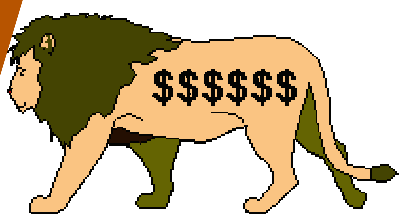
VISUA... ...NG IN

...etrieval and how one may use ...ioural patterns of information ...lysis of search strategies.

...Card., 1995), searching for information is, in many ...n strategy between the two focuses on the trade-off ...hieving the goal at hand (Figure 1). When we search ...if we are using a modem and the size of the file to be




**...uses on risks as well as the profitability of a target.**

...n foraging process is therefore similar to the analysis of a traditional ...ars ago. It is the judgement of the profitability versus the risk that ..., and other behavioural patterns. When we deal with an information ...ocument in terms of its intrinsic values, such as relevancy to our tasks at ...ng and retrieval a document in terms of our subscription fees, per item ...r Internet connections.

...trieval process as information foraging, we may gain additional insights ...haviour patterns in the context of spatially and semantically organised ...n estimate the profitability with the proportion of relevant documents in a ...ded by the time it will take to read all the documents within this area. In their ...irolli and Card found that even a simplified model of information foraging ...be influenced. Their study suggests that users are likely to search widely in an information space if the query is simple, and more focused if the query is harder (1995). According to the profitability principle, harder queries entail higher cost to resolve and the profitability of each document is relatively low. In general, users must decide whether or not to pursue a given document on the course of navigation based on the likelihood profitability of the document.

The optimal information foraging theory provides a useful framework for the modelling and analysis of behavioural patterns involving visual navigation. In particular, users are concerned about how to maximise the potential profitability, where to search, and how long to search in a specific area.

In this article, we present a conceptual framework for modelling behavioural patterns in visual information foraging. In particular, this visual information foraging takes place in a spatial-semantic interface of a thematic information space. This framework constitutes the optimal information foraging theory, Hidden Markov Models, spatial-semantic interfaces, and a taxonomy of visual navigation. In this article, we focus on the role of information foraging in our approach. A related study focusing on Hidden Markov Models is reported in (Chen, Cribbin, Kuljis, & Macredie, 2000). The overall approach is illustrated through an example in which visual navigation data were drawn from an information retrieval experiment. Finally, implications of this approach for understanding users' navigation strategies are discussed.

## 2. INFORMATION FORAGING IN VISUAL NAVIGATION

In order to study users' information foraging behaviour, we constructed four thematic spaces based on news articles from the Los Angeles Times newspaper retrieved from the Text Retrieval Conference (TREC) test data. Each thematic space contains the top 200 news articles retrieved through a single keyword query to the document collection. The four keywords used were *alcohol*, *endangered*, *game*, and *storm*. Corresponding spaces were named accordingly by these keywords.

The user interface for each thematic space is designed based on a popular organisation principle, which is closely connected to Gestalt laws in cognitive psychology. Gestalt laws are several principles developed in earlier 20th century concerning how people perceive patterns. For example, the proximity principle says that people tend to see natural groupings purely based on proximity relationships displayed (Figure 2). In this article, we focus on searching behaviours associated with spatial-semantic interfaces designed based on relevant Gestalt laws. We expect that users' search behaviours should reflect this spatial-semantic feature.

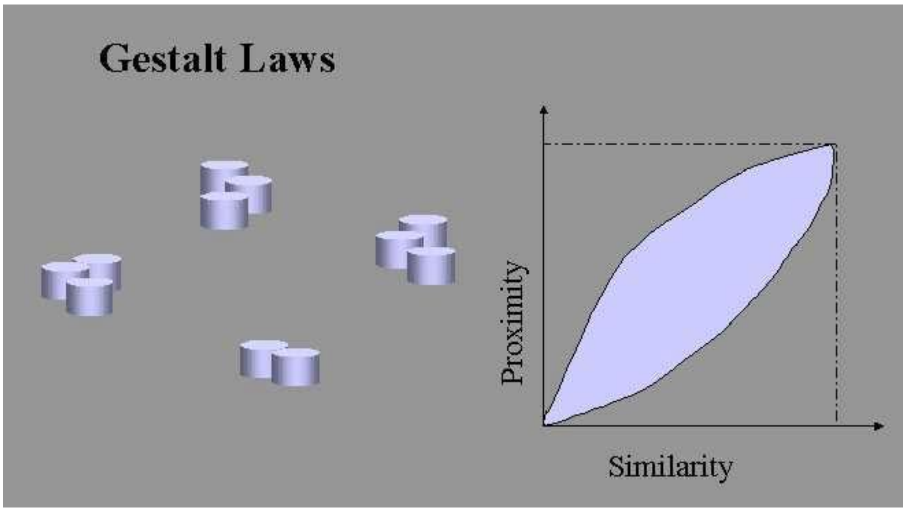



**Figure 2. A spatial-semantic mapping exploits the Gestalt law of proximity.**

For each thematic space, we generated a document-to-document similarity matrix using Latent Semantic Indexing (LSI) (Chen & Czerwinski, 1998; Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990). A few types of spatial-semantic interfaces were produced, including Pathfinder networks (PF) and minimum spanning trees (MST) (see Figure 3). In MST-based visualisation, *N-1* explicit links connect all the documents together. Users can see these links on their screen. In PF-based visualisation, additional explicit links are allowed as long as the triangular inequality condition is satisfied. Detailed descriptions of the use of these techniques for information visualisation can be found in (Chen, 1999).

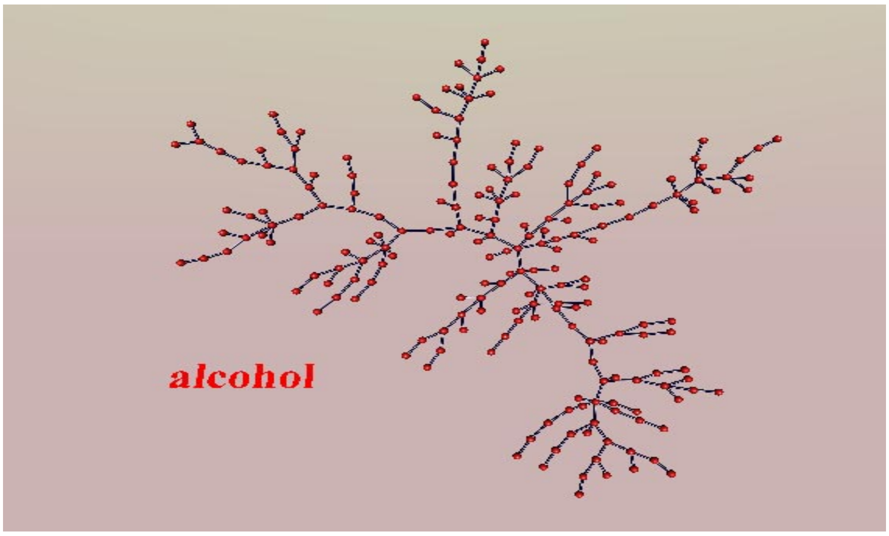



**Figure 3: A spatial-semantic interface of the ALCOHOL thematic space.**

Users in this study are self-selected students at Brunel University. They performed a series of tasks in each thematic space through spatial-semantic interfaces. Usage data was logged to a computer file in each session, including the occurrence of an event, the time stamp of the event, and the target document on which the event takes place.

The design of the tasks follows Shneiderman's mantra: overview, zoom, filter, details on demand, which highlights users' cognitive needs at various strategic stages in visual information retrieval. At the top level with Task A, users need to locate and mark documents relevant to a search topic (Figure 3). For example, users were asked to locate and mark any documents that mention an incident of drink driving. 20~25 documents were judged as relevant by experts for TREC conferences. Task B is more specific than Task A, and so on. We expect that users would narrow down the scope of their search from Task A through Task D and that this should be evident in their trails of navigation.

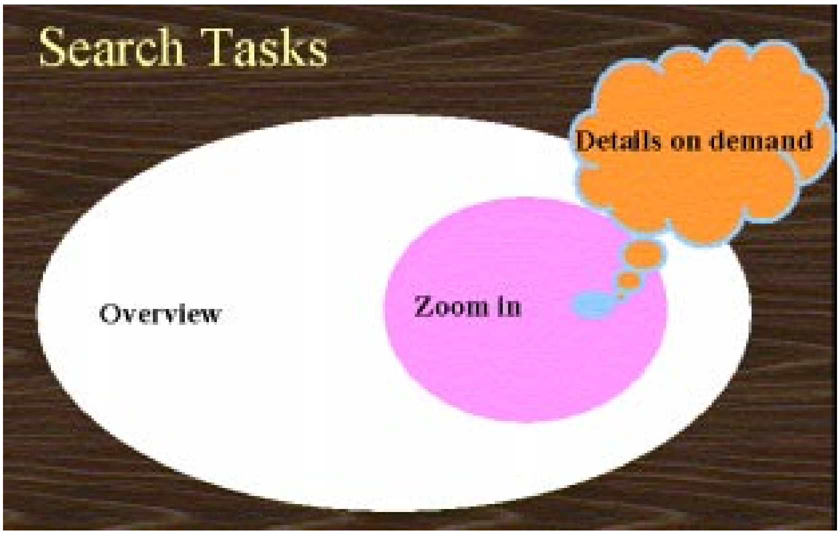



**Figure 4. Shneiderman's mantra is used in the characterisation of interactive tasks.**

The screen display is designed as follows. Documents in red are not relevant to the search tasks. The course of navigation appears as dotted yellow links. Relevancy judgements made by experts are provided in the TREC test data. Documents relevant to the original search are marked with a bright yellow dot in the centre. If the user marks a document as relevant in a search session, this document will be coloured in blue. Upon the user visits a document, a dark circle is drawn around the current document. The time spent on a document is denoted by a growing green belt until the user leaves the document. If the user comes back to a previously visited document, we will see a new layer of dark circle and an additional layer of green belt will start to be drawn. One can choose to carry these discs grown from

one task into the next task and a red disc indicates how long the user has spent on it in the previous task.
Based on the above reasoning, we expect to observe the following patterns concerning users' navigation strategies:

1. Spatial-semantic models may reduce the time spent on examining a cluster of documents if the spatial-semantic mapping preserves the latent semantic structure.
2. Spatial-semantic models may mislead information foragers to over-estimate the profitability of a cluster of documents if the quality of clustering is low.
3. Once users locate a relevant document in a spatial-semantic model, they tend to switch to local search.
4. If we use the radius of disc to denote the time spent on a document, the majority of large discs should fall in the target area in the thematic spaces.
5. Discs of subsequent tasks are likely to be embedded in discs of preceding tasks.

## 4. RESULTS

Figure 5 shows user *jbr*'s navigation trail for Task A in the ALCOHOL space, who performed the best in this group. Task A corresponds to the initial overview task in Shneiderman's taxonomy. Users must locate clusters of relevant documents in the map. Subsequent tasks are increasingly focused.
As shown in the trajectory map, user *jbr* started from the node 57 and moved downwards along the branch. Then the trajectory jumped to node 105 and followed the long spine of the graph. Finally, the user reached the area where relevant documents are located. We found an interesting trajectory pattern - once the user locates a relevant document, he tends to explore documents in the immediate neighbouring area. This confirmed our expectation. The frequency of long-range jumps across the space decreased as the user became familiar with the structure of the space. The trajectory eventually settled to some fine-grained local search within an area where the majority relevant documents are placed, and it didn't move away from that area ever since, which was also what we expected.

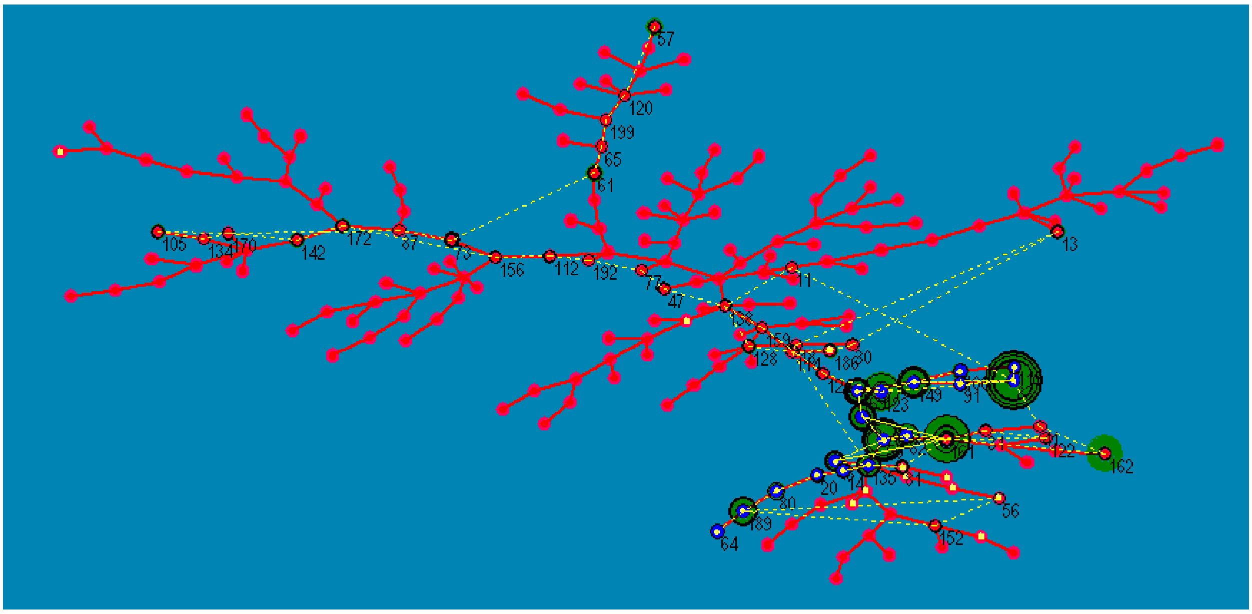


**Figure 5: User *jbr*'s trails in searching the ALCOHOL space (Task A).**

In the trajectory replay, the time spent on a document is animated as the radius of a green disc growing outward from where the document is located. This design allows us to find out whether the majority of large green discs appear in areas with a high density of relevant documents, and whether areas with a low density of relevant documents will only have sporadic passing navigation trails.

## 5. DISCUSSION AND CONCLUSION

This is an integrated approach to the study of behavioural semantics. The thematic space was exposed to users for the first time in Task A. Apart from the structural model, no navigation cues were readily available to users. Users must first locate areas in the thematic space where they can find documents relevant to the task. The optimal information foraging theory provides an appropriate description of this type of processes.
Our framework allows us to introduce descriptive and normative modelling techniques. We have conducted an in-depth study of visual information foraging strategies utilising Hidden Markov Models (HMMs) (Chen et al., 2000). The visual inspection of information foraging trails is encouraging. Animated trails and optimal paths generated by

HMMs have revealed many insights into how users were dealing with the tasks and what are the prevailing characteristics and patterns. Replay and animate HMM-paths over actual trails allow us to compare transition patterns in the same context.

In conclusion, many of our expectations have been confirmed in the visualisation and animation of trails of information foragers in thematic spaces. The task we have studies is a global information foraging in nature. The initial result is promising, especially with the facilities to animate user trails within the thematic spaces. Future studies should expand the scope of tasks to cover a fuller range of information foraging activities. Visual-spatial interfaces should be carefully designed for future studies so that fundamental issues can be addressed.

ACKNOWLEDGEMENTS

The work was supported in part by the Council for Museums, Archives and Library in the UK (RE/089).

REFERENCES

Chen, C. (1999). *Information Visualisation and Virtual Environments*. London: Springer-Verlag London.

Chen, C., Cribbin, T., Kuljis, J., & Macredie, R. (2000). Footprints of information foragers: Behaviour semantics of visual exploration. *(Submitted for Publication).*

Chen, C., & Czerwinski, M. (1998, June 1998). *From latent semantics to spatial hypertext: An integrated approach.* Paper presented at the 9th ACM Conference on Hypertext and Hypermedia (Hypertext '98), Pittsburgh, PA.

Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science, 41*(6), 391-407.

Pirolli, P., & Card., S. K. (1995). *Information foraging in information access environments.* Paper presented at the Conference on Human Factors in Computing (CHI '95), Denver, CO.