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
0268-4012/

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

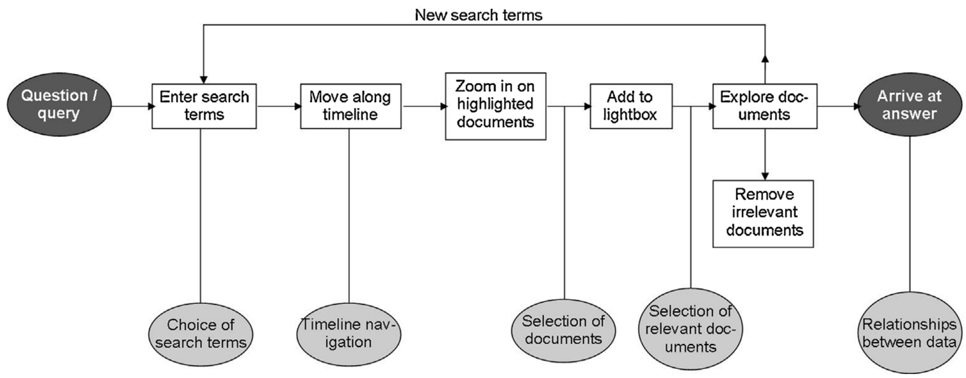

Fig. 1. SIZL system process with human interactions and 'human cog' inputs.

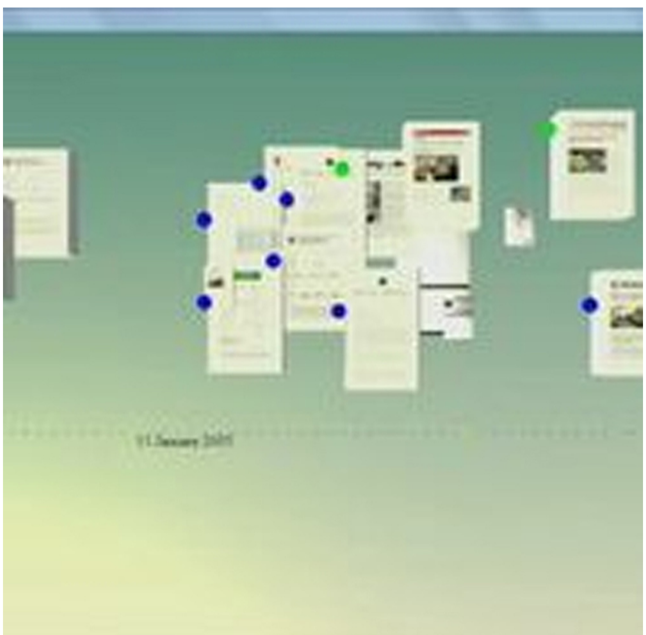

Fig. 2. Entering search terms into SIZL highlights relevant documents along the zooming timeline.

Fig. 3. The SIZL interface provides a visual summary of each document in the dataset.

objects which are then presented to the user in the 'timeline' section of the Zooming User Interface (ZUI)—see Fig. 2. Objects are presented as 3D blocks with differing depths, providing the user with an 'at a glance' pictorial summary of the content of the document each model represents (see Fig. 3). The user can focus in detail on the objects they deem relevant by dropping them into the 'lightbox' section of the ZUI (see Fig. 4). The user will remove interface options that are not considered to be relevant and replace them with objects that relate to the direction the user is taking within the application, allowing them to focus on a limited number of chosen documents. The system's functionality is summarized in a You-tube video (http://www.youtube.com/watch?v=PeEbfwKQu30)/ Video 1 which accompanies this paper..

The SIZL system was designed to evaluate four key elements:

Timeline: *provides instinctive chronological flexibility*.

Time is a form of metadata applicable to almost all information. The timeline class consists of multiple arrays of time-scales—days, hours and minutes. The user has the ability to set the upper and lower limits of the timeline which will in turn limit the availability of the documents to those that fall within the set limits. The user can also expand or contract the timeline allowing them to manipulate the volume of displayed documents.

Zoom: *emulates time, facilitates user controlled data convergence.*

Zooming provides an infinite landscape. The zooming functionality has been created by fixing a camera to the 2D *x*-plane in the 3D environment. This allows users to zoom in and out of the *Z*-axis which changes the distance of the camera plane in relation to the displayed objects. This allows the user to control (in conjunction with the timeline) the amount of documents currently on display by the system. Furthermore, it allows the user to focus on clusters of documents that may be of particular interest.

Visualizations: *inter-document relationships.*

Human search is powerful because we notice patterns, oddities and depth/distance. The database contains a vast range of data which, with the multi-search functionality, is used by the system to create arrays of objects which have underlying connections. These relationships are used to generate the default ZUI. The goal of exploiting the underlying relationships is to allow the user to quickly retrieve information based on whatever knowledge they

Fig. 4. The lightbox feature enables users to collect and analyze relevant documents.

have relating to the search. For example, knowing the name of a person or place but not knowing the title of a document.

Interaction: *concepts to match input/output devices.*

The ZUI is based around an interactive timeline which can be customized by the user. The ZUI allows the user to zoom in and out of documents; manipulate document positions and scales; retrieve key information relating to a document; and, open original documents.

Section 2.3 discusses how the SIZL system was used to test the research hypothesis. It should be noted that the SIZL system is a prototype under development. It was designed to support the study of how advanced visual interfaces impact upon a user's ability to find individual items and identify patterns within subsets of data, and therefore it is the interactions enabled by SIZL that were being evaluated, not the usability of the software itself.

2.2. Methodology

An experimental methodology was adopted so that the SIZL system could be used to evaluate the use of human cognition in effective data management. The methodology progressed in four phases:

- Requirements
- Design
- Experiments (the evaluation process is discussed further in Section 2.3)
- Documentation, demonstration and dissemination

Very few previous studies of visualization methods have included adequate user evaluation (Ellis & Dix, 2006), meaning it is difficult to make conclusions regarding their effectiveness and applicability. The use of an experimental methodology to conduct user evaluation is a significant contribution of this research.

Perhaps one of the reasons behind this lack of user evaluation is the considerable number of challenges the task poses, some of which are discussed in the literature. One of the main complications in this kind of study is the use of custom developed software. The software itself is not tested as part of the research, however, inevitably bugs and unfamiliarity with the system can have an impact upon participant performance and experience (Andrews, 2006). Similarly, the unfamiliarity of a new system can create a bias for any existing systems that are included in the experiment (Andrews, 2006). These issues were observed in the SIZL evaluation, however their effects were alleviated through the use of two pilot experiments and the presence of the SIZL software developer to assist participants with any software-related problems during the tasks.

Another criticism of visualization evaluations is the use of students (Andrews, 2006; Ellis & Dix, 2006), as they will not have the necessary experience and expertise to carry out the evaluation tasks asked of them, thus affecting their performance and quality of feedback. Eighteen design engineering students were used in this study (2 for the pilot and 16 for the experiments). However, rather than tackle unfamiliar tasks, the students in this study were asked to use SIZL to answer questions about a Hollywood spy movie. The decision to use a popular culture reference meant that students were more engaged with the task and able to digest instructions more easily, regardless of whether they had seen the movie or not. Furthermore, there were advantages to using a student demographic: this age group is typically very familiar with visualization systems, in particular games technology, and as such they could use and learn the new software relatively quickly.

According to the literature (Andrews, 2006; Ellis & Dix, 2006; Sebrechts, Cugini, Laskowski, Vasilakis, & Miller, 1999), a good evaluation of a visualization system will provide training in the new system, use appropriate tasks for testing that provide meaningful results for the kind of tasks the new system is designed for, and use appropriate measurements. Section 2.3 will discuss the evaluation process that was carried out for this research.

2.3. Evaluation process

In order to evaluate the effectiveness of exploiting the human cog in the system for effective data management, experiments were conducted with the SIZL system. Microsoft's 'File Manager' system—the predominant data management system used in industry today—was used so a comparison could be made between the speed and accuracy of human interaction against a baseline. The fictional story from the 2007 movie '*The Bourne Ultimatum*' was used as a case study for the evaluation.

Firstly, SIZL and File Manager were populated with identical sets of documents (104 documents in total) relating to *The Bourne Ultimatum* story, including newspaper articles, receipts and airline passenger manifests. Six questions were set relating to the movie—two 'easy', two 'medium' and two 'difficult'. Participants were asked three questions for SIZL and three questions for File Manager. Which question was asked of which system was alternated between tests. The difficult tasks (questions 3 and 6) were specifically designed to test user's ability to retrieve file relationships. Questions included:

- Easy: e.g. What weapon is used by the assassin who was not working for the CIA?
- Medium: e.g. Which of his aliases did Bourne use to travel from Moscow to Paris?
- Difficult: e.g. *Who gave the order to assassinate Simon Ross?—Give your evidence.*

Following two pilot tests, 16 participants were invited to take part in the study. Participants were university students, some of whom had seen the movie and some who had not. In a short survey conducted before the test, all participants stated that they were confident with their IT skills. All spent at least an hour a day on a

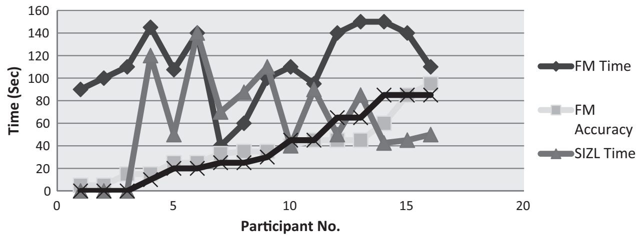

Fig. 5. Participants' time and accuracy when tested on both File Manager and SIZL.

computer or online, with 7 participants stating that they spent over seven hours a day. This survey indicated that the participants were in general highly familiar with visualization systems.

Firstly, participants were given a basic File Manager tutorial and a SIZL tutorial to familiarize themselves with the software. Next, participants each spent 20 min answering three questions on File Manager (acting as an unorganized filing system) and 20 min answering three questions on SIZL. They answered an easy, medium and difficult question for each system. The number of correct documents found (accuracy) and the time it took them to answer questions was recorded (summative testing). Participants were also encouraged to 'think aloud' and their thoughts were recorded throughout (formative testing). A facilitator was always available to assist with use of the software.

Participants were then asked for immediate feedback through a semi-structured questionnaire. They were asked about the ease of finding information and determining relationships between documents. Some were later invited back for a focus group where they were presented with the overall results. They were then asked to reflect on these results, and discuss their views the benefits and drawbacks of both File Manager and SIZL.

2.4. Analysis

For each experiment conducted, both quantitative and qualitative data were gathered: the measurement of task accuracy and time; the collection of user experiences from facilitator observations and the 'think aloud' exercise; and the results of the questionnaire participants were asked to complete at the end of their test. This allowed the research to explore both the effectiveness of the visualization methods utilized in SIZL and users' preferences for the system.

3. Results

This section presents the findings from the 16 experiments, and the focus group that was conducted at a later date. Recordings of the time and accuracy to complete a set of tasks (answer three questions using SIZL and three questions using File Manager); the qualitative data from observations and the participant survey were analyzed to identify key findings. The focus group was used to validate these findings.

3.1. Quantitative findings

During testing, facilitators recorded both the times in which participants answered the three questions for SIZL and the three questions for File Manager, and the accuracy of their answers (i.e. the number of relevant documents the participant retrieved). The results are displayed in Fig. 5.

By plotting time and accuracy together (Fig. 6), it can be seen that many participants were able to achieve the same accuracy in SIZL as File Manager, in significantly less time. This relationship was particularly clear for the difficult questions (3 and 6), which were specifically developed to evaluate participants' ability to identify relationships between documents. No correlation was found between participants' performance in File Manager and SIZL.

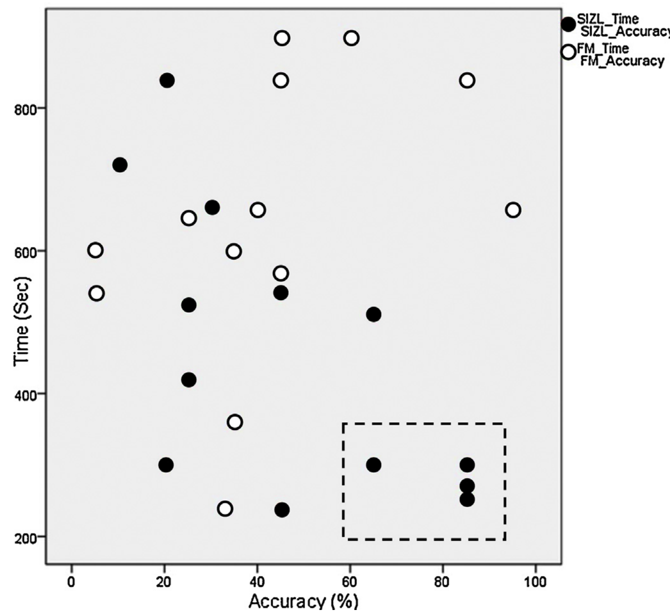

Fig. 6. Some participants were able to achieve high accuracy in less time using SIZL (highlighted).

3.2. Observation findings

During testing, facilitators recorded both their observations of participant behaviour, and any comments made by participants during each session (using the 'think aloud' method). These notes were then analyzed using NVivo software to identify the most beneficial and popular features of the SIZL system, as well as the most common issues and criticisms. Notes were coded when a particular feature of the SIZL system was recorded as being used to identify the correct answer to a question, or was commented upon positively by a participant. Notes were also coded when participants expressed difficulty or dislike for a particular aspect of the system.

Firstly, considering the most useful features of the SIZL system, by far the most commonly used feature was the multi-search functionality. Participants normally began their search for data by entering a number of search terms into the system, and would often use trial and error to retrieve the correct documents. The second feature most commonly referred to in the notes was the 'lightbox', which participants used to examine promising documents in greater detail. Based on the testing notes, it would appear that the zoom and timeline functions were used less frequently in determining the correct answers to the questions.

Focusing on participant comments, again the multi-search functionality received much praise. For example:

> '*The multi search was great, it was very useful for seeing documents.*'

The overall visualization method of the SIZL system also scored highly, with participants expressing a preference for the visual overview of the dataset that SIZL provides, and commenting on how much easier it was to visualize relationships between documents. The zoom and timeline features also received some positive commentary.

> '*SIZL provided a good way to visualize the overall dataset.*'

> '*I found SIZL very effective in highlighting relationships - in File Manager I felt I had to hold a lot of info on relationships between*

documents in my head, and then forget. Visually seeing the relationships made it much easier to narrow down the docs to find what I was looking for.'

Overall, comments on the SIZL system were highly positive. As predicted, the vast majority of negative comments concerned the SIZL software design and on-screen layout of information; aspects not being tested in this study. In particular, some participants expressed difficulty in identifying highlighted documents and difficulty moving documents around the screen with the keyboard. These problems could be alleviated with improvements to the software prototype, and are not directly linked to the developed visualization system. Although the design of the SIZL software was not being tested in this study, it inevitably had an impact upon participants' experiences and ability to retrieve the correct answers to the questions. The second most common negative comment relating to SIZL was unfamiliarity with the system. This is a common problem with the testing of any new visualization system, and was in part alleviated by providing each participant with a SIZL (and File Manager) tutorial prior to testing; and access to the systems developer during experiments. Another problem situation that occurred on several occasions was participants getting 'lost' in the system—going on tangents and struggling to get back on the right track.

As well as providing an insight into the most beneficial aspects of the SIZL system, the observation data also demonstrated the 'human cog' at work when participants were using SIZL. The facilitators regularly recorded use of reasoning and 'real world thinking' when participants were navigating the system for answers, and participants were often observed taking notes while problem-solving. One participant was also observed searching for words that were not included in the question, illustrating how the human mind plays a significant role in the search for relevant information when using the SIZL system.

In summary, the key findings from participant observation were:

- The multi-search function was the most useful and most popular feature of the SIZL system.
- Participants responded positively to the visualization provided by SIZL.
- The SIZL system utilizes the 'human cog' in the system: participants used the system to navigate their own problem-solving processes.

3.3. Survey findings

Participants were also asked to complete a survey at the end of the test. This Likert scale survey asked participants about their experience of using SIZL and their views on its effectiveness as a visualization system. Survey results are presented, in descending order of positive response, in Fig. 7. Corresponding to the findings from participant observation, the question relating to the SIZL system's multi-search functionality (Q6) received the most positive response. Participants also largely responded positively to the usability of the SIZL functions (timeline, zoom, lightbox, etc.) and visual layout of the system. Although unfamiliarity was a common complaint during participant testing, the survey results show that the majority of participants felt that the SIZL system was easy to learn, suggesting this problem would quickly dissipate with experience.

Most of the 'negative' responses to the survey were in response to negatively keyed questions. However, a significant number of participants responded negatively to the statement 'I found I could quickly recover when errors occur in this system' (Q4). This could be linked to problems with the software design, as discussed in Section 3.2, but could also be linked to the issue of several participants getting 'lost' in their search for relevant documents.

In summary, the key findings from the participant survey were:

- Participants responded highly positively to the multi-search function in particular.
- The majority of participants expressed a preference for the visual layout of the SIZL system.
- Some participants struggled to recover when errors occurred in the system.

3.4. Focus group findings

Once all data had been gathered and analyzed, participants were invited to attend a focus group to discuss the results and reflect on their experience of the SIZL system. During the focus group, the participants were presented with the key findings from the research and asked to discuss a variety of questions and topics, including:

- Did you have a preference for either system?
- For each system give the benefits and drawbacks.
- In terms of looking for information in large datasets, please comment on SIZL's multi-search/timeline/zoom features.
- Please comment on SIZL's ability to support the visualization of inter-document relationships.

This process contributed to the validation of the research results, by reaffirming the key findings from research and providing further insight into the reasons behind these results. The key outcomes from the focus group were:

- In correlation with the qualitative data findings, the multi-search function was the most popular feature of the SIZL system. The timeline and zoom features, on the other hand, were not used as readily with some participants unaware of these features.
- Also mirroring the key findings from the tests, errors in the SIZL software were highlighted as a key issue. Although the usability of the software design was not tested as part of the research, it inevitably had an impact upon the results. For example, it is possible that the lack of interest in the timeline and zoom features could be attributed to the design of the SIZL software.

4. Discussion

4.1. Significance of research findings

This study set out to test the hypothesis that 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. This was done through the evaluation of a visualization system, SIZL, which enables users to discover relationships and patterns within large datasets.

A small, preliminary evaluation was conducted, and findings suggest that the SIZL system, using a zooming timeline and multi-search visualization methods, does indeed enable users to achieve similar accuracy in less time when compared to traditional text-based searching. Also, participants expressed a preference for SIZL when asked to identify relationships between documents. These findings would suggest that people do find certain data mining tasks more intuitive when searching within a 2.5D environment. The findings from the qualitative data suggest that the multi-search feature in particular has high potential

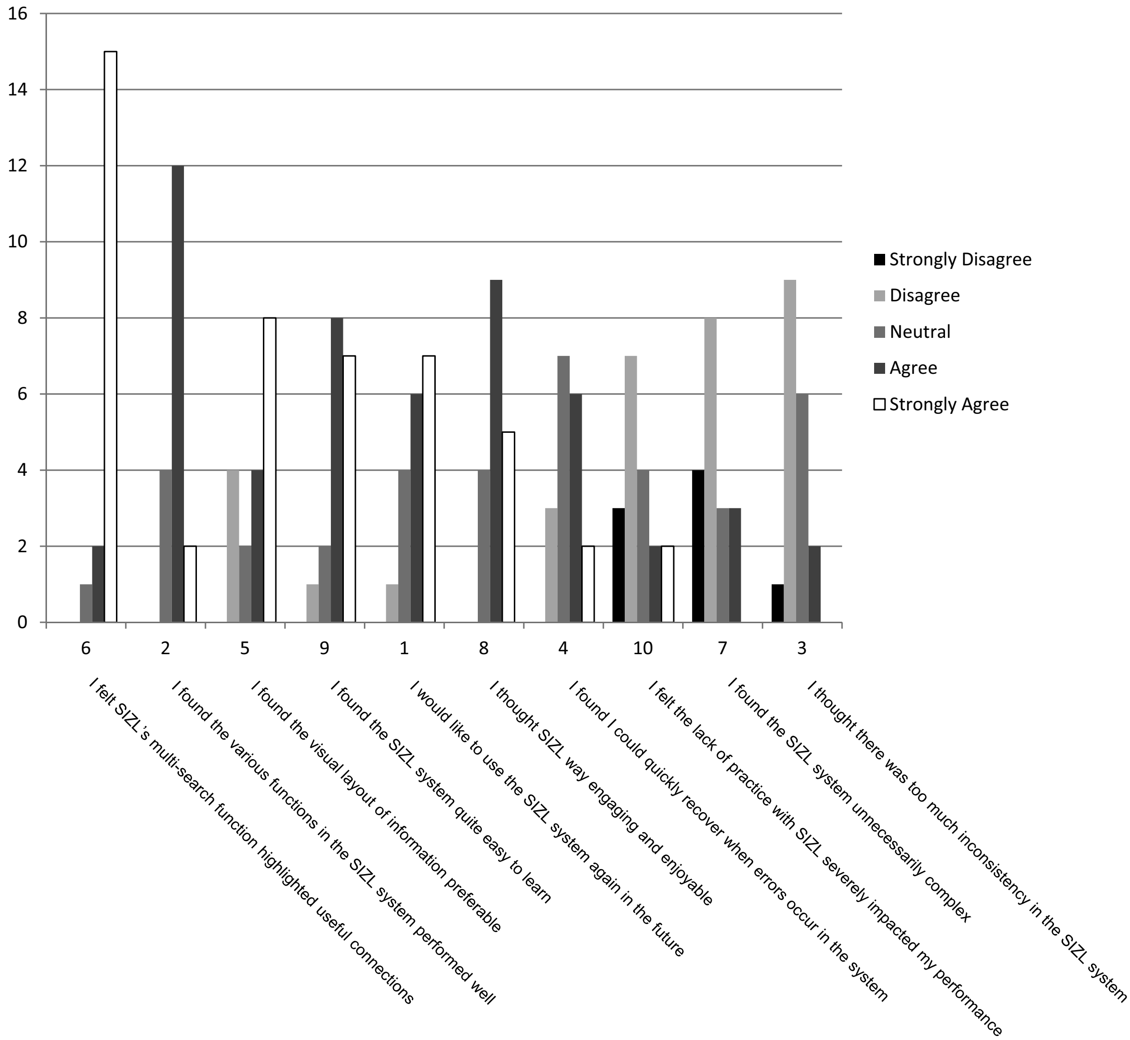

Fig. 7. Participant survey results.

to improve users' data mining experience. The findings also highlighted some key areas for improvement—better usability of the zooming timeline functionality and an improved navigation experience for users who make wrong decisions and get 'lost' in the system. Improvements to the usability of the SIZL software would also enable more accurate testing of the visualization method.

One of the key contributions of this study was the methodology—the use of user evaluation experiments to explore the effectiveness of a new data visualization system. It has been established that there is a lack of evidence of user evaluation in the visualization literature, so it is anticipated that this method may be refined and replicated to build upon the preliminary findings of this study, and may also act as a case study for future evaluations of visualization systems.

4.2. Research limitations

This paper has presented the results from the preliminary evaluation of a new data visualization system. During this phase the testing process was improved through the use of two pilot tests, however there were still a number of limitations to the research. Some can be attributed to the early stage of development of this system, others are common challenges faced when attempting to evaluate any visualization system.

4.2.1. Number of participants

Early evaluation of the system involved only 16 participants. Whilst this is considered an acceptable number for the qualitative elements of the methodology (observations and focus group), a sample size of 50–100 participants is more appropriate for identifying statistically significant findings (Andrews, 2006). Therefore, although the statistical analysis of the 16 experiments provided interesting and significant insights into the merits of the SIZL system, further evaluation with a larger number of participants is required to provide more conclusive results.

4.2.2. SIZL software design

Within the constraints of the project, the SIZL software was designed to be easy to use and accessible, with a view to ensuring the SIZL interface had as little negative impact upon participants' experience as possible. This is because it was the visualization method that was being evaluated, not the system's interface. However, it was very difficult to completely remove the influence of software design from the evaluation. Qualitative data collected suggested that unfamiliarity with the SIZL software had an impact on participants' ability to complete tasks and fully experience the visualization system—for example, some participants reported not noticing the timeline functionality, or finding the zoom function difficult to use, even having reacted positively to the concept. Such feedback will enable the SIZL software to be improved for future evaluation.

4.3. Implications for industry

As global partnerships and collaborative workspaces, not to mention the 'big data' phenomenon of increasingly digitized businesses, lead to increasingly large and complex datasets, traditional text-based search systems are inadequate for effective information management. The functionality provided by the SIZL system has the potential to improve the data mining experience in a wide range of industries with a need to manage large numbers of documents, understand the timing of these documents and the relationships between them.

There are a great variety of industry sectors that could benefit from an improved method of mining relevant data. Three key sectors identified for this research were Defence, IT and Manufacturing. The UK's defence export market exceeded £7bn for the first time in 2009 and today accounts for nearly 20% of the global market. Similarly manufacturing is an important part of the UK economy. It accounts for 12.8% of UK gross domestic product (GDP) and 55% of total exports. The software industry although smaller by comparison is pivotal to the products of many other enterprises. The proposed SIZL system could support business needs of all three sectors. This is particularly relevant in the defence industries where security is paramount, and information has to be stored and retrieved—in near real time—safely from increasingly diverse and distributed sources.

One of the key barriers to implementing a new system in any workplace is 'familiarity bias': even if the new system eventually provides an improved experience, the effort and time required to become familiarized with the new system means that people tend to express a preference for the old, familiar system (Andrews, 2006). This was also an issue during the evaluation of SIZL, and, although the software design was not being tested as part of the study, it inevitably had an effect upon participants' experience. The problem of familiarity bias could be alleviated by combining the implementation of the new visualization with a high standard of user-friendly interface design.

5. Conclusions

This paper has presented the findings from the early phase of a study of exploiting the human 'cog' within a data visualization system; testing the hypothesis that 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. The paper also presented the experimental methodology that was used to evaluate the functionalities of the SIZL system: data visualization and searching software that was developed specifically for this study. Although further evaluation is required to provide conclusive results, this first phase has indicated that the SIZL system does help users search for and identify relationships between documents in large datasets, when compared to a traditional text-based system. The key findings from this preliminary study can be summarized as follows:

- Participants were able to achieve the same accuracy using SIZL, in less time, when compared to results for File Manager.
- The multi-search functionality was the most commonly used and most popular feature of the SIZL system during testing.
- Participants stated a preference for the overall visualization of the SIZL system.
- The zoom and timeline functionalities were not used as often as expected.
- Some participants experienced problems of getting 'lost' in the SIZL system.

The next stage in this research is a refinement of the evaluation process and further testing with improved software and a larger pool of participants to provide in-depth and reliable insight into the benefits of using a data visualization system based on the SIZL model, as well as a greater understanding of the potential applications for a system that utilizes the human mind to provide a richer user experience.

Acknowledgements

The authors would like to acknowledge the EPSRC and Dstl who funded this study as part of the Data Information Systems (DaISy) Programme (EPSRC Grant EP/J020338/1) and also the system developers Bryan Young and Dr Andrew Lynn and the project evaluator, Jennifer Griffith, all of the University of Strathclyde, Glasgow.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijinfomgt.2014.12.003.

References

Alvarez, W., & Saekow, R. (2014). ChronoZoom. [Online] Available from: http://www.chronozoom.com/. [Accessed: 27th March 2014].

Andrews, K. (2006). Evaluating information visualisations. In *Proceedings of the 2006 AVI workshop on beyond time and errors: Novel evaluation methods for information visualization (BELIV '06)*. New York: ACM.

Ark, W. S., Dryer, D. C., Selker, T., & Zhai, S. (1998). Representation matters: The effect of 3D objects and a spatial metaphor in a graphical user interface. In *Proceedings of HCI on people and computers XIII*. Springer-Verlag.

Baumgartner, S., Ebert, A., Deller, M., & Agne, S. (2007). 2D meets 3D: A human-centered interface for visual data exploration. In *CHI '07 extended abstracts on human factors in computing systems*. San Jose: ACM.

Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., & Whittaker, S. (2008). Improved search engines and navigation preference in personal information management. *ACM Transactions on Information Systems (TOIS)*, *26*, 4.

Chen, C., & Yu, Y. (2000). Empirical studies of information visualization: A meta-analysis. *International Journal of Human-Computer Studies*, *53*, 851–866.

Cockburn, A. (2004). Revisiting 2D vs 3D implications on spatial memory. In *Proceedings of the fifth conference on Australasian user interface (Vol. 28)*. Australian Computer Society, Inc.

Dragland, A. (2013). Big Data - for better or worse. [Online] Available from: http://www.sintef.no/home/corporate-news/Big-Data–for-better-or-worse/. [Accessed: 27th March 2014].

Ellis, G., & Dix, A. (2006). An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: Novel evaluation methods for information visualization*. ACM.

Google Inc. (2014a). Google Earth. [Online] Available from: http://www.google.co.uk/intl/en_uk/earth/. [Accessed: 27th March 2014].

Google Inc. (2014b). Google Images. [Online] Available from: https://images.google.com/. [Accessed: 27th March 2014].

Hornbæk, K., Bederson, B., & Plaisant, C. (2001). Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *9.4*, 362–389.

Keim, D. (2001). Visual exploration of large data sets. *Communications of the ACM*, *44*, 38–44.

Keim, D., Andrienko, G., Fekete, J., Gorg, C., Kohlhammer, J., & Melancon, G. (2008). *Visual analytics: Definition, process, and challenges*. Berlin, Heidelberg: Springer.

Keim, D., Mansmann, F., Schneidewind, J., & Ziegler, H. (2006). Challenges in visual data analysis. In *Tenth international conference on information visualization*. IEEE.

Kosara, R., Hauser, H., & Gresh, D. L. (2003). An interaction view on information visualization. In *State-of-the-Art Report. Proceedings of EUROGRAPHICS*

Pfitzner, D., Hobbs, V., & Powers, D. (2003). A unified taxonomic framework for information visualization. In *Proceedings of the Asia-Pacific symposium on information visualisation (Vol. 24)*. Australian Computer Society, Inc.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., & Van Dantzich, M. (1998). Data mountain: Using spatial memory for document management. In *Proceedings of the 11th annual ACM symposium on user interface software and technology*. ACM.

Sebrechts, M. M., Cugini, J. V., Laskowski, S. J., Vasilakis, J., & Miller, M. S. (1999). Visualization of search results: A comparative evaluation of text, 2D, and 3D interfaces. In *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval*. ACM.

Shneiderman, B. (2003). Why not make interfaces better than 3D reality? *Computer Graphics and Applications*, *23*, 12–15.

Wiza, W., Walczak, K., & Cellery, W. (2004). Periscope: A system for adaptive 3D visualization of search results. In *Proceedings of the ninth international conference on 3D web technology*. ACM.

Hilary Grierson studied Architecture and Design at the University of Strathclyde, Glasgow, and at the Graduate School of Design, Harvard University. She now works at DMEM, University of Strathclyde and has published over 40 journal and conference papers, on global design education, information management and creative design methods. Dr Grierson worked on the DIDET project with the Center for Design Research at Stanford University, researching into global design education. More recently she has worked on the EPSRC funded DAISY (DAta Intensive SYstems) project—investigating the visualization and handling of large secure data sets for defence industries.

Jonathan Corney is a Professor of 'Design and Manufacture' at the University of Strathclyde. Prior to this appointment he was in the Mechanical Engineering Department at Heriot-Watt University and Edinburgh University's Artificial Intelligence Department. His research interests centred around mechanical CAD/CAM include: feature recognition, 3D content based retrieval, cloud manufacturing and user interface design. He is also active in various areas of advance manufacturing (cold forming processes) and design (high pressure pumps and prosthetic covers). Within the University he is involved in the management of the department's EngD centre, the Weir Advance Research Centre and Scottish Institute for Remanufacturing.

Gillian Hatcher studied Product Design and Innovation at the University of Strathclyde, Glasgow. Her research interests include sustainable design and manufacture, and her PhD research focused on product design for remanufacture. She has also worked on the GREENET (Globally Recoverable and Eco-friendly E-equipment Network) Project, investigating the remanufacture of electrical and electronic equipment. Dr. Hatcher is currently working in communications in the environmental conservation sector.