International Journal of Information Management 35 (2015) 244–252

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Using visual representations for the searching and browsing of large, complex, multimedia data sets

H.J. Grierson*, J.R. Corney, G.D. Hatcher

University of Strathclyde, Design Manufacture & Engineering Management, 75 Montrose Street, Glasgow G1 1XJ, United Kingdom

ARTICLE INFO

Article history:
Available online 28 January 2015

Keywords:
Searching
Browsing
Information visualization
Information management
Zoomable

ABSTRACT

Recent years have seen a huge increase in the digital data generated by companies, projects and individuals. This has led to significant challenges in visualizing and using large, diverse collections of digital information. Indeed the speed and accuracy with which these large datasets can be effectively mined for information that is relevant and valuable can have a significant effect on company performance. This research investigates the feasibility of using visual representations for the searching and browsing of large, complex, multimedia data sets. This paper introduces the SIZL (*Searching for Information in a Zoom Landscape*) system, which was developed to enable the authors to effectively test whether a 2.5D graphical representation of a multimedia data landscape produces quantifiable improvements in a user's ability to assess its contents. The usability of this visualization system was analyzed using experiments and a combination of quantitative and qualitative data collection methods. The paper presents these results and discusses potential industrial applications as well as future work that will improve the SIZL data visualization method.

1. Introduction

The quantity of data generated and stored globally is increasing at a phenomenal rate. In 2013 it was said that 90% of the world's data were generated in the past two years (Dragland, 2013). The digitization of business, global industry partnerships and the increasing presence of the Internet in our lives have all contributed to datasets so large that they have become highly challenging to manage effectively. Companies that operate from a digital platform, for example internet retailers and social networks, face great challenges in capturing, storing, analyzing and protecting the huge volumes of data their businesses generate. Even 'traditional' industries such as engineering and construction are facing challenges. Large, global, collaborative projects also generate huge volumes of data, from design documentation to supply chain management to communication records. The speed and accuracy at which these large datasets can be effectively mined for information that is relevant and valuable can have a significant effect on company performance (Keim et al., 2008). Therefore there is a need for systems which enable the effective management of this 'big data'.

This paper presents the preliminary findings from an evaluation of a data visualization system designed to enable faster and more accurate searching and understanding of large datasets. The first section introduces the concept of data visualization systems and presents the aim and hypothesis of the study. The second introduces the SIZL visualization system and defines the research methodology used to evaluate its effectiveness. The third section presents the findings from the quantitative and qualitative data gathered during evaluation, and the fourth section discusses the significance of these results as well as the limitations of the research at this stage. The paper concludes with details of further work required to refine the research and provide further insights into the benefits of the data visualization system.

1.1. Data visualization

The rate at which data can be collected and stored is outgrowing the rate at which it can be analyzed (Keim, Mansmann, Schneidewind, & Ziegler, 2006). As the size of datasets grows exponentially, there is increasing risk that much of the valuable and relevant information stored is being lost due to ineffective systems for data exploration and visualization (Keim, 2001).

Traditional, 2-dimensional methods of data visualization include charts, graphs and plots. These visual ways of displaying

* Corresponding author. Tel.: +44 0141 548 5939.
E-mail address: h.j.grierson@strath.ac.uk (H.J. Grierson).

http://dx.doi.org/10.1016/j.ijinfomgt.2014.12.003

data have been designed to communicate information in a way that humans can more easily understand and analyze. Chen and Yu (2000) conducted a meta-analysis of information visualization research focused on users, tasks and tools. The research revealed that given a consistent level of cognitive abilities, individuals showed a preference for simple visual–spatial interfaces. In other words, processing visual information is more intuitive to humans than processing other types of information such as text or numbers. Over the years, methods of data visualization and interactive graphics have become increasingly sophisticated. This progress has also resulted in an increase in the use of 3D visual displays to present greater complexity in datasets and enable a more interactive, intuitive data searching experience. Some well-known examples of data visualization include Google Earth, Google Images and ChronoZoom (Alvarez & Saekow, 2014; Google, 2014a, 2014b), which allow users to search for information in a visual, interactive and multi-dimensional environment.

1.2. State of the art

There has been much debate in the literature over the merits of 3D visualization systems—do they genuinely improve the effectiveness of information retrieval and analysis? Earlier studies in particular promoted 3D visualization as more intuitive (Ark, Dryer, Selker, & Zhai, 1998; Robertson et al., 1998). However, later studies have questioned this assertion. Cockburn (2004) evaluated data storage and retrieval tasks in 2D and 3D visualizations. The study concluded that whilst 3D systems emulate a more 'natural' environment, their benefits are task-specific. Kosara, Hauser, and Gresh (2003) also state that 3D visualizations can have detrimental effects on users such as increased workload, occlusion and disorientation. Schneiderman (2003) highlights that 3D visualizations can simplify tasks and improve interactions only if properly implemented. Clearly 3D is not without merits, but its application must be carefully considered to ensure it is truly providing benefits to the desired task(s).

Users can process visualizations faster than text, and inexperienced users can navigate 3D interfaces more intuitively than 2D interfaces. However, several issues affect 3D visualization, such as context, interpretation, cognitive and dimensional overload (Pfitzner, Hobbs, & Powers, 2003). The fine balance between beneficial and gratuitous use of 3D in data visualization has led several researchers to recommend the use of hybrid or 2.5D interfaces (Baumgartner, Ebert, Deller, & Agne, 2007; Wiza, Walczak, & Cellery, 2004). Such environments can provide users with the cognitive/spatial advantages of 3D whilst retaining the refined interactions of 2D (Baumgartner et al., 2007), therefore reducing the chance of users becoming 'lost' in the system.

Other studies have explored in more detail the preferred functionalities for effective data visualization systems. Bergman, Beyth-Marom, Nachmias, Gradovitch, and Whittaker (2008) found that users show a preference for navigation over searching when locating files that have a set structure, for example folders or e-mails, and argue that navigation reduces cognitive workload, because individuals are psychologically programmed from childhood to store and retrieve objects from locations. Whereas searching relies on an individual's ability to associate attributes to an object, for example the file name of a document. Exploring navigation further, Hornbæk, Bederson and Plaisant (2001) studied the use of overviews in user interfaces. Participants showed a preference for a navigation overview which allowed them to keep track of their actions, however the researchers found that this overview slowed down performance, possibly due to increased workload. They propose the implementation of a 'zoomable' interface to overcome these issues.

1.3. Research aims

Over the years there have been continuing advances in low cost, high performance 2D and 3D display and manipulation technologies, as well as ever-increasing computation power. At the same time, the huge increase in data generated by companies, projects and even individuals has led to great challenges in visualizing and searching for information. This project emerged from the idea that exploiting the human "cog" within these systems provides an opportunity to redress the balance between high volume information/data storage and effective navigation. Thus finding information more easily.

Therefore, the overall aim of this research was to investigate the feasibility of using visual representations for the searching and browsing of large, complex, multimedia data sets. Drawing upon prior research in this field, the following hypothesis was tested:

> *Human beings find the recall and recognition of 2D and 3D shapes and environments so intuitive and effortless that any system for the effective management and use of data should make use of this fact.*

In addition to this hypothesis, a number of more specific questions were raised during the early stages of the research, including:

- Can a new system allow an "at a glance" pictorial summary of its content?
- Can the information interface allow users to spot relationships in data more easily by "illustrating" the contents of files through icon representations that reflect the context of information?
- Can an advanced visual interface (a 'zooming timeline') impact on a user's ability to quickly and accurately find individual items and identify relationships within subsets of the data?

A system was developed that would enable the researchers to answer these questions and effectively test whether 2.5D environments can benefit effective data management.

2. System overview and research methodology

2.1. Introduction to SIZL

The SIZL (*Searching for Information in a Zoom Landscape*) system was created to evaluate user interaction and experience with data in 2.5D environments, and enable the researchers to evaluate the effectiveness of this method. This software prototype combines a zooming user interface (ZUI) and a timeline—a zooming interface to a visual information landscape—and was designed with the capability to extract data from numerous document types such as word documents, spreadsheets, PDFs and image files. The software has a multi-search functionality, allowing users to search within the dataset for multiple keywords or phrases that are highlighted simultaneously using different colours. Captured data can then be moved to the 'lightbox' area to be compared and contrasted, enabling the user to identify document relationships. The SIZL system process is summarized in Fig. 1. The system is database-driven and facilitates the creation of dynamic user interfaces in response to user inputs. It was developed in the Net environment using C#, connecting to a MySQL database, using Sphinx searching technology to index and search the system content.

Based upon findings from the literature regarding the advantages and disadvantages of both 2D and 3D visualizations, SIZL was designed as a 2.5D (or hybrid) environment. The user can interact with the system through direct searching, for example key words or document browsing (i.e. the multi-search functionality). Relevant documents are extracted by the system and used to generate