Visual-Spatial Exploration of Thematic Spaces: A Comparative Study of Three Visualisation Models

Timothy Cribbin[1], Chaomei Chen

Vivid Research Centre, Department of Information Systems and Computing, Brunel University, Uxbridge, UK. UB8 3PH.

ABSTRACT

Scatter graphs (e.g. SPIRE Galaxies, Bead, VR-VIBE) are a popular medium for visualising spatial-semantic structures derived from abstract information spaces. For small spaces (i.e. less than one hundred nodes), such graphs can be an effective means of reducing high-dimensional information into two or three spatial dimensions. As dimensionality increases, representing the thematic diversity of documents using spatial proximity alone becomes less and less effective. This paper reports an experiment designed to determine whether, for larger spaces, benefits are to be gained from adding visual links between document nodes as an additional means of representing the most important semantic relationships. Two well known algorithms, minimum spanning trees (MST) and pathfinder associative networks (PFNET), were tested against both a scatter graph visualisation, derived from factor analysis, and a traditional list-based hypertext interface. It was hypothesised that visual links would facilitate users' comprehension of the information space with corresponding gains in information seeking performance. Navigation performance and user impressions were analysed across a range of different search tasks. Results indicate both significant performance gains and more positive user feedback for MST and PFNET visualisations over scatter graphs. Performance on all visualisations was generally poorer and never better than that achieved on the text list interface, although the magnitude of these differences was found to be highly task dependent.

Keywords: information visualization, information retrieval, evaluation, pathfinder associative networks, minimum spanning trees

1. INTRODUCTION

The explosion in on-line storage and communications is providing more and more users with access to diverse information resources. Search engines and web directories provide 'a way in' to this vast store of information, but traditional hypertext interfaces are clearly becoming inadequate for the kinds of search tasks that many users now wish to perform. Although many engines employ quite sophisticated retrieval algorithms, results are still generally displayed in a one-dimensional format. For simple queries, this may adequate, however for more complex or ill-defined search tasks, however, this format is often counter-intuitive. This is because the thematic structure of even at few hundred items is likely to be highly diverse. Whilst ordered in terms of query similarity (the sense of which the computer does not understand), it is highly likely that items conforming to the intended sense will be distributed almost arbitrarily across the list. Over the last decade there has been increasing interest in the potential of Information Visualisation as a means of facilitating the comprehension and retrieval of textual information. Visual Information Retrieval Interfaces (VIRI), particularly those that use virtual environments, provide a means of both visualising and directly interacting with large collections of, not only conventional hypermedia, but also unstructured text.

It has long been known that spatial ability is a good predictor of search skills[6]. The ease with which users are able to comprehend and learn the structure of the information 'space' seems to be a critical factor in determining the usability of a retrieval system. It is thought that visualisations help the user by moving part of the mental load from cognitive to perceptual processes, which are much faster and less resource hungry[13]. They may, therefore, be particularly beneficial for users who have poor spatial skills[12, 17,18] and/or are unfamiliar with the topic domain[11].

[1] Correspondence: Email: Timothy.Cribbin@brunel.ac.uk; WWW: http://www.cribbin.co.uk/infovis.htm

Many systems have been or are currently in development (see [1, 9] for reviews) and a wide range of metaphors and interaction methods have been implemented. This is no bad thing, as most have been developed with particular contexts in mind and it is unlikely that a single design will become an optimum solution for all tasks or indeed all users. There is, however, still a strong technology bias within VIRI development and, in general, the relative benefits of different approaches are still unknown.

Existing studies have compared either multiple VIRIs[24], variations of the same interface[16, 22] or specific VIRIs with traditional interfaces[4, 16,19,21.] This study aimed to combine these paradigms into a single experiment that compared the utility of multiple visualisations alongside a traditional text based interface using a range of information retrieval tasks. Like Wiss and Carr[24], Shneiderman's[15] visual information search taxonomy was used to define task types.

1.1. Graph Drawing Algorithms

The design of a VIRI can significantly effect users' performance on and liking of it. Wiss and Carr[24] found strong effects of visualisation type when searching directory structures. Even dimensionality or subtle changes in fidelity of the structure can impact heavily on performance and workload [22].

Whilst approaches to information visualisation are diverse, at a fundamental level it is possible to identify a number of common elements, particularly in terms of graphing techniques used to represent information spaces to the user.

Perhaps the one of the most common is the scatter graph (e.g. Bead [2], SENTINEL [8], SPIRE [23]). These graphs are produced in two- or three-dimensions with documents being represented by visual objects or nodes. Similarity between nodes is implied by their spatial proximity.

For abstract textual information spaces, some form of text analysis is applied to the information space in order to extract the salient semantic structure. This analysis typically generates a high dimensional semantic 'space', represented in the form of matrix of inter-document semantic similarity coefficients.

In order to represent this structure visually, dimensionality must be reduced. This is done using statistical methods that are able to produce an approximation of the original high-dimensional space within two or three dimensions. There are a variety of methods available for this purpose, for example Multidimensional Scaling and Factor Analysis. Whilst working on different principles, they all have basically the same objective. As a consequence the graphs produced usually appear very similar.

Whilst these approaches are adequate for smaller spaces, as the number of documents increase, so too does the conceptual diversity of the space. It becomes practically impossible to usefully approximate high dimensional spaces within two or three spatial dimensions. Scatter graphs containing more than 100 nodes often seem amorphous in appearance with a tendency for many nodes to cluster around the origin point [3]. One option is to apply links between document nodes as a means of disambiguating this structure.

Representing all semantic relationships using links can produce a cluttered display, particularly in large information spaces, with a confusing array of crossing links. Minimum spanning trees (MST) get round this problem by preserving only the most critical or informative links. Given a network of nodes, the link between each pair of nodes has a weight. The algorithm typically begins by ranking all the links in weight order (e.g. semantic proximity). Starting with the lowest weights the algorithm works through the list, deciding in each case whether to accept or reject the link. Links are rejected if an alternative path between the two nodes is already available. The end result is that each node is linked to at least one other node but the total links is always equal to N-1, the bare minimum for a tree graph.

The problem with MST is that because some links will have equal weights, the order in which these tied ranks are placed will often influence the final structure. For this reason, there is often more than one possible MST for a given network. A minimum cost pathfinder associative network (PFNET) takes this into account and can essentially be described as the union combination of all possible MSTs [14]. What this means is that, unlike MST, two nodes can be connected by more than one link, but only if they all posses the same weight. This is known as the triangle inequality rule. Hence PFNETs are similar in appearance to MSTs but tend to have slightly more than N-1 links.

In this study, three different algorithms were used to create visualisations. These were Principal Components Analysis (PCA), MST and PFNET. Figure 1 shows examples of visualisations that these algorithms produce. Each was created using the same proximity data, a Latent Semantic Indexing (LSI) solution based on 200 documents [5]. Node co-ordinates for PCA were derived from loadings on the top two factors. MST and PFNET appear almost identical. Only close inspection reveals the expected additional links in the form of closed loops.

1.2. Hypotheses

It was hypothesised that the links provided by MST would serve to disambiguate the spatial-semantic structure, leading to an overall improvement in semantic comprehension, retrieval performance and, in turn, user satisfaction when compared to the PCA visualisation. Furthermore, additional links provided by PFNET would enhance this effect still further.

Studies comparing VIRIs to more traditional interfaces have found a variety of results. Some have found performance benefits [18,21] whilst others have either found no appreciable difference [16, 19] or even deficits [4]. Our hypotheses are as follows:

H1: MST and PFNET will result in more efficient navigation and more effective document retrieval than the linkless PCA graph interface.

H2: Users will develop richer mental models of an information space when using MST and PFNET as opposed to PCA.

H3: VIRIs, especially MST and PFNET, will become relatively more useful in comparison to text as users progress through the structured tasks.

a) PCA b) MST c) PFNET

Figure 1. The 'Alcohol' topical space as visualised by VIRIs used in the study.

2. METHOD

2.1. Development of the testbed

2.1.1. Data sets

A testbed of four document sets was developed, each comprising short news articles retrieved from the Los Angeles Times newspaper archives stored on the Text Retrieval Conference (TREC) vol.5 CD-ROM. In each case, articles were retrieved based on a single keyword query. In order to minimise the effects of reading time during tasks, only articles between 250 and 750 words in length were retained. Of the remaining document pools, in each case the top 200 hits, based on keyword frequency, were selected. Keywords used were "alcohol", "endangered", "gaming" and "storm".

2.1.2. Tasks

It was intended that the tasks should reflect a variety of search activities. It is becoming an increasingly common empirical approach [16] and the current TREC-9 Interactive Track to present users with a range of task types in order to gain a more generalisable impression of the capabilities of a retrieval engine or interface. We chose Shneiderman's [15] taxonomy of

visual information search as a framework for the design of our tasks. His mantra of overview, zoom, filter, details on demand describes the typical [2]stages of user activity that can be observed during visual information retrieval. These task types were translated into four experimental task definitions (Figure 3). Task A, conforms to the overview type as the user is simply orienting themselves within the space by identifying a general topic area. The user is required to locate and mark documents relating to a given topic that is defined by a relatively broad set of criteria. In the 'alcohol' example this was "Locate and mark any documents that mention an incident of drink driving". Documents relevant to this topic numbered between 20-25 or 10-13% of all items in information space. Task B is simply a refinement of the criteria from the previous task. For example "…an incident where a death has occurred as result of drink driving and the blood-alcohol level of the driver is described". Relevant documents form a sub-set of those relevant to task A (6-10 documents). Hence the user must zoom in to more specific regions of the information space, whilst filtering out documents that are no longer relevant. Task C introduces further criteria refinement by asking the user to extract a thread of two directly related documents from the task B domain. In the 'alcohol' space these were two articles describing different chronological stages of the same court trial.

Tasks A to C constitute the structured tasks, as the user is guided gently from a general exploratory browse into the final directed search task by progressively more specific criteria. Task D, the unstructured task, like C is a thread, but also involves a complete change of track (new topic). In the case of alcohol, this thread comprised letters to the editor regarding a book that was banned from schools because of references to recreational alcohol consumption. This task is more difficult than C because the user must reorientate themselves within the space without the benefit of progressive query refinement of the kind offered in tasks A and B. Table 1 summarises the definitions of each task type. The full range of tasks was developed for each data set.

Figure 2. Document browser software.

Task A (structured) Topic – general theme in the information space (20-25 documents)
Task B (structured) Sub-topic – a more specific aspect of the topic (6-10 documents)
Task C (structured) Thread – pair of articles, within the sub-topic, that discuss the same or two directly related incidents.
Task D (unstructured) Thread – pair of articles, from a different topic, that discuss the same or two directly related incidents.

Figure 3. Overview of tasks

2.1.3. Software

Inter-document similarity matrices were computed for each data set using Latent Semantic Indexing [5]. Visualisations were then created in VRML and displayed using the CosmoPlayer™ browser plug-in. Each node (sphere) in the space represents a single document (Figure 2.). For the control condition nodes were represented as list of hypertext titles, in their original rank order. In all conditions, moving the mouse over a node causes the associated document title to be displayed both in a proximal ToolTip and also in a text box at the top of the screen. Clicking once on a node brings up that document in the

[TM]Visual Basic, Internet Explorer and ActiveX are trademarks of Microsoft Corporation. Cosmoplayer is a trademark of Cosmo Software.

adjacent frame. Current documents are marked by pressing control key. A Visual Basic™ program was written to present the tasks and to record a log of user activity. The VRML/HTML component was integrated into the VB program by means of the Internet Explorer™ ActiveX control.

2.2. Experimental Design

A two-way within-subjects (Interface x Task) design was used. Interface conditions comprised three VIRIs based on MST, PFNET and Principal Components Analysis (PCA). In addition a conventional text list was included as a control condition (TEXT). A block of four associated search tasks was devised, according to the criteria in section 2.1.2., for each of the four data sets. Each interface could therefore be presented with any of the four datsets and task blocks.

2.3. Participants

The sample was drawn from the Brunel University student population. Participants (P) voluntarily applied to take part following a poster and email advertising campaign. Sixteen Ps (9 male, 7 female) completed all four blocks. Median age range was 25-29 years. Each also completed a brief questionnaire regarding their computing experience. Results indicated a self-selection bias meaning that the majority of Ps fell in the upper range of both on-line and general computing experience.

2.4. Procedure

Ps attended for the computer tasks in groups of between one and three. After a brief verbal introduction to the tasks Ps were presented with a practice task in order to both familiarise them with the concept of VIRIs and the CosmoPlayer software. This task comprised an information space composed of the top ten articles from each of three queries – Tourism, Hubble and Suicide. The graph coordinates were produced using PCA from LSI data resulting in a simple space composed of three tight clusters. Ps were instructed to browse each cluster in turn and to write down in one or more words what they believed to be the dominant topic that was causing articles to group together.

Each condition block ran automatically from beginning to end. At the beginning of each block, Ps were presented with a brief description of the nature of the information space including the number of documents and the search term used to retrieve them. Tasks were always presented in the same sequence (A, B, C, D). At the start of each task, instructions were presented comprising document relevance criteria and the number of relevant documents within the space. Ps were asked to read these instructions carefully before clicking on the START button. Once P clicked on the START button the information space (see figure 2) they had five minutes in which to complete the task. At all times they were able to toggle back and forth between the navigation and instruction screens.

Once five minutes had passed, the task ended immediately and post-task user was feedback elicited (12 items). Responses to the usability items were all based on a seven point scale (1=Disagree Strongly, 7=Agree Strongly). A new instruction screen was then presented for the next task.

At the end of the block a series of 29 more detailed usability items were presented. These asked questions about utility and acceptability of the interface along with information space comprehension.

2.5. Dependent measures

Performance was analysed from two perspectives. Firstly, overall task success was measured in terms of retrieval performance or relevant documents marked. Recall and precision indices were combined into the well known F-measure[20]. This metric, Fmk, represents the intersection between the relevant and the retrieved document pools and has the attractive property whereby high scores are only possible when both recall and precision are high. The formula is detailed in Figure 4. In addition time taken to retrieve the first relevant document was also taken into account and will be referred to as FRMK.

The second perspective focuses on navigational performance and takes into account all nodes that were at least visited (title read). The F-measure was used again, and took into account documents visited (mouse over node events). This will serve as a general measure of navigational efficiency and effectiveness and will be referred to as Fno. A lostness measure, LOST, was also included. This again uses the same combination formula but the two variables represent the number of relevant events as a proportion of all events and the degree of backtracking over non-relevant nodes.

Construct validity for LOST as a measure of lostness was established by post-hoc analysis. For all interfaces and tasks (4x4), the items "I often felt lost", "I often felt disorientated", "I often felt frustrated" and "It took too long to find what I wanted" were correlated with their associated LOST variable. The mean correlations, by usability item, ranged from -.31 to -.41 (n=16), with t-scores showing that all these means differed significantly from zero (df=15, p<.01). We would expect negative correlations, given that a low LOST score indicates that the user is lost and high score on the usability items indicates agreement

(a) $F_{pr} = \frac{2pr}{p+r}$ (b) $NAV = \frac{2(D/T) \bullet (R/E)}{D/T + R/E}$

a) F_{PR} measure

P=Precision
R=Recall

b) LOST measure

D=different non-relevant nodes visited
T=total non-relevant nodes visited
R=number of relevant events logged
E=total events logged.

Figure 4. Formulae for objective performance measures used in the analysis

3. RESULTS

All sixteen participants completed all four conditions. Due to a software problem, at least one task failed during administration of the TEXT condition for three cases. For this reason, where TEXT is included in the analysis, N=13.

3.1. Task Effects

There was a significant linear effect of task on all measures. In fact, task effects were much more influential than those of interface. This was particularly true for the navigation measures (LOST, F1,9)=21.01, p=.001; Fno, F(1,9)=32.73, p<.001). This confirms the validity of the experimental design where tasks were intended to become progressively more difficult. There were no interactions between interface and task.

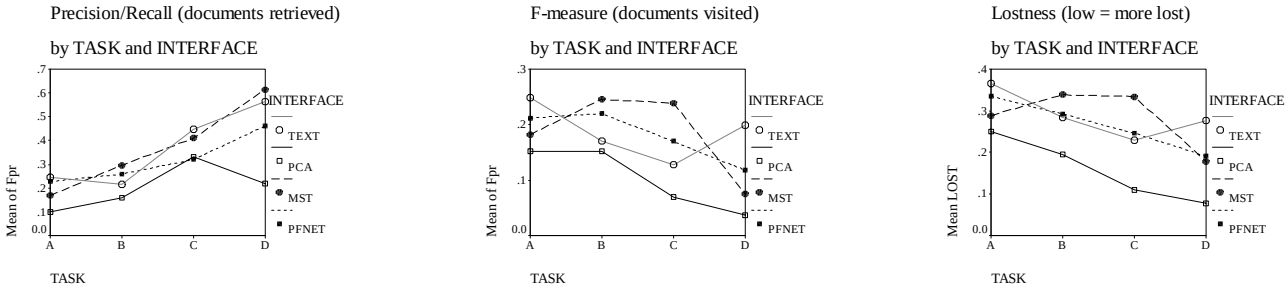
3.2. Differences in retrieval and navigation between visualisations

For all performance measures the trend was for similar scores on both MST and PFNET visualisations and consistently poorer performance when using the PCA based VIRI (Table 1). The effect was most clear for the navigation measures, Fno and LOST, with significant differences between PCA and both MST and PFNET. The interface effect was less powerful in terms of overall retrieval performance. Nevertheless, the general trend remained the same.

Dependent measure	**PCA**	**MST**		**PFNET**	
	Mean	**Mean**	**F(1,15)**	**Mean**	**F(1,15)**
LOST	.16	.26	6.39*	.26	4.37*
Fno	.10	.17	5.59*	.18	5.58*
FRMK (seconds)	216.9	201.2	.895^{ns}	167.4	4.50*
Fmk	.21	.36	5.91*	.35	2.37^{ns}

Table 1: Comparison of mean performance scores on MST/PFNET vs PCA. * p (2-tailed) <.05, ns: not significant

An examination of graphs (a) to (c) in figure 5 show a further breakdown of interface effects by task. Figures 5b and 5c show the two navigation (nodes visited) measures. One can see that the gap between PCA and the minimum cost networks is at a maximum during task C, the final stage of the structured tasks. In terms of precision/recall of documents retrieved, the biggest gap seems to occur on task D, where participants must find a 'thread' based on a new sub-topic. When using MST/PFNET participants were scoring around 0.5, compared with an average score of only 0.22 on PCA.



(a) Precision/Recall, documents retrieved (b) Precision/Recall, documents visited (c) Lostness when navigating

Figure 5. Retrieval and Navigational Performance across tasks and by interface

3.2.1. Comprehension of spatial-semantic structure

After each task participants responded to brief statements regarding their personal experience and their perceived success. Generally, participants reported feeling less disorientated, less frustrated and less confused when using the MST or PFNET visualisations, in comparison to PCA, p(1-tailed)<.05. They also felt they had completed the tasks more successfully when using the minimum cost networks. Furthermore, participants gave lower levels of agreement in response to the items "It took too long to find what I wanted" and "I needed more time to complete the task", but this was only significant for MST.

At the end of each task block, participants responded to a number of more detailed statements relating to usability. With respect to "Much of the time I had no idea where I was going next", participants were more likely to agree when using the PCA interface, than when using either MST, F(1,15)=3.14, p(1-tailed)<.05 or PFNET, F(1,15)=3.99, p=.03. Participants found the PCA interface more cluttered than MST, F(1,15)=3.00, p(1-tailed)=.05 and PFNET, F(1,15)=2.62, p(1-tailed)=.06. They also agreed to a greater extent that documents were clustering into recognisable themes on the MST interface, but PFNET was not significantly different from PCA in this respect.

In support of H2, we therefore found that participants seemed to report a slightly greater level of comprehension when using MST and PFNET. That said, there were other related items such as "The layout of the information space made good sense" or "I found it easy to find my way back to articles I'd visited earlier", where no statistical differences were found between interface conditions.

3.3. Visual vs non-visual interface performance

The textual interface resulted in consistently better performance than that seen with the PCA visualisation. This effect was found for Fmk, F(1,12)=7.50, p=.02, Fno, F(1,12)=11.74, p<.01, and LOST, F(1,12)=13.26, p<.01. Time to mark first relevant document (FRMK) was a different story. Users retrieved their first relevant document faster when using TEXT compared with both PCA, F(1,12)=11.56, p<.01, and MST, F(1,12)=6.29, p=.03. Users were also slightly slower when using PFNET, but not significantly so, F(1,12)=2.135. No main effects were evident between TEXT and either MST or PFNET for the other three measures.

3.3.1. Task effects

H3 hypothesised that VIRIs may be relatively more useful as users progressed through the structured tasks i.e. zoomed into and became familiar with a particular topical area. One can see from the graphs in Figure 5 that the performance differentials between MST, PFNET and TEXT were generally small, with PCA remaining inferior to all three other interfaces. There are no signs of any task * interface interactions when it comes to documents retrieved. On the navigation

measures, Fmk and LOST, there seems to be a trend for MST and PFNET to become more useful, relative to TEXT, for tasks B and C.

To explore this effect further, ratios of relative performance were computed, by task, between text and mean score on the VIRI conditions. Means can be seen in table 2, with scores above 1 indicating the text interface enabled superior performance to the VIRIs. Single sample t-tests were computed for each cell, using a comparison value of 1. An asterisk indicates a significant t value. The left half of the table shows ratios computed where all VIRI conditions were taken into account. On the right, the PCA data was excluded, leaving only the minimum cost networks, MST and PFNET.

Considering all VIRI conditions, the clearest differences can be seen on task A, with the traditional text interface facilitating superior performance on all measures. These differences only approached significance, however, in all cases. As expected the ratios drop for all measures except FRMK, with no significant or near significant differences occurring. One can see that FRMK remains consistently lower for the TEXT condition across all tasks.

Come the track change on task D and TEXT regains the advantage and resulted in more precise complete navigation of the relevant document set. Participants also showed less evidence of being lost, although this only approached significance.

Many of the differences described can be attributed to the effect of the PCA condition. When this data was excluded from the ratios, for example, no significant differences occurred on any tasks for Fmk, Fno and LOST. In line with H3, the ratios frequently dropped below 1 on tasks B and C, particularly for the navigation measures (Fno, LOST). These differences were never significant, however. No changes were evident in the pattern for FRMK.

So, as predicted by H3, it seems the relative value of VIRIs improves somewhat as users become more familiar with the nature of a topical locality. That said, users were still taking much longer to begin finding relevant documents, even on the latter structured tasks (B & C). Some of this difference disappeared when the PCA data was excluded, but orientation time remains a problem area even when using these VIRIs.

Ratio		**TEXT : Mean of all VIRI conditions**				**TEXT : Mean of MST and PFNET**			
Task		**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Fmk	*Mean*	1.08[as]	.99	1.09	1.16	1.05	.97	1.10	1.13
	SD	.15	.15	.32	.42	.15	.20	.35	.48
FRMK	*Mean*	3.58 [as]	2.03*	6.01	2.68*	3.86[as]	1.66*	5.13	2.23*
	SD	4.53	1.47	10.28	2.33	5.63	.90	8.56	1.63
Fno	*Mean*	1.06 [as]	.97	.98	1.12*	1.05	.96	.95	1.10 [as]
	SD	.11	.09	.16	.18	.12	.11	.18	.18
LOST	*Mean*	1.07 [as]	1.01	.99	1.12 [as]	1.06	.99	.97	1.09
	SD	.12	.15	.18	.20	.14	.18	.19	.21

Table 2. Ratios of VIRI:TEXT performance. *Value of >1 indicates better performance on TEXT. * p< .05, [as] approached significance.*

3.3.2. General user impressions

Post-block feedback given by participants provided a few interesting insights into both the relative value and shortcomings of the VIRIs used here, when compared to using a more traditional interface. Items that returned as significant main effect of interface are detailed in Table 3.

The first item shows that participants tended to find themselves losing their place in the space much more so when using the VIRIs. The problem seemed to be that users would click on a node, read the document then be unable to re-find the node in question (e.g. for marking). This problem occurred to a similar extent on all of the visual interfaces. Nevertheless, participants noticed that documents were tending to group thematically and they agreed quite strongly that logic of the layout became more apparent the more they explored it. For MST and PFNET this agreement was significantly stronger than that reported for TEXT.

It is interesting to see that PCA was perceived as the most cluttered/unclear interface. MST and PFNET were seen as slightly more cluttered than TEXT but not significantly so. Participants also reported finding the text much harder to read when using the VIRIs. One can assume they are referring to the title text as the format of full-text presentation was identical for all conditions. On the VIRIs, titles appeared as ToolTip labels only when the mouse pointer was held over a node. Furthermore the font was slightly smaller than that used in the TEXT list.

In terms of interaction, users reported finding the controls harder to use for the VIRIs. This was not surprising given most reported having no prior experience using CosmoPlayer.

	TEXT	**PCA**	**MST**	**PFNET**	**F(3,45)**
I kept losing my place (in the space)	4.0	5.8**	5.5*	5.5**	6.192**
The longer I used it, the more the layout of the space started to make sense	3.6	4.6	5.3*	4.9*	2.536[as]
It seemed like articles were clustering into recognisable themes or concepts	2.4	3.8*	4.8**	3.6[as]	5.049**
I found the controls fiddly	2.7	3.9*	3.4	3.8*	3.108*
The text was easy to read	5.4	4.4*	4.8	4.5[as]	2.299[as]
The display was clear and uncluttered	4.2	2.9*	3.6	3.5	2.600[as]

Table 3. Response means from post-block usability items. *1=Disagree strongly, 7=Agree strongly. * p<.05, **p<.01, [as] approached significance*

4. DISCUSSION

4.1. Value of minimum cost networks

Overall, the data presented here indicate that there are benefits to be gained from implementing an economical set of visual links between nodes to highlight the most important semantic relationships within a document set. Performance figures show significant gains in terms of both document retrieval and the ease with which users were able to remain focused on the relevant thematic areas. There was also some evidence on the subjective level that users found MST and PFNET visualisations more intuitive and useful. Although significant differences did not occur for all items relating to comprehension, there was never any evidence to suggest a negative impact of visual links.

4.2. Comparison with the text list interface

It seems apparent that VIRIs are not immediately more useful than a traditional text list interface. Even taking different tasks into account we found that, at best, MST and PFNET visualisations enabled performance to a similar level as that seen using the text list. That said, there was a weak trend for VIRIs to become relatively more useful in the more advanced stages of the zoom and filter process.

4.3. Improving VIRI usability

The purpose of this study, however, was to test the simple effect of graph style on the usefulness of a spatial-semantic structure. As such these test VIRIs were intentionally poor of navigational aids. Because of this, certain actions were made unnecessarily difficult. For example, there was a strong tendency for VIRI users to lose their place in the space as soon as they diverted their gaze from the visualisation. Such a problem would be easily remedied by providing a visual highlight

over the current document. Furthermore, static visual cues to show which documents have been marked, read or recently visited would no doubt make a tremendous positive impact on usability.

The data indicates that the greatest difficulties occurred during tasks that required orientation to a new sub-topic. This is unsurprising given no indication was given to the user regarding where they should begin their exploration. This factor probably accounts for most of the observed differences between the VIRIs and TEXT in terms of time taken to find and mark the first relevant document. Relevant documents were distributed arbitrarily through the text list so the probability of discovery early on was much higher. In contrast, the benefits of a spatial-semantic structure only become available once the thematic 'hot-spot' is located. In some cases, this could use up almost all of the allotted task time. Much of this burden could be removed by providing appropriate context labelling. Many algorithms are currently in development to enable auto-summarisation of salient categories within information spaces. Such a feature would enable the user to quickly appreciate the high level thematic structure and would considerably narrow the choice of starting point(s).

Allowing further refinement of the initial query would also accelerate orientation, particularly during the later tasks. Such a feature would provide the facility to gradually filter out more and more irrelevant documents as the user becomes more directed and confident in their search strategy. Indeed, many of our participants suggested that a within-results query search facility would have been highly useful on the VIRIs. An intelligent algorithm could also be implemented to suggest alternative query terms to the user based on their previous behaviour. This could be achieved for example, by extracting the key terms from documents that the user has indicated most interest in.

As the Internet continues to grow, information visualisation has the potential to empower the currently overloaded information seeker. Before this can be achieved researchers must identify all the important sources of cognitive load that occur when navigating an abstract, virtual information space. This is no easy goal given, as seen here, the powerful effect of task type alone. Nevertheless the results identified in this paper provide a useful guide for future VIRI design and evaluation studies.

5. CONCLUSIONS

In summary, our data show clear benefits of providing visual-semantic links between document nodes portrayed within an spatial-semantic visualisation. There were no significant benefits, in terms of either retrieval or general navigational performance, to be gained from using visual interfaces, as opposed to a standard text list interface, on any of the tasks. It is hypothesised that the implementation of certain navigational cues into these interfaces will significantly and positively impact on the support provided by visualisation to information search tasks.

ACKNOWLEDGEMENTS

This study was supported by a grant provided by the Library and Information Commission.

REFERENCES

1. Card, S., Mackinlay, J. & Shneiderman, B. *Readings in Information Visualization: Using vision to think.* Morgan-Kaufman, 1999.
2. Chalmers, M. & Chitson, P., "Bead: Explorations in information visualization," *15th Annual International SIGIR '92*, 330-337, 1992
3. Chen, C., "Visualising semantic spaces and author co-citation networks in digital libraries," *Information Processing and Management*, **35**, 401-420, 1999.
4. Cockburn, A. & McKenzie, B., "An evaluation of cone trees," *In People and Computers XV. Proceedings of the 2000 British Computer Society Conference on Human-Computer Interaction*, University of Sunderland, 4--8 September, 2000.
5. Deerwester, S., Dumais, S., Landauer, T., Furnas, G. & Harshman, R., "Indexing by latent semantic analysis," *Journal of the American Society for Information Science*, *41*(6), 391-407,1990.
6. Egan, DE. (1988) Individual Differences in Human-Computer Interaction. In M. Helander (ed*.) Handbook of Human-Computer Interaction*, Elsevier (North-Holland), 1988.

7. Fekete, J. & Plaisant, C., "Excentric labeling: dynamic neighborhood labeling for data visualization," *CHI '99*, 15-20 May 1999, 512-519, 1999.
8. Fox, KL., Frieder, O., Knepper, MM. & Snowberg, EJ., "SENTINEL: A multiple engine information retrieval and visualization system," *Journal of the American Society for Information Science*, **50(7)**, 616-625, 1999.
9. Herman, I., Melancon, G. & Marshall, M.S., "Graph visualization and navigation in information visualization: A survey," *IEEE TVCG*, **6(1)**, 24-43, 2000.
10. Lin, X., "A self-organizing semantic map for information retreival," *Proceedings of SIGIR '91* (pp. 262-269)(Chicago, IL, 13-16 October 1991), ACM Press, 1991.
11. McDonald, S. & Stevenson, RJ., "Navigation in hyperspace: An evaluation of navigational tools and subject matter expertise on browsing and information retrieval in hypertext," *Interacting with Computers*, **10**, 129-142, 1998.
12. Paez, LB., daSilva, JB, Marchionini, G., "Disorientation in electronic environments: A study of hypertext and continuous zooming interfaces," *Proceedings of the Asis Annual Meeting*, **33**, 58-66, 1996.
13. Robertson, GG., Mackinlay, JD. & Card, SK., "Cone Trees: Animated 3D visualizations of hierarchical information," *In Proceedings of the ACM SIGCHI'91, Conference on Human Factors in Computing Systems*, pp.189-194, 1991.
14. Schvaneveldt, R. W. (Ed.). *Pathfinder Associative Networks: Studies in Knowledge Organization.* Norwood, NJ: Ablex, 1990.
15. Shneiderman, B., "The eyes have it: A task by data type taxonomy for information visualisation," *Proceedings of IEEE Conference on Visual Languages*, (Boulder, CO, Sept. 3-6, 1996) 336-343, 1996.
16. Sebrechts, MM., Vasilakis, J., Miller, MS., Cugini, JV. & Laskowski, SJ., "Visualization of Search Results: A Comparative Evaluation of Text, 2D and 3D Interfaces," *SIGIR '99*, 3-10, 1999.
17. Sein, M.K., Olfman, L., Bostrom, R.P. & Davis, S.A., "Visualization ability as a predictor of user learning success," *International Journal of Man-Machine studies*, **39**, 599-620, 1993.
18. Stanney, KM. & Salvendy, G., "Information Visualization: Assisting low-spatial individuals with information access tasks through the use of visual mediators," *Ergonomics*, **38(6)**, 1184-1198, 1995.
19. Swan, RC & Allan, J., "Aspect Windows, 3-D Visualizations and indirect comparisons of Information Retrieval systems," *Proceedings of SIGIR '98* (pp. 173-181)(Melbourne, Australia), ACM Press, 1998.
20. Van Rijsbergen, C.J. *Information Retrieval.* Butterworth & Co., Boston, MA, 1979.
21. Veerasamy, A. & Heikes, R., "Effectiveness of a graphical display of retrieval results," *Proceedings of SIGIR 97*, pp 236-245 (Philadelphia, USA), 1997.
22. Westerman, S.J. & Cribbin, T., "Mapping semantic information in virtual space: Dimensions, variance, and individual differences," *International Journal of Human-Computer Studies*, **53(5)**, 765-788, 2000.
23. Wise Jr, JA., Thomas, JJ., Pennock, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V., "Visualising the non-visual: Spatial analysis and interaction with information from text documents," *In Proceedings of InfoVis '95, IEEE Symposium on Information Visualization*, New York, 51-58, 1995.
24. Wiss, U & Carr, D., "An empirical study of task support in 3D Information Visualisation," *IEEE International Conference on Information Visualisation*, London, 392-399, 1999.