

1. INTRODUCTION

This document constitutes a report from the Centre for HCI Design, City University to the JISC for the results of a foundation study in Information Visualisation in response to the invitation to tender “Funding 11/03: Human-Computer Interaction Design and Visualisation Foundation Studies”.

JISC offers wide-ranging services and resources from bibliographic databases to digital maps to the UK's further and higher education sectors and the research community. The study investigated how Information Visualisation could best be applied to these services and resources, with particular reference to the JISC Information Environment.

1.1. Theoretical Foundations

Information Visualisation has been defined as “the use of computer supported interactive, visual representation of abstract data to amplify cognition” (Card et al, 1999). In recent years a number of influential Information Visualisation techniques have been developed and found application in various domains (e.g. starfield displays (Ahlberg & Shneiderman, 1994a), Bead (Chalmers & Chitson, 1992) and self-organizing maps (Lagus et al, 1996)).

With the spread in the use of the World Wide Web new challenges in Information Visualisation emerged. Many visualisations have been developed for the purpose of browsing large collections of documents, or large data sets, such as FilmFinder (Ahlberg & Shneiderman, 1994) and issues of different information visualizations for navigation (Spence, 2001) have been studied (Zaphiris, Shneiderman & Norman, 2002).

1.2. Current Practice

As pointed out in our recently completed Usability Studies for JISC services we believe that Digital Libraries have the opportunity to develop and support user's idea creation and information seeking processes. These generally involve initial ideas that expand as the task develops. Digital Libraries can support this process by incorporating visualisation techniques to help users develop and widen their search. The KartOO (<http://www.kartoo.com/>) is a

characteristic example where the textual interface of traditional search engines is transformed to an information visualisation where search results are displayed in an interlinked network.

1.3. Programme of Work

The programme of work was structured into 4 activities, as detailed below.

Activity 1: JISC Requirements for Information Visualisation

The aim of this activity was to articulate JISC's requirements for Information Visualisation in greater depth. This involved firstly understanding and characterising the broad range of JISC services to establish their specific Information Visualisation requirements (e.g. geospatial systems will have different requirements from portals) and, secondly, investigating current Information Visualisation practices within JISC to identify best practice and opportunities for improvement. Interviews, focus groups and questionnaires were conducted with staff and users of existing JISC services, other digital libraries developers/managers to elicit this information.

Activity 2: Literature Review

An extensive review of literature on Information Visualisation and especially literature that links this discipline with the design of Digital Libraries was carried out to investigate existing theories, frameworks and practices. The review focused on issues relevant to the services offered by JISC, as identified in Activity 1.

Activity 3: Study of Current Practice

This activity was intended to complement the literature review carried out in Activity 2. We interviewed domain experts from the UK and abroad to establish Information Visualisation criteria and best practices for services and resources similar in nature to those offered by JISC. Also we attended the Information Visualisation workshop/tutorial at the British HCI conference and the HCI International conference where we shared our views and also were exposed to a series of views and opinions on the domain of Information Visualisation and HCI. This led to a comparative assessment of similar practices. Our discussions with JISC, domain experts, and our own research guided us in the selection of similar services and resources.

Activity 4: Applying Theory to Practice

The last activity of this project had three main objectives. Our first objective was to evaluate existing information visualisation techniques used by JISC. A number of different JISC services' information visualisation techniques were evaluated in a series of expert (cognitive walkthroughs, heuristic evaluations) and user based (usability testing) exercises with an emphasis on identifying best practice that could be taken up more generally by JISC services and projects.

Our second goal was to engage in the development of Information Visualisation low-fidelity prototypes.

Our team has been very active in applying participatory design for virtual learning environments (Zaphiris and Zacharia, 2001).

Participatory design (PD) (often termed the "Scandinavian Challenge" (Bjerknes, et al., 1987)) refers to a design approach that focuses on the intended user of the service or product, and advocates the active involvement of users throughout the design process. User involvement is seen as critical both because users are the experts in the work practices supported by these technologies and because users ultimately will be the ones creating new practices in response to new technologies (Blomberg & Henderson, 1990).

In this stage of this project, we engaged in innovative participatory focus-group design exercises with users with the goal of evaluating existing information visualisation interfaces/techniques and especially designing low-fidelity prototypes of information visualisation techniques for JISC services. This helped us in recommending adaptations of visualisation techniques and methods used by JISC and others and make them applicable to web-based, end-user-oriented JISC services. Apart from providing low-fidelity prototypes an analysis of the requirements and feasibility for taking these adaptations and designs forward to final designs and implementation is provided.

Our third goal was to develop JISC specific guidelines for designing and evaluating Information Visualisation techniques for JISC services. Building on the existing literature and our prior work in developing usability guidelines for JISC services a comprehensive taxonomy of

guidelines for designing and evaluating Information Visualisation is provided. Emphasis was also placed in developing links between these Information Visualisation guidelines and pedagogical models (e.g. constructivism (Cobern, 1993) and social constructionism (Gergen, 1995)).

In all four activities described above, emphasis was given to the key aspects that JISC emphasises in their call for tender. More specifically, building on our research (Zaphiris & Zacharia, 2001; Armitage, 2002) in investigating the linkage between HCI principles and pedagogical models of web-based learning (e.g. linkage between navigation aids and HCI participatory techniques to enhance learning) our activities above also addressed issues that are directly related to portals and virtual learning environments, emphasizing on how good Information Visualisation techniques/practices can enhance the challenges of online learning for diverse subjects and diverse data types.

1.4. Report Structure

This report summarises the results of Activities 1, 2, 3 and 4 above. For that reason a separate chapter is provided for each activity. Finally the report is concluded with a conclusions and discussions chapter where final remarks are noted.

Throughout this report for consistency we use the term digital libraries to refer to the services funded and supported by JISC Information Environment.

2. LITERATURE REVIEW

2.1. What is Information Visualisation (IV)?

Card et al (1999) defines visualisation as “the use of computer-supported, interactive, visual representations of **data** to amplify cognition” and information visualization (IV) as “the use of computer-supported, interactive, visual representations of **abstract data** to amplify cognition.”

Furthermore, the University of Illinois at Urbana-Champaign Digital Libraries Initiative defines information visualisation as “a method of presenting data or information in non-traditional, interactive graphical forms. By using 2-D or 3-D colour graphics and animation, these visualisations can show the structure of information, allow one to navigate through it, and modify it with graphical interactions.” (UIUC, 98)

To be concise, “information visualisation is a highly efficient way for the mind to directly perceive data and discover knowledge and insight from it.” (Pacific Northwest National Laboratory, 1999)

“The purpose of visualisation is insight, not pictures” (Card et al., 1999). The main goals of IV are discovery, decision making, and explanation. Information visualisation is useful to the extent that it increases our ability to perform these and other cognitive activities. In visualisation, abstractions are based on physical space, whilst non-physical information such as financial data, business information, collections of documents, and abstract conceptions – may also benefit from being cast in a visual form.

Literature makes the distinction between scientific and information visualization. In Scientific visualisation what is represented visually is a physical thing (visualisation is applied to scientific data). Whilst in information visualisation, visualisation is applied to abstract data and it is concerned with abstract concepts such as price, stress, exchange rates, baseball data etc (Spence, 2001).

In the information age, where online information is increasing at an incredible rate as is the diversity of users of such resources; there is a need for new forms of presentation and manipulation of electronic data. Tufte (2001) described two fundamental rules for visual display:

- Maximise the data-ink ratio, i.e. every drop of ink, or pixel on your screen, ought to be information bearing. Anything that appears simply for decoration should be removed.
- Maximise Information Density, i.e. prefer displays with more rather than less information.

Tufte (2001) also observes that old-fashioned maps are a superior means of applying these rules, as they convey more information per square unit of display area than other presentation techniques.

Spatial ability is an important predictor of effectiveness and efficiency when performing common information (i.e. textual) search tasks (Egan, 1988; Vicente & Willeges, 1988; Hook et al., 1996). It is thought that visualising information by representing the semantic relationships (through metaphors like spatial proximity and visual links), can facilitate the development and application of the user's cognitive map of the information space. This makes the task of searching and browsing for information similar to that of real world navigation. For example, a map overviews (as an aid to hypertext navigation) have been found to benefit users (Sein et al., 1993; Stanney & Salvendy, 1995; Vicente & Willeges, 1988). Furthermore, the use of virtual environments takes the notion of visualisation one step further by providing full, real-time interactivity and the ability to view the relationships between objects from an unlimited number of perspectives (Cribbin, 2003).

2.2. Information Visualisation and Digital Libraries

In recent years, the information superhighway, the Internet, has become a global gateway for information dissemination. With the ability to share worldwide collections of information, Digital Library's (DL) have become one of the common means for sharing and disseminating information by individuals or groups that select, organise and catalogue large numbers of documents.

Digital Libraries (DLs) are content rich, multimedia, multilingual collections of documents that are distributed and accessed worldwide. Given the fact that they are becoming the main repository of mankind's knowledge, the design of useful interfaces to access, understand, and manage DL content has become an active and challenging field of study (Fox & Urs, 2002).

DLs, generally referred to as 'collections of information that are both digitised and organised' (Lesk, 1997), give us opportunities we never had with traditional libraries or even with the web.

According to Borner and Chen (2002), our primary means of accessing DLs today are search engines that retrieve very large amounts of relevant documents to our search terms. Search interfaces lack the ability to support information exploration, making it increasingly difficult for scientists and practitioners to gain a macro view of DLs, to locate germane resources, to monitor the evolution of their own and other knowledge domain and to trace the influence of theories within and across domains.

Since new information and knowledge emerge at lightening speed, these floods of information require librarians and information professionals to efficiently and effectively catalogue the large amount of documents produced (Borner and Chen, 2002). There is a need for new tools that will assist scientists, researchers and practitioners across all types of industries to identify and manage all these useful information resources.

By applying information visualisation to DLs' interfaces, the aim is to shift the user's mental load from slow reading to faster perceptual processes such as visual pattern recognition.

According to Borner and Chen (2002), there are three common usage scenarios for visual interfaces to DLs:

- To support the identification of the composition of a retrieval result, understand the interrelation of retrieved documents to one another, and refine a search.
- To gain an overview of the coverage of a digital library and to facilitate browsing.
- To visualise user interaction data in relation to available documents in order to evaluate and improve DL usage.

Furthermore, Borner and Chen (2002) give suggestions on effective design solutions towards information visualisation and how visualisation could assist digital libraries:

- Research based on a detailed analysis of users, their information needs, and their tasks.
- Provide new means to interact with data, for example, to provide an overview of what is covered by a DL, enable users to filter out relevant documents and to examine relationships

among those documents.

- Good labelling to ensure that the selection of meaningful words and their display would not be over plotted.
- To provide multiple perspectives to one data source.
- To support searching of text, image, video, spatial data
- Personal baskets for users to store previously selected document sets
- Usability and usefulness studies are needed to improve interfaces and to specify what does and does not work.
- Strong collaboration among librarians and programmers will help to improve the design usability of interfaces considerably.

2.3. The application of Information Visualisation to Digital Libraries

When designing digital libraries, designing meaningful overviews so that patterns can be easily recognised, creating comprehensible interfaces to specify what they want, and providing effective displays of search results are real challenges. The visual information seeking mantra: "Overview first, zoom and filter, then details on demand" provides a good starting point. By interpreting this mantra, a remarkably diverse set of research and commercial interfaces (Shneiderman, 1998; Card, Mackinlay and Shneiderman 1999; Reed & Heller, 1997) demonstrating the use of IV in DLs have been produced.

Searching

Search results from DLs are often displayed as a textual list, with 10 to 20 items per page. Shneiderman et al (1998) proposed a framework that increases clarity and user control while it reduces inconsistencies in text-search user interfaces. This allows designers of specific systems to offer a variety of features in an orderly and consistent way. Users begin the search process by considering their information needs and clarifying their search goals, after which they

are ready to employ a computer-based system for the four phases: formulation (what happens before the user starts a search); action (starting the search); review of results (what the user sees resulting from the search); and refinement (what happens after review of results and before going back to formulation). Incorporating this framework into the design of digital libraries increases clarity and user control and reduces inconsistencies in text-search user interfaces.

Shneiderman et al (1999) suggested the use of a two-dimensional display with continuous variables to view several thousands search results at once. It has been applied to a digital video library, a legal information system and to a technical library using the ACM Computing Classification System.

Navigation

In terms of navigation, research in information visualisations (Bederson & Shneiderman, 2003; Shneiderman, 1998; Card, Mackinlay and Shneiderman, 1999) produced a remarkably diverse set of research and commercial interfaces.

Two important strategies are:

1) Two-dimensional visualisations:

Two-dimensional visualisations with meaningful axes have been an effective visualisation strategy because thousands of items can be shown at once. Typical axes make use of continuous variables (e.g. relevance or publication date) or categorical variables (e.g. language or name of journal). The labels on such axes are cues to searchers and may contain valuable information if they are ordinal variables.

2) Browsers for hierarchical data sets.

Hierarchical structures are common nowadays and often successful, (e.g. Yahoo!). Hierarchies have the potential to reduce complexity by organising related information into comprehensible structures. Hierarchies are typically shown by textual lists of one level at a time, by node-link diagrams, cone trees (Hearst and Karadi, 1997) or an outliner in which levels and branches can be expanded and contracted (Nation et al, 1997). GRIDL (Shneiderman et al, 1999) (Graphical Interface for Digital Libraries) adopts these two important strategies by

combining the use of hierarchical browse tool with a two-dimensional visualisation. This combination preserves visual overviews and enables users to rapidly comprehend search results.

Browsing

According to Shneiderman (1997), tasks can range from specific fact-finding to more unstructured open-ended browsing of known databases and exploration of the availability of information on a topic:

1) Specific fact-finding (known item search)

e.g. find the Library of Congress call number of Future Shock

2) Extended fact-finding

e.g. What kinds of music is Sony publishing?

3) Open-ended browsing

e.g. Is there a new work on voice recognition in Japan?

4) Exploration of availability

e.g. Can NASA data sets show acid rain damage to soy crops?

This focus on tasks lead to the Object Action Interface Model (OAI) (Shneiderman, 1997) that emphasizes objects and actions.

Several techniques have been developed in the last few years focusing on interfaces that allow users to browse and search collections of information. Example include: the dynamic queries (Ahlberg et al, 1992), visual information seeking (Ahlberg & Shneiderman 1994) and query previews (Doan et al., 1996; North et al., 1996).

Dynamic query interfaces extend direct manipulation to search tasks by closely coupling search specification with the display of results. These interfaces feature a visual representation of a database (typically with a scatterplot or starfield display), a visual representation of a query using a collection of widgets (e.g. sliders), and tight coupling between these two components. Users browse the database by interacting with the interface widgets. Each change produces a new query, the results of which are immediately and continuously shown in the display, supporting a progressive refinement of the search, continuous reformulation of goals, and visual

scanning to identify results. Query previews also deal with large distributed databases, and use previews of the data to maintain real time feedback and limit access to the network.