In Second SIAM International Conference in Data Mining, Arlington VA, 85-101

Stephen, P., Hornby, S. (1997) Simple Statistics for Library and Information Professionals. London: Library Association.

Swift, S., Tucker, A., Vinciotti, V., Marin, N., Orengo, C., Liu, X., Kellam, P. 2004. Consensus Clustering and Functional interpretation of gene-expression data. *Genome Biology*. 5:R94, http://genomebiology.com/2004/5/11/R94.

Torkzadeh G.,Van Dyke, T.P. 2002. Effects of training on Internet self-efficacy and computer user attitudes. *Computers in Human Behavior, 18,* 479-494.

Tarpin-Bernard, F., Habieb-Mammar, H. 2005. Modeling Elementary Cognitive Abilities for Adaptive Hypermedia Presentation. User Modeling and User-Adapted Interaction. 15, 459 - 495

Uebersax, J.S. 1987. Diversity of decision-making models and the measurement of interrater agreement. *Psycological Bulletin*. 101, 140-146.

Valiquette, C. Lesage A., Cyr, M., Toupin J. 1994. Computing Cohen's kappa coefficients using SPSS MATRIX. *Behavioral Research Methods, Instruments and Computer*s. *26*, 60-61.

Vansickle, S. L. 2000. *Tenth graders' search knowledge and use of the World Wide Web*. Unpublished PhD Dissertation. Georgia State University.

Wang, P., Hawk, W.B., Tenopir, C. 2000. User's interaction with world wide web resources: an exploratory study using a holistic approach. *Information Processing and management* 36, 229-251.

Webb, A. 1999. *Statistical Pattern Recognition*. Arnold.

Weller, H. G., Repman, J., Rooze, G. E. 1994. The relationship of learning, behavior, and cognitive styles in hypermedia-based instruction: Implications for design of HBI. *Computers in the Schools, 10*, 401-420.

Witkin, H. A., Moore, C. A., Goodenough, D. R., Cox, P. 1977. Field-dependent and field independent cognitive styles and their educational implications. *Review of Educational Research*, *47*, 1-64.

Witten, I.H., Frank E. 1999. *Data Mining. Practical Machine Learning Tools and Techniques with JAVA Implementations*. Morgan Kaufman Publishers.

Wolfinger, R.D, Gibson, G., Wolfinger, E.D., Bennet, L. Hamadeh, H., Buschel, P., Afshari, C., Paules, R.S. 2001.Assesing Gene Significance from cDNA Microarray Expression Data via Mixed Models. *Journal of Computational Biology. 8*, 625-637.

Yi, M.Y. Hwang, Y. 2003. Predicting the use of web-based information systems, self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies,* 59, 431-449.

Zukerman, I., Albrecht, D.W., And Nicholson, A.E. 1999. Predicting Users Request on the WWW. In *Proceedings of the 7th International Conference on User Modeling*, UM99, Banff, Canada, 275-284.

Authors' Vitae

Enrique Frias-Martinez is a Research Fellow in the School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, U.K. He obtained a Ph.D. degree in Computer Science from the Universidad Politecnica de Madrid in Spain in 2000 and a Ph.D. degree in Information Systems from Brunel University, Uxbridge in UK in 2007. His current research interests include soft computing, data mining, machine learning, and human-computer interaction, areas in which he has published extensively. Dr. Frias-Martinez was the recipient of the Best Ph.D. Thesis Award of the School of Computer Science 2001, Universidad Politecnica de Madrid

Sherry Y. Chen is a Senior Lecturer in the School of Information Systems, Computing and Mathematics at Brunel University. She obtained her PhD from Department of Information Studies, University of Sheffield, UK in 2000. Her current research interests human-computer interaction, data mining, digital libraries, and educational technology. She has published widely in these areas. Dr. Chen is the editor of several research books, including Adaptable and Adaptive Hypermedia Systems and Advances in Web-based Education: Personalized Learning Environments. She is a member of the editorial boards of five computing journals.

Robert D. Macredie is Professor of Interactive Systems and Pro-Vice-Chancellor, Brunel University. He has worked with a range of organizations, ranging from large, blue-chip companies, through small businesses, to government agencies. Robert's key interest lies in the way in which people and organizations use technology, and his research aims to determine how work can be more effectively undertaken by improving the way that we understand how people and technology interact in organizational (and social) settings. He has undertaken work on a range of issues associated with people, technology and organizations and has over 180 published research contributions in these areas.

Xiaohui Liu is Professor of Computing at Brunel University where he directs the Centre for Intelligent Data Analysis, conducting interdisciplinary research concerned with the effective analysis of data. He was the founding chair of the international conference series on IDA (1995), a keynote speaker at the International Conference of the Royal Statistical Society (2002), and vice chair of the IEEE International Conference on Data Mining (2004). Professor Liu has over 150 refereed publications in data mining, bioinformatics, intelligent systems and time series. He was appointed as Honorary Pascal Professor at Leiden University in the Netherlands in 2004.