College of Information Science and Technology

Drexel E-Repository and Archive (iDEA)
http://idea.library.drexel.edu/

Drexel University Libraries
www.library.drexel.edu

Please direct questions to archives@drexel.edu

Footprints of information foragers: behaviour semantics of visual exploration

Chaomei Chen, Timothy Cribbin, Jasna Kuljis and Robert Macredie
The VIVID Research Centre, Department of Information Systems and Computing, Brunel University, Uxbridge UB8 3PH, UK. email: chaomei.chen, timothy.cribbin, jasna.kuljis, robert.macredie@brunel.ac.uk

Social navigation exploits the knowledge and experience of peer users of information resources. A wide variety of visual–spatial approaches become increasingly popular as a means to optimize information access as well as to foster and sustain a virtual community among geographically distributed users. An information landscape is among the most appealing design options of representing and communicating the essence of distributed information resources to users. A fundamental and challenging issue is how an information landscape can be designed such that it will not only preserve the essence of the underlying information structure, but also accommodate the diversity of individual users. The majority of research in social navigation has been focusing on how to extract useful information from what is in common between users' profiles, their interests and preferences. In this article, we explore the role of modelling sequential behaviour patterns of users in augmenting social navigation in thematic landscapes. In particular, we compare and analyse the trails of individual users in thematic spaces along with their cognitive ability measures. We are interested in whether such trails can provide useful guidance for social navigation if they are embedded in a visual–spatial environment. Furthermore, we are interested in whether such information can help users to learn from each other, for example, from the ones who have been successful in retrieving documents. In this article, we first describe how users' trails in sessions of an experimental study of visual information retrieval can be characterized by Hidden Markov Models. Trails of users with the most successful retrieval performance are used to estimate parameters of such models. Optimal virtual trails generated from the models are visualized and animated as if they were actual trails of individual users in order to highlight behavioural patterns that may foster social navigation. The findings of the research will provide direct input to the design of social navigation systems as well as to enrich theories of social navigation in a wider context. These findings will lead to the further development and consolidation of a tightly coupled paradigm of spatial, semantic and social navigation.

© 2002 Elsevier Science Ltd. All rights reserved.

KEYWORDS: social navigation; behaviour semantics; information visualization; hidden Markov Models; information foraging

1. Introduction

Social navigation exploits the knowledge and experience of peer users of information resources. The concept of social navigation was introduced in 1994 (Dourish &

Chalmers, 1994), in which they distinguished three models of information navigation: spatial navigation, semantic navigation, and social navigation. In its original form, social navigation refers to the notion of navigation guided by the choice of like-minded people. As explained in *Footprints in the Snow*, the ideas of social navigation build on a general concept that interacting with computers can be seen as navigation in information space (Munro, Hook & Benyon, 1999). There has been an increasing interest over the last few years in social navigation. The scope of social navigation has been substantially expanded from collaborative recommendation on the Internet to the use of the metaphor of footprints in various systems (Hill, Stead, Rosenstein & Furnas, 1995; Wexelblat & Maes, 1997).

A wide variety of visual–spatial approaches become increasingly popular as a means to optimize information access as well as to foster and sustain a virtual community among geographically distributed users. In particular, information landscape has become one of the most appealing design metaphors for representing and communicating the essence of distributed information resources to users (Chen & Carr, 1999a; Rennison, 1994; Waterworth, 1999; Wise Jr. *et al*., 1995).

The advances in information visualization have introduced a fresh perspective to investigate social navigation in virtual environments, especially the ones that are built on strong spatial–semantic couplings (Chen, 1999a; Chen & Davies, 1999). On the one hand, the increasingly wide spread of graph drawing algorithms means that more and more users are likely to deal with graphs visualized in one way or another. On the other hand, until recently it is still very much uncharted water in terms of how users actually interact and respond to various structural cues. For example, what topological features in visual–spatial interfaces influence users' search behaviour, what navigation cues should be incorporated to help users and how can social navigation benefit from insights into users' trails?

Our interest in social navigation is influenced by Erickson's work (Erickson, 1993). We have been exploring a fresh angle to conceptualize a deeper relationship between information visualization, spatial hypermedia, and three-dimensional multi-user virtual environments. Information visualization provides a new dimension to the construction and navigation of three-dimensional multi-user virtual environments. Virtual environments should provide explicit representations of underlying semantics of an information space. Over the last few years, we have been pursuing two lines of interrelated research: information visualization (Chen, 1999a) which concentrates on cognitive concerns of individual users, and multi-user spatial–semantic virtual environments (Chen & Davies, 1999; Chen, Thomas, Cole & Chennawasin, 1999) which adapts the philosophy of social construction of knowledge. Searching for a synergy between information visualization and virtual environments has profound implications on social navigation.

The most fundamental principle of social navigation is to learn from others. By understanding others' navigation trails and their behaviour in visual information retrieval, one can improve their own understanding of the information source they are dealing with and subsequently improve search strategy. In this article, we address this issue concerning social navigation: in a visual information retrieval setting, how users find their way in a visual–spatial interface, what we can learn from the successful users,

and how we can make such heuristics available to peer users so as to foster social navigation.

The rest of this article is organized as follows. First, we introduce the notion of trails and footprints in an abstract information space and various works on this topic. Second, we introduce a conceptual framework for accommodating and recommending trails. Then, we describe how we re-construct user behaviour in an experimental study and build Hidden Markov Models. Finally, we generate optimal paths using the Viterbi algorithm and conclude the article.

1.1. SOCIAL NAVIGATION

Landmarks and well-trodden footpaths are among the most commonly used navigation cues in the real world. The idea of edit wear and read wear is one example of how such cues can be adapted for social navigation in virtual worlds (Dieberger, 1997; Hill, Hollan, Wroblewski & McCandless, 1992). In this article, we focus on the counterpart of footpaths in virtual environments.

The growing number of three-dimensional multi-user virtual environments opens opportunities for researchers and practitioners to exploit the role of new information in such environments for social navigation. For example, the use of avatars in virtual environments allows us to study how people co-ordinate the use of shared information space in terms of the positioning of avatars, their orientation and proximity to each other (Jeffrey & Mark, 1999). Some systems are also designed with indirect information on positioning and access. In the Knowledge Garden system, the design follows the metaphor of a garden, where one can plant, prune or touch (Crossley, Davies, McGrath & Rejman-Greene, 1999). The Knowledge Garden is a collection of document references clustered by similarity. Social navigation cues are provided by swinging branches of a cluster whenever someone is accessing the files.

A footpath becomes increasingly wider as more people follow it. Spatial structures are being re-organized as far as these travellers are concerned. Space transformation is indeed a familiar concept in adaptive and self-improving information systems. For example, in information retrieval, an indexing space should become more and more efficient through its underlying space transformation in response to various relevance feedback received (Salton, 1975).

Figure 1 shows the interface design of the StarWalker virtual environment—a prototype developed to explore the interrelationship between spatial–semantic models and the dynamics of avatars' movements (Chen *et al.*, 1999). The central component of the StarWalker environment is the virtual world, which is a visualization of a semantic space, typically based on a collection of scientific publications. The organizational principle of the visual–spatial design is to use spatial proximity to reflect semantic similarity. Documents are mapped into the virtual world such that similar documents are near to each other, whereas unrelated documents are further apart—a widely accepted heuristic in information visualization (Chalmers, 1992; Kamada & Kawai, 1991; Wise Jr. *et al.*, 1995).

What are the possible implications of such proximity–similarity mapping in a spatial–semantic interface on users' navigation behaviour? Is it possible to establish a strong coupling between spatial–semantic metaphor and users' navigation behaviour? Figure 2
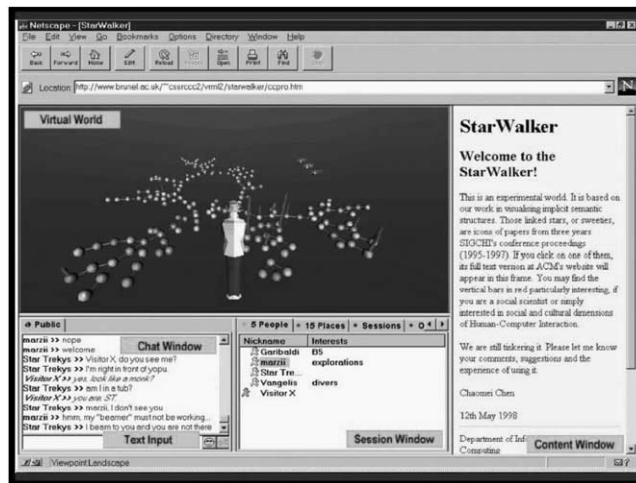
FIGURE 1. Interface design of the StarWalker virtual environment.

reveals a screenshot from StarWalker in action. In an open session in StarWalker, several avatars appeared to be attracted to the area with red poles. What would happen if more avatars were here and what if a longer duration was observed? What if users are required to search for specific information?

Behavioural semantics is an approach to modelling and analysing user behaviour in a shared semantic space. It aims to provide increased intelligibility among users and foster social navigation. It is a complementary approach to content analysis approaches such as discourse analysis, conversation analysis, and social episode studies. In this article, we describe our research in our attempts to answer these questions and introduce a conceptual framework for the study of behavioural semantics in this context.

In this article, we focus on the notion of trails of information foragers as they search information through a thematically organized visual–spatial interface. If one can make sense of the position and various movement and transition patterns of other users in a shared virtual world, then it is more likely that such information will increase mutual intelligibility, thereby increasing the chance of social navigation.

1.2. TRAILS AND FOOTPRINTS

Visualizing and animating the dynamics of user activities is a challenging issue with profound significance. The history of the notion of trails can be traced back to several decades, whereas the use of footprints has been particularly rooted in the study of social navigation.

In 1945, Bush (1945) envisaged a global and persistent device for storing and retrieving information — Memex. The Memex of more than half a century ago and the now household name of the World Wide Web have a lot things in common. For example, the concept of association to Memex then is like hyperlinks to the web today. For both Memex and the web, there is an underlying information space that users want
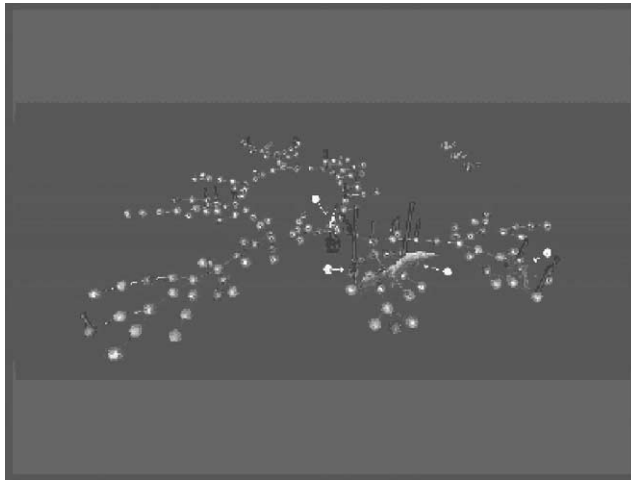
FIGURE 2. Is this a sign of social navigation?

to access. However, a crucial concept raised by Bush has not yet been fully realized on today's web. On today's web, users are surfers, whereas in Bush's Memex, users are not only surfers, but also builders with their own trails in the information space. Bush called this type of users trailblazers.

Because in Memex everything is about association, trails allow new comers to the ever-growing information space to experience what others have previously experienced and what associations have been made. The central idea of trailblazing is to preserve such valuable information and make it an integral part of the information space. To some extent, we began to see some forms of trailblazing on the web. For example, on-line bookstore Amazon can tell their customers about books they might find tempting to buy, simply based on what books are there in their shopping baskets now and the buying habits of those customers in the past who have bought the same books. In this case, the *footprints* of previous customers are invisible to new customers.

The vision of Bush has inspired several examples of visualizing trails and pathways intellectual or behavioural in nature. For example, the path model described in Chalmers, Rodden and Brodbeck (1998) provides a good example of revealing a perceived structure of an information space through users' navigation. The notion of intellectual pathways has been explored in trailblazing scientific literature (Chen, 1999b; Chen & Carr, 1999b; Small, 1999). The number of users who followed a hyperlink connecting two documents has been used to estimate the degree of relatedness between the two documents (Pirolli, Pitkow & Rao, 1996).

Alan Wexelbalt's Footprints system developed at the Media Lab at MIT is one of the pioneering systems that emphasise the dynamic structure as conceived and experienced by users on the web (Wexelblat & Maes, 1997). The Footprint system[†] visualizes the paths of users travelling from one document to another on a web locality. To make the cut on the Footprints system, a document must have been visited at least once. Such

[†] The latest version of Footprints is available at http://footprints/media.mit.edu.

visualization is useful for webmasters to spot broken links and other abnormality in users' access patterns. In Knowledge Garden, although users' trails are not directly visible, one can instead see the plant swinging in real-time whenever someone else is reading the corresponding document in the garden (Crossley *et al.*, 1999).

1.3. INFORMATION FORAGING AND EFFECTIVE VIEW NAVIGATION

Information foraging refers to a wide variety of activities in information search, including exploring, locating and assessing information. One may think of the user as someone who is hunting information and wants to maximize the rate of retrieving relevant information. This analogue is originated from the so-called optimal foraging theory developed in biology and anthropology. A mathematical model for information foraging can be found in Pirolli and Card (1995). In order to predict users' behaviour in information foraging, the profitability of a given document can be defined as the proportion of relevant documents in a cluster of documents divided by the time it will take to process the cluster, for example, to read all the documents in the cluster. Pirolli and Card (1995) found that even a crude, approximate model of information foraging predicts the observation that the harder the query is, the fewer clusters the users will select. According to the theory, profitability becomes lower if users have to spend more time to examine a set of documents.

The theory adopts optimal foraging theory in biology and anthropology, which analyses the value of food-foraging strategies and whether they should be adapted given a particular situation. Information foraging theory applies similar trade-off analytical techniques in modelling the value of information gains against the costs for the user.

In general, users must decide whether or not to pursue a given document upon encountering it, the time to spend processing a collection of documents, the choice of moves to make in navigation, the choice of strategy under uncertainty, or the degree of resource sharing.

In this article, we introduce the optimal information foraging theory in order to set our approach to modelling and animating users' trails in a broad context. From a social navigation point of view, trails of users in an information space can be used as an indicator of the profitability of a document cluster. The longer the users spend their time in a cluster, the more likely that the cluster contains relevant documents. The more frequently the users mark documents as relevant in an area, the more profitable this area is likely to be.

Another potentially relevant perspective is called effective view navigation (Furnas, 1997). In view navigation a user moves about an information structure by selecting something in the current view of the structure. Furnas (1997) explores implications of basic requirements for effective view navigation. He explored some basic issues concerning finding things in various information structures, be they webs, trees, tables or even simple lists.

1.4. HIDDEN MARKOV MODELS

Hidden Markov Models (HMMs) have been widely used in signal processing and speech recognition (Rabiner & Juang, 1993). HMMs provide a suitable means for us to

model user behaviour in information search. In particular, given the basis of optimal information foraging, HMMs can be used to enhance our theory about how users are expected to behave.

HMMs consists of states, which are not observable, and observations, which are probabilistic functions of states. State transitions are governed by a stochastic process. Observations are also a stochastic process. An HMM can be defined as follows: N denotes the number of hidden states; Q denotes the set of states $Q = \{1, 2, \ldots, N\}$; M denotes the number of symbols, or observations; V denotes the set of symbols $V = \{1, 2, \ldots, M\}$ and A denotes the state-transition probability matrix.

$$A = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_1 & \ldots & \mathbf{a}_{1N} \\ \ldots & \ldots & \mathbf{a}{ij} & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{a}_{N1} & \mathbf{a}_{N2} & \ldots & \mathbf{a}_{NN} \end{bmatrix},$$

where $\mathbf{a}_{ij} = P(q_t = |q_{t-1} = i)$, $1 \leqslant k \leqslant M$.

B denotes the observation probability distribution:

$$B_j(k) = P(o_t = k | q_t = j), \quad 1 \leqslant k \leqslant M.$$

π denotes the initial-state distribution:

$$\pi_i = P(q_1 = i), \quad 1 \leqslant i \leqslant N.$$

λ denotes the entire HMM model, $\lambda = (A, B, \pi)$

An HMM is completely defined by $\lambda = (A, B, \pi)$, which in turn are called parameters of the model. HMMs typically address the following three basic problems:

1. Given observation $O = (o1, o2, \ldots, oT)$ and model $\lambda = (A, B, \pi)$, efficiently compute $P(O|\lambda)$. Given two models $\lambda 1$ and $\lambda 2$, this can be used to choose the better one.
2. Given observation $O = (o1, o2, \ldots, oT)$ and model λ find the optimal state sequence $q = (q1, q2, \ldots, qT)$.
3. Given $O = (o1, o2, \ldots, oT)$, estimate model parameters $\lambda = (A, B, \pi)$ that maximize $P(O|\lambda)$.

A well-known algorithm from Viterbi has been widely used to find the most likely path through the HMM for each sequence, although for small state spaces it is possible to work out the answer using a brute-force approach.

In order to apply HMMs to the sequences of user activities observed in information foraging within a thematic space, we derive the transition matrix and observation probability as follows. The state space consists of all the documents in a thematic space. Each document corresponds to a unique state:

$$S_1 = d_1, S_2 = d_2, \ldots, S_N = d_N$$

A user's trail refers to the sequence of documents processed by the user in the entire session to complete a given task, including reading and marking documents. Each trail in the thematic space corresponds to a state sequence $\mathbf{S} = \{d_{i1}, d_{i2}, d_{i3}, \ldots\}$. The state transition probability matrix is derived from the sequence of documents visited by a

user in his/her session:

$$a_{ij} = \frac{d_i \to d_j}{d_i}.$$

The observation probabilities are defined as follows. Three observation symbols are defined: o_k, $K = 1, 2, 3$. O_1 denotes the user moves the mouse over a document. O_2 denotes the user clicks on the document. O_3 denotes the user marks the document as relevant. A sequence of observed symbols could be $\boldsymbol{O} = \{1, 1, 1, 2, 1, 1, 2, 3, \ldots\}$. The observation probability is also estimated from the log files:

$$b_{ik} = \frac{o_k(d_i)}{d_i}.$$

The reason we choose the document space as the state space and the three actions as the observation symbols is based on the consideration that the sequence of activities such as node over, node click and node mark is a stochastic process. This observable process is in turn the function of a latent stochastic process—the traversal of the thematic space by the user because which document the user will move to in his/her next step is very much opaque to observers.

1.5. THEORETICAL FRAMEWORK

In order to produce an integrated and cohesive theory, we introduce a theoretical framework for modelling trails of information foragers in thematic spaces. Figure 3 shows a flow chart of this framework. The framework starts with a proximity–similarity mapping, which is a common technique in information visualization. Then users are asked to complete a series of tasks designed with reference to Shneiderman's information visualization taxonomy (Shneiderman, 1996). Users' trails in each session are logged and modelled as HMMs. Users' profiles are also established in terms of their cognitive abilities and other possible criteria. Finally, synthesized user trails are generated from the HMMs and animated in context through the same graphical interfaces to the underlying thematic spaces.

2. Information foraging in thematic spaces

2.1. DESIGN OF THEMATIC SPACES

In order to study users' information foraging behaviour, we have constructed four thematic spaces. Users are asked to conduct a series of tasks in each thematic space. Each session corresponds to the completion of one task. In each session, the following information is logged to a computer file: the occurrence of an event, the time stamp of the event and the target document on which the event takes place.

Four thematic spaces were generated from the Los Angeles Times' news articles included in the Text Retrieval Conference CDROM. Each document space consisted of the top 200 documents retrieved through a single keyword query to the document collection. The length of these articles was controlled between 250 and 750 words. The
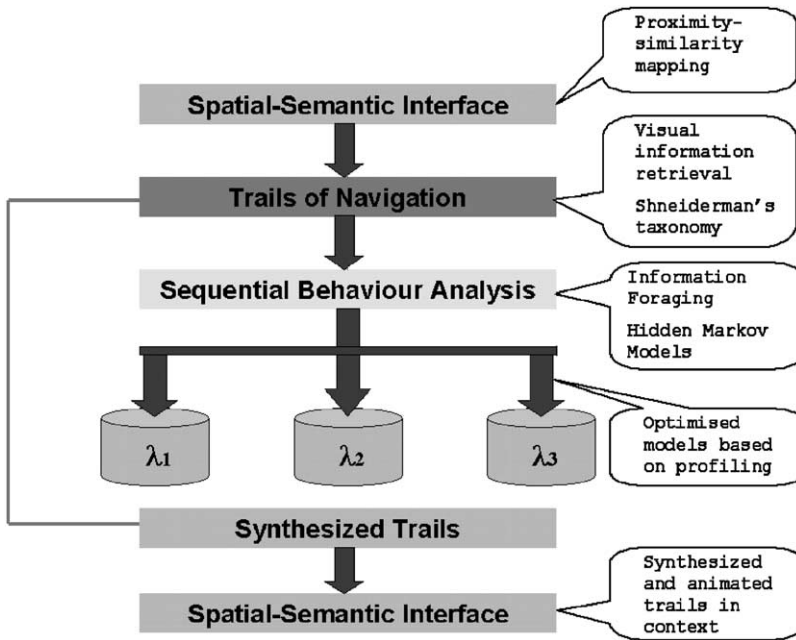
FIGURE 3. Modelling trails of information foragers in thematic spaces.
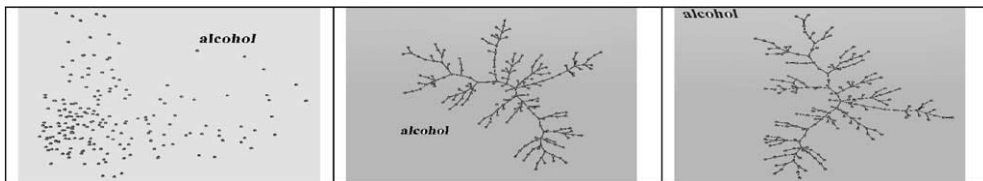

FIGURE 4. The ALCOHOL thematic space in three visualizations, from left to right, Principle Component Analysis, Minimum Spanning Tree and Pathfinder.

four keywords used were *alcohol*, *endangered*, *game* and *storm*. Corresponding spaces were named accordingly by these keywords.

Figure 4 shows the ALCOHOL thematic space visualized by three layout mechanisms, namely Principle Component Analysis (PCA), Minimum Spanning Trees (MST) and Pathfinder Networks (PF). In PCA-based visualization, the connectivity is not denoted by explicit links. Instead, one must infer the semantic similarity between two documents solely based on their spatial proximity. In MST-based visualization, $N - 1$ explicit links connect all the documents together. Users can see these links on their screen. In PF-based visualization, additional explicit links are allowed as long as the triangular inequality condition is satisfied. In our examples, the PF version tends to have only a handful of extra links at most in comparison to the MST version. Detailed descriptions of the use of these techniques for information visualization can be found in Chen (1999*a*).

2.2. DESIGN OF TASKS

The design of the tasks was based on Shneiderman's mantra: Overview, Zoom, Filter, Details on Demand, which highlights users' cognitive needs at various strategic stages in visual information retrieval.

In Task A, users are required to locate and mark documents relevant to a search topic. For example, in the Alcohol space, 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. From Task A through Task D, the scope of the target areas becomes smaller and smaller. In particular, from Task A through Task C, the target area at each level is a proper subset of the one for the immediately precedent task. The depth of Task C is designed to be equivalent to that of Task D. This inclusion relationship is made clear in the instructions to users who participated in the experiment. We expect that this should be also reflected through the trails of users — we should be able to observe that the trails of users become increasingly focused on smaller areas in the underlying thematic spaces.

We can study sequential patterns in these trails through visual representations of both the thematic spaces and the trails. If documents relevant to a task are marked in the visual interface, we should observe that the trail of a successful information forager should lead to the target area. In addition, once users' enter the target area, their trails should indicate whether they explore documents within the target area more purposefully than their trails outside the target area.

2.3. HIDDEN MARKOV MODELS

In order to characterize sequential patterns of users' trails, we construct HMMs based on the actual trails logged from sessions of the experiment. HMMs are both descriptive and normative — one can not only describe what happened with information foraging sessions, but also predict what might happen in similar situations. On the one hand, the state transition probability matrix is derived from the actual trails of users. On the other hand, behavioural patterns captured by HMMs provide insights into how users would behave as they are exposed to the same type of structural and navigational cues in the same thematic space.

In terms of the three basic problems that HMMs address, we can reword the basic problems as follows. The first basic question states that given observation $O = (o1, o2, \ldots, oT)$, which is a sequence of information foraging actions of a user, and model $\lambda = (A, B, \pi)$, efficiently compute $P(O|\lambda)$. Given two models $\lambda 1$ and $\lambda 2$, this can be used to choose the better one. We first derive an HMM model from the log files of two users: one has the best performance score, but without any node click events; the other has all types of events. We use $\lambda_{\log}$ to denote this model. Given an observation sequence, it is possible to estimate model parameters $\lambda = (A, B, \pi)$ that maximize $P(O|\lambda)$, we use λ_{seq} to denote this model. We choose the navigation path of the user with the best performance as the input sequence.

The second basic question states that given observation $O = (o1, o2, \ldots, oT)$ and model λ find the optimal state sequence $q = (q1, q2, \ldots, qT)$. In this case, we submit the navigation paths of users to the model $\lambda_{\log}$ and animate the optimal state sequences

within the thematic space. In this way, we can compare the prevalent behavioural patterns. For the purpose of social navigation, such animation will allow other users to study not only the actual trail with the best performance, but also the optimal state sequence based on the collective behavioural patterns. For those users whose data were not included in the training data, it is particularly interesting to see their behavioural patterns transformed by the HMMs.

Finally, the third basic question states that given observation $O = (o1, o2, \ldots, oT)$, estimate model parameters $\lambda = (A, B, \pi)$ that maximize $P(O|\lambda)$. We focus on the user with the best performance score in searching the thematic spaces. For example, if we observe a user clicking and marking to a great extent, it is likely that the user has successfully located the cluster of relevant documents.

2.4. VISUAL REPRESENTATIONS OF TRAILS

In order to capture behavioural as well as contextual characteristics and patterns, we have decided to animate users' navigation time series against the backdrop of the corresponding topical space. There are three types of actions that users can apply to a document in the space: mouse over to reveal the title of the document mouse click to view the content of the document, and mark the node as relevant to a given retrieval task.

Figure 5 is an annotated screenshot of the graphical interface design, which explains the animation design in detail. Non-relevant documents are marked as red. Traversal paths are defined as the transitions from one document to another. These paths appear in the animation as dotted yellow links. The ALCOHOL topical space consists of 200 documents relating to the topic of alcohol. Relevancy judgements from experts are available from the TREC test data. Documents judged to be relevant to the original search are marked with a bright yellow dot in the centre. If the user marks a document as relevant in his/her search session, this document will be coloured in blue. When the user visits a document, a dark circle is drawn around the current document. The time
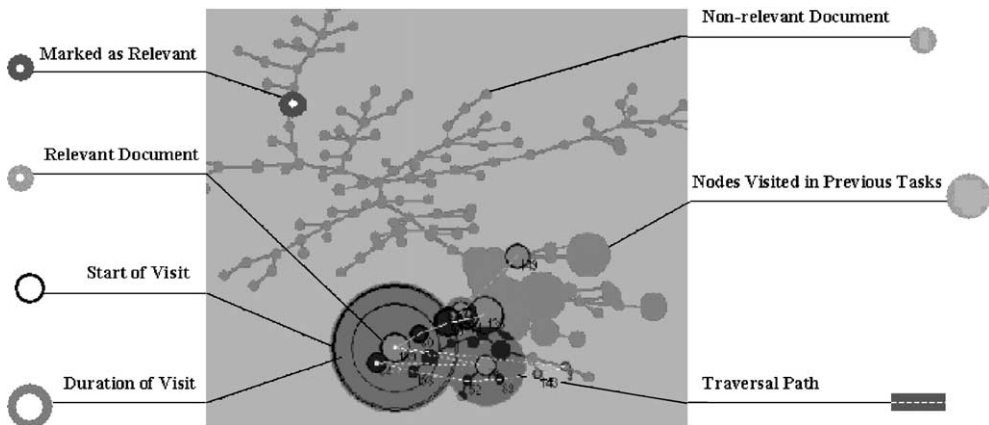

FIGURE 5. Information foraging trails.

spent on a document is denoted by a growing green belt until the user leaves the document. If the user comes back to a document he/her had visited before, 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 in one session into the next session. In this case, a red disc will outline the profile of previous session.

Based on the above reasoning, we can formulate our expectations in the following hypotheses.

(a) Spatial–semantic models may reduce the time spent on examining a cluster of documents if the proximity–similarity mapping preserves the latent semantic structure.

(b) Spatial–semantic models may mislead information foragers to over-estimate the profitability of a cluster of documents if the quality of clustering is low.

(c) Once users locate a relevant document in a spatial–semantic model, they tend to switch to local search.

(d) 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.

(e) Discs of subsequent tasks are likely to be embedded in discs of preceding tasks.

3. Results

The performance results have showed that users performed better with interfaces based on MST visualizations. We will focus on trails of users in these spaces, especially in the Alcohol thematic space. Four users participated in sessions for Task A in this space. Figure 6 shows an overview map of the ALCOHOL space in which documents relevant to Task A are marked with bright yellow dots in the centre. All the relevant documents are clustered in the branch located at the lower right-hand corner of the map, with the
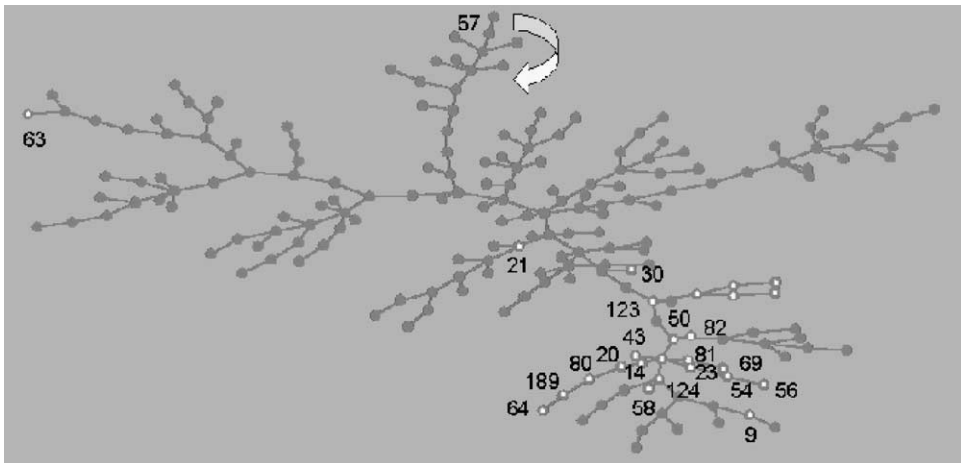

FIGURE 6. The distribution of target documents for Task A in the ALCOHOL space (MST).

exceptional documents being number 63 and number 21. Another special node in the map is number 57. As will be shown subsequently, three out of four users we studied chose this node as the starting point for their navigation.

3.1. TRIALS OF INFORMATION FORAGERS

Because of the superior performance results with MST-based interfaces, in the following discussion, we focus on information foraging behaviour in the MST version of the ALCOHOL topical space. Four users took part in the experiment and searched the ALCOHOL space visualized as an MST.

Table 1 lists the details of these four users, including their psychometrics and retrieval performance measure F_{pr}, which is defined as follows:

$$F_{pr} = \frac{pr}{p + r}$$

where p is the precision and r the recall

In the psychometrics block in the table, measures above the average are highlighted in bold. Various factor-referenced tests are included in the experiment, MV—Visual Memory, MA—Associative Memory, VZ—Spatial Ability, FI—Ideational Fluency, P—Perception Speed, FA—Association Fluency, S—Spatial Orientation (Eckstrom, French, Harman & Derman, 1976). User jbr has above-average measures in most cognitive ability tests, except in Perceptual Speed (P). In contrast, user nol has a strong P, but relatively weak in others. In terms of the F_{pr} measures breakdown by tasks, user jbr has the best performance in Task A ($F_{pr} = 0.54$), whereas user nol has $F_{pr} = 0.00$ in Task A.

Figure 7 shows the trail of user jbr in Task A in the thematic space on alcohol because he has the best performance in this group. Task A is designed in relation to the initial overview task in Shneiderman's taxonomy. In this task, users must locate clusters of relevant documents in the map. Subsequent tasks become more and more focused. Each trajectory map corresponds to the navigation and information foraging actions of one user.

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

TABLE 1
Cognitive ability scores of users completed Task A in the alcohol-MST condition

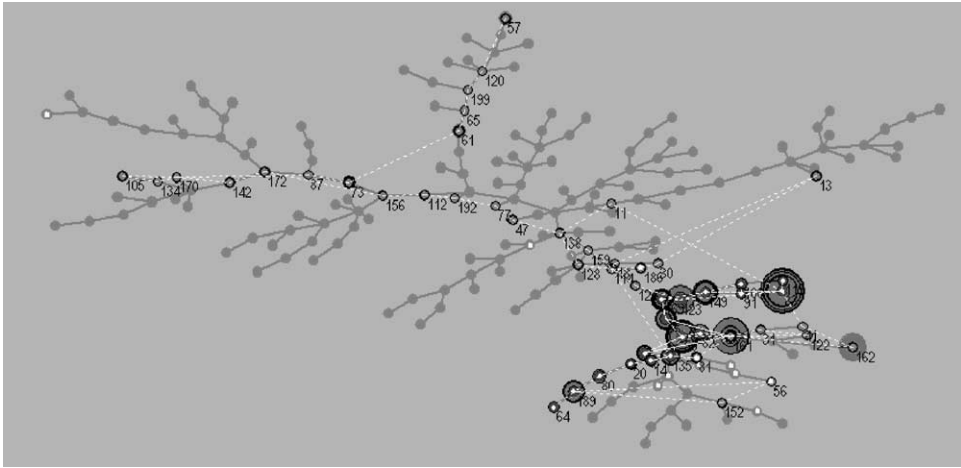| User	MV	FA	S	MA	VZ	FI	P	Sex	F_{pr} Measure (MST)			
									T.A	T.B	T.C	T.D
jbr	**22.00**	**38.00**	**102.00**	**27.00**	**18.75**	**49.00**	52.00	M	**0.54**	**0.32**	**0.80**	**0.00**
msh	9.00	18.00	55.00	**22.00**	−1.25	21.00	25.00	M	**0.14**	**0.49**	**0.36**	**0.11**
mas	20.00	**37.00**	79.00	**19.00**	3.00	**32.00**	44.00	M	**0.08**	**0.07**	**0.03**	**0.00**
nol	18.00	15.00	73.00	14.00	7.25	**31.00**	59.00	F	**0.00**	**0.00**	**0.13**	**0.09**
Average	21.05	21.62	102.19	16.81	11.52	28.19	57.00					
S.D.	6.62	7.47	31.17	6.84	5.41	8.11	13.83					

FIGURE 7. Trajectory of user jbr as he is excelling on Task A in the thematic space on alcohol.

long spine of the graph. Finally, the user reached the area where relevant documents are located. One of the most interesting trajectory patterns is that once the user locates a relevant document, he tends to explore the immediate neighbouring area. This confirms our expectation. Long-range jumps across the space become increasingly rare as the user gets familiar with the structure of the space. The trajectory eventually remained within the area with the most relevant documents, and it did not go outside that area ever since. This also confirms our expectation.

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 will allow us to verify whether the majority of large green discs appear in areas with relevant documents and whether areas with few relevant documents will only be connected by sporadic navigation paths.

We observed that users were able to mark certain documents extremely fast. For example, user jbr apparently spent almost no time to determine the relevancy of documents 80, 20 and 64 and marked them in blue. It appears once users have identified two relevant documents, they tend to identify relevant documents in between very quickly. Explicit links in the visualization play a crucial role in guiding the course of navigation of users. Users not only follow these links in their navigation, but also make their relevance judgement based on the cues provided by these visible links. In other words, users have taken these explicit links into account when they access the profitability of a document cluster.

Figure 8 shows the trajectory of another user conducting the same task in the same space. However, in this case, the traversal did not reach the target area. Instead, the user has spent a considerable amount of time foraging in areas that contain no relevant documents. In particular, the user repeatedly visited node 120 and even marked it in blue. This is shown in the map as the huge green disc on the top.

Figure 9 shows the trajectories of the other two remaining users in the same *alcohol* space for Task A. One can tell from the path of their trails that their performance is not
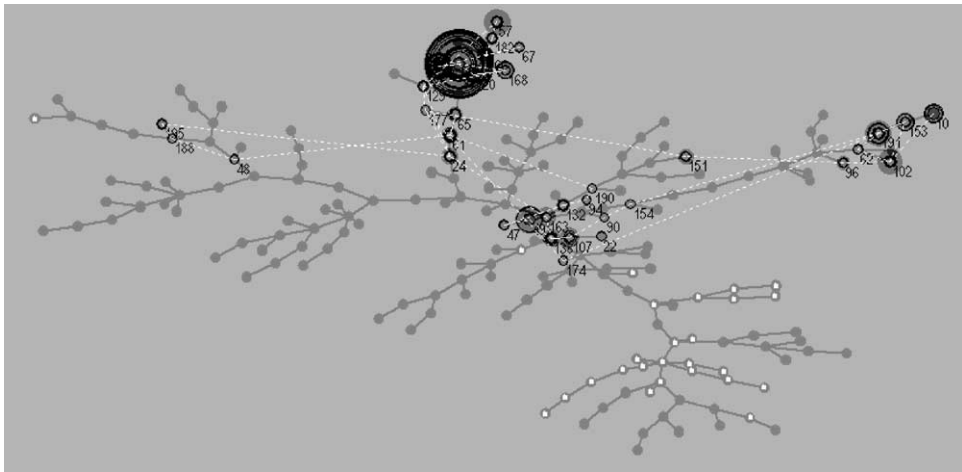
FIGURE 8. Trajectory of user nol conducting Task A in the thematic space on alcohol.

user mas   user msh
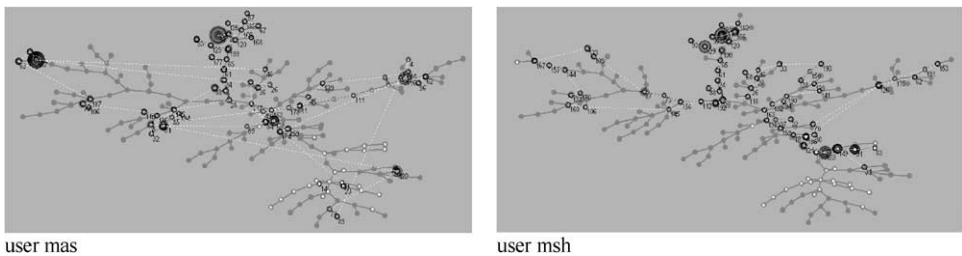
FIGURE 9. Trajectories of two other users in the ALCOHOL space.

as successful as the first user jbr. Documents in the target area were hardly visited. Interestingly, all four users started with the same branch of documents in the thematic space — the one starting with node 57 on the top and downward.

Each user participated four retrieval tasks A–D. Tasks A–C are about the same topic, but require increasingly specific and more detailed information from the space. Task D is on a different topic at the same level as Task C. Trajectory maps are designed so that an outline of the trajectory from the previous task can be preserved and carried over to the next task. For example, if the user spends a long time at a document in Task A, the accumulative trajectory map starts with the information. Following the same logic on the relationship between relevance judgement and the time spent on a document, we expect to see the scope of areas within which the user spends considerably longer time in their navigation will be gradually narrowed down. In addition, as users become more and more familiar with the structure and content of the trajectory map, we expect that there is no need for them to re-visit areas that seem unlikely to be profitable.

Figure 10 shows the trajectories of the same user jbr for four tasks. It is clear from the maps that the user spent longer and longer time in areas with relevant documents. In the last trajectory map for Task D, the user began to forage information in new areas.
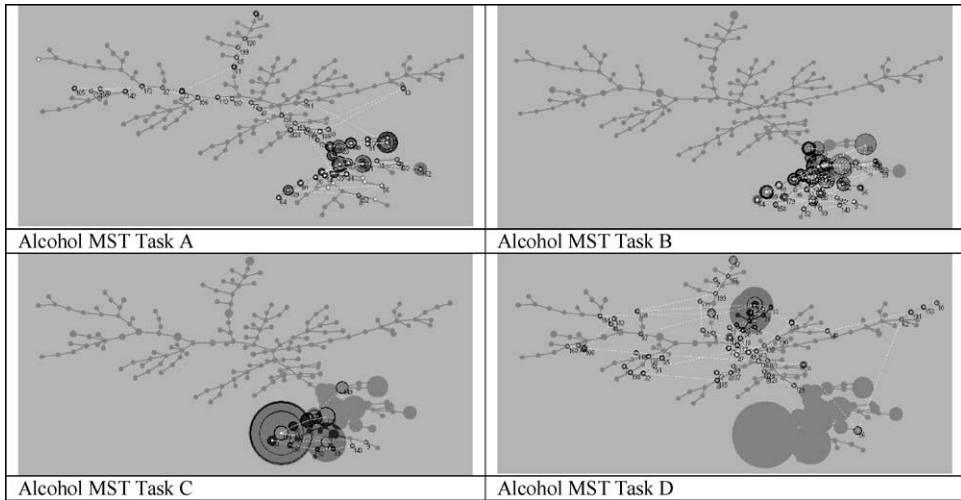
FIGURE 10. Accumulative trajectory maps of user jbr in four consecutive sessions of tasks. Activated areas in each session reflect the changes of the scope.

3.2. HIDDEN MARKOV MODELS

Trajectories of individual users have revealed many insightful findings. The next step is to extract behavioural patterns from the group of users as a whole. From a social navigation point of view, one has not only to understand the characteristics of the trajectory of individual users in a spatial-semantic space, but also to identify the commonality across individuals' behavioural patterns.

HMMs allow us to describe and predict sequential behaviour characteristics of users foraging information in thematic spaces. We categorize users' information foraging actions into three types of action events.

(1) Node over.
(2) Node click.
(3) Node mark.

When the user moves his/her mouse over a document in the thematic space, the title is flashed out on the screen. When the user clicks on the document, the content of the document becomes available. When the user has decided that the current document is relevant for the task, he/she can mark the document. Table 2 lists the breakdowns of these actions across the four users. User jbr marked 17 documents as relevant. The other three users marked considerably less.

The following example illustrates user jbr's original sequence of information foraging actions, or observations in terms of HMMs. The total length of the sequence is 120, among which 103 were node overs and the remaining 17 were node marks. No node marking was recorded during this session. This is in part due to a temporary problem with logging node click events. However, it is still a valid representation from the HMM point of view, and after all, we are interested in node marking events.

TABLE 2
Foraging actions of users observed in the MST-version of the thematic space on alcohol

Users	Node over	Node click	Node mark	Total
jbr	103	0	17	120
mas	104	9	5	118
msh	97	4	2	103
nol	78	26	4	108

TABLE 3
State transition matrix of the HMM λ_{state} derived from the training data for the ALCOHOL topical space ($N = 200$, $M = 3$)

λ_{state}	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
s1	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
s2	0.000	0.625	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
s3	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
s4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
s5	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
s6	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
s7	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
s8	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
s9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.800
s10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Observed sequence of symbols for user jbr ($T = 120$)

```
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 3 1 1 1 1 1 1 1 1 1 3 1 3 1 1
1 1 1 1 1 1 1 3 1 3 1 3 1 1 1 1 3 1 3 1 1 1 1 1 1 1 1 1 3 1
1 1 1 1 1 1 1 3 1 1 1 1 1 3 1 3 1 3 1 3 1 1 1 1 1 3 1 3 1
```

First, we use two users' trails as the training set to build the first HMM λ_{state}. We choose users jbr and nol because one marked the most documents and the other clicked the most number of times. Table 3 shows a small part of the state transition matrix—only transitions among the first 10 states are shown.

Table 4 shows the observation probability distribution over three information foraging actions. The entire matrix is not shown due to its size. For example, as far as the document number 4 is concerned, it is certain that the only information foraging action is node over.

The third parameter of an HMM is the initial distribution, denoted as π. Table 5 shows a part of the initial distribution. Intuitively, this is the likelihood that users will start with a given document for their information foraging.

In addition to the above approach, one can derive an HMM by using an algorithm called the Baum–Welch algorithm based on a given sequence of observed actions. We

TABLE 4
*Observation probability distribution derived from users' log files (*jbr *and* nol*)*

State	o_1	o_2	o_3
s1	0.333	0.333	0.333
s2	0.625	0.312	0.062
s3	0.333	0.333	0.333
s4	1.000	0.000	0.000
s5	0.333	0.333	0.333
s6	0.333	0.333	0.333
s7	0.333	0.333	0.333
s8	0.333	0.333	0.333
s9	0.333	0.333	0.333
s10	0.600	0.200	0.200

TABLE 5
Initial distribution of the HMM model

$\pi 1$	$\pi 2$	$\pi 3$	$\pi 4$	$\pi 5$	$\pi 6$	$\pi 7$	$\pi 8$	$\pi 9$	$\pi 10$
0.000	0.080	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.025

TABLE 6
State transition matrix of the HMM λ_{obs} by the Baum–Welch algorithm

λ_{obs}	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10
s1	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
s2	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
s3	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
s4	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
s5	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
s6	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
s7	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
s8	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
s9	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006
s10	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006	0.006

use user jbr's action sequence as the input and generate an HMM. This model is denoted by λ_{obs}. Table 6 shows a part of the state transition matrix of this model.

Similarly, the observation probability distribution is shown in Table 7 in part. The initial distribution of the second HMM is shown in part in Table 8.

The following example illustrates the difference between a first-step Markov chain and an HMM. Suppose we have a transition path from node d_1 to d_9 and eventually mark d_9 as relevant.

$$\rho_1 = (d_1, d_4, d_8, d_9, M_9).$$

TABLE 7
Observation distribution of the HMM λ_{obs}

State	o_1	o_2	o_3
s1	0.997	0.001	0.004
s2	0.997	0.001	0.004
s3	0.997	0.001	0.004
s4	0.999	0.001	0.002
s5	0.997	0.001	0.004
s6	0.997	0.001	0.004
s7	0.998	0.001	0.003
s8	0.997	0.001	0.004
s9	0.997	0.001	0.004
s10	0.997	0.001	0.004

If we only consider the transition probability of the path, we have

$$p(\rho_1) = p(d_1)p(d_1 \rightarrow d_4)p(d_4 \rightarrow d_8)p(d_8 \rightarrow d_9)p(d_9 \rightarrow b_3)$$
$$= 0.007 \times 0.006 \times 0.006 \times 0.006 \times 0.004$$
$$= 6.05 \times 10^{12}.$$

If we know the probability of visiting node d_k leading to the marking of d_9, we can compute the likelihood that the node d_9 will be marked if the user follows this path. Assume that the following probabilities are known, in fact, they can be estimated from the sequential data in the log files:

$$p(d_1)p(d_1 \rightarrow d_4)p(m_9|d_4)p(d_4 \rightarrow d_8)p(m_9|d_8)p(d_8 \rightarrow d_9)p(m_9|d_9)$$
$$= 0.007 \times 0.006 \times 0.020 \times 0.006 \times 0.024 \times 0.006 \times 0.004$$
$$= 2.91 \times 10^{15}.$$

Using the HMM derived from user jbr's and user nol's actual sequences, we can verify the internal structure of the model using the well-known algorithm — the Viterbi algorithm. Given a HMM λ and a sequence of observed symbols, the Viterbi algorithm can be used to generate a sequence of states. One can examine this state sequence and compare it with the original sequence of events log from the user.

The following is the state sequence generated by the Viterbi algorithm based on the HMM λ_{state}, which returns the sequence of states that is most likely to emit the observed symbols, i.e. the information foraging sequence. Relevant documents in the state sequence are highlighted in bold. This sequence is identical to the original sequence recorded in the session.

State sequence generated by the HMM for user jbr

67 57 120 199 65 61 61 61 73 73 73 87 170 134 105 170 142 172 156 112
192 77 47 138 128 114 186 30 13 13 18 114 135 **50** 161 **50 43 50** 66 **50**
50 66 161 66 66 169 66 66 **169 169** 123 123 83 **149 169 169** 123 123 **149**
149 **83** 11 138 159 121 123 **149 149 100 100 91 91 83 83 119** 83 83 119
119 83 41 162 162 **82 50 82 82 82 82** 161 122 31 **43 135 81** 161 **43 43**
135 81 81 135 14 **135 135 14 14 20 20 80 80 189 189** 152 **56 189 189 64 64** 158

TABLE 8

Initial distribution of the HMM model λ_{obs}

$\pi 1$	$\pi 2$	$\pi 3$	$\pi 4$	$\pi 5$	$\pi 6$	$\pi 7$	$\pi 8$	$\pi 9$	$\pi 10$	...
0.007	0.007	0.007	0.004	0.007	0.007	0.006	0.007	0.007	0.007	...

TABLE 9

State sequence and information foraging sequence of user nol *in the thematic space on alcohol (length=108)*

State sequence

4 10 10 10 10 10 102 153 153 153 191 62 191 191 62 102 102 102 102 96 96 151 151 177
174 191 153 191 191 154 90 107 138 163 163 24 177 129 57 57 57 182 182 166 166 67 97
97 2 2 2 2 97 97 97 97 2 2 120 120 168 120 120 2 97 2 2 97 2 2 120 2 2 2 2 2 168 129
24 24 195 195 195 188 48 61 65 65 61 61 190 94 132 132 47 47 49 49 49 49 49 138
138 138 107 107 22 22

Symbol sequence

1 1 1 2 1 3 1 1 1 2 1 1 1 2 1 1 1 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 2
1 1 1 1 2 1 1 1 2 1 2 1 3 1 2 1 2 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 3 1 1 1 1 1 1 1 1 1
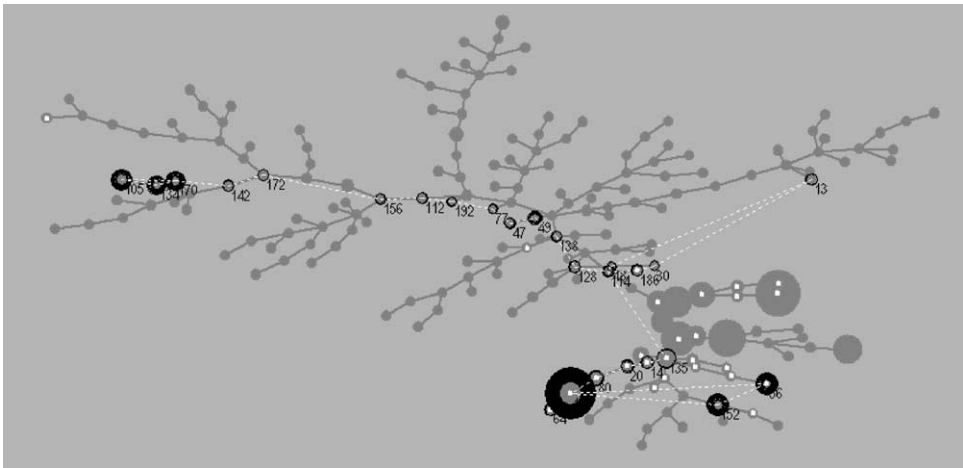1 2 1 2 1 1 1 2 1 1 1 1 2 1 3 1 1 2 1 2 1 2

FIGURE 11. Trajectory of the optimal path over the original path of user jbr.

As a comparison, Table 9 shows the state sequence and information foraging action sequence for user *nol*. A total of 78 node over actions were made, 26 node clicks and only four node mark actions.

Based on the HMM λ_{state}, we used user jbr's observed information foraging action sequence as the input and applied the Viterbi algorithm to generate the optimal state transition path. Figures 11 and 12 shows the trajectory of the path generated by the Viterbi algorithm. The path started from the left-hand side of the thematic space and
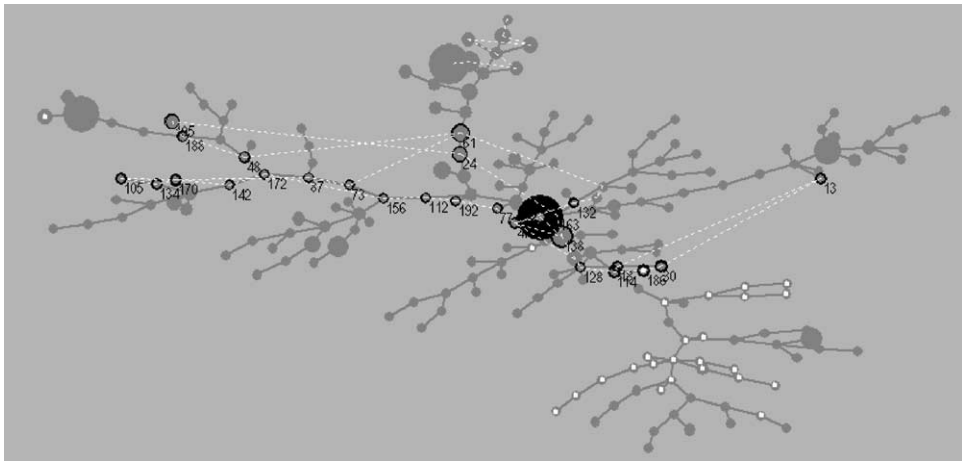
FIGURE 12. Trajectory of the optimal path over the original path of user nol.

traced the horizontal spine across the map and reached the target area. The path finished in the target area with several extended visits to relevant documents in this area. The optimal path is drawn on top of the original trail of the same user. By showing the two versions of the trails on the same thematic map, it becomes clear where the discrepancies are and where the conformance is. Since this is a novel way to represent paths in an HMM, many characteristics are yet to be fully investigated. Even though, the synthesized path appears to be promising and it moves straight to the target area some wandering in the original trail has been filtered out. For social navigation, the optimal path is likely to provide an enhanced profile for this group of users.

4. Discussion

This is the first step towards the development of an integrated approach to the study of behavioural semantics. The number of users who conducted Task A with the same visualization of the same thematic space is four. We focused our analysis on the thematic space on alcohol in MST because on the average, users scored relatively higher with MST visualizations. 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.

We have made an assumption that this is an information foraging process and it is also a stochastic process because much of the judgements and decisions made by users in their exploration and foraging of relevant documents are implicit and difficult to externalize. The introduction of HMMs allows us to build descriptive and normative models so that we can characterize sequential behaviour of users in the context of information foraging. The visual inspection of information foraging trails is encouraging. Animated trails and optimal paths generated by HMMs have revealed

TABLE 10
Coverage of the study reported in this article

Space\visualization	PCA	MST	PF
Alcohol	No	Yes	No
Endangered	No	No	No
Game	No	No	No
Storm	No	No	No

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

Due to the small sample size that we can use in the modelling process, we have not been able to gather trails from a sufficient number of users and classify trails according to cognitive ability scores such as Spatial Ability, Associative Memory and Perceptual Speed. We are planning to conduct an experiment which focuses on individual differences in information foraging behaviour in the same thematic space by increasing the sample size and extending the duration of sessions. Although each foraging session lasted for 30 min, if we break down to tasks, each user has only 5 min.

The levels of tasks are related to the scope of search. In this article, we studied Task A, which is by nature a global information foraging within the entire thematic space. Evidence shows that subsequent tasks have changed to local search to a great extent. This raises an interesting question as to how one can extend the study to tasks involving different search strategies. We have touched upon the shrinking-scope tendency in this article, but studies of the full range of tasks with reference to Shneiderman's task-data type taxonomy should lead to deeper insights into how users interact with visual–spatial interfaces.

As far as the resultant HMMs are concerned, a clearer understanding and interpretation of various characteristics manifested by paths selected by HMMs are certainly desirable. We have only analysed a small portion of the data generated from our experiment. Among 12 combinations of visual–spatial interfaces and underlying thematic spaces, we have only studied one—Alcohol in MST. Table 10 lists the remaining combinations that have not been studied in this article.

There are a number of possible ways to incorporate social navigation cues into a virtual environment with the behaviour semantic approach. HMM allow us to generate optimal paths and characterize various transition patterns in a thematic space. In addition to animations of trails and HMM-paths, one can use ghost avatars to traverse the thematic space along with the real users. Ghost avatars can travel along HMM-generated paths as well as actual trails, which will in turn inspire other users and draw their attention to profitable areas in information foraging.

5. Conclusion

In conclusion, many of our expectations have been confirmed in the visualization and animation of trails of information foragers in thematic spaces. The task we have studied

is a global information foraging in nature. The initial integration of the optimal information foraging and Hidden Markov Models (HMMs) is promising, especially with the facilities to animate user trails within the thematic spaces.

Visual representations of information foraging have led to valuable insights into how users explore and navigate through thematic spaces. The only visible navigation cues for users in these spaces are structures resulted from a proximity–similarity mapping. Labelling in its own right is a challenging issue—how to generate the most meaningful labels and summarize unstructured documents. However, because the aim of this study is to investigate information foraging behaviour, it has been decided not to label document clusters for users in the experiment. Precisely because of this reason, the optimal information foraging theory and HMM are appropriate tools. Users have indeed raised the issue concerning labelling local areas in the thematic space.

In future studies, there are several possible routes to pursue. One can repeat the study with a large sample size of users and classify users according to their cognitive abilities or other criteria. Then one can compare HMMs across different user classes and make connections between information foraging behaviour of users and their individual differences. 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.

As far as social navigation is concerned, one can enhance navigation cues in a multi-user virtual environment with various actual trails and synthesized paths. The findings of the research will provide direct input to the design of social navigation systems as well as to enrich theories of social navigation in a wider context.

Our goal is to improve the understanding of how such information can help users to learn from each other, for example, from the ones who have been successful in retrieving documents. Additional navigation cues, guided tours and even ghost avatars are likely to provide a richer source of information for users to maintain awareness, to judge the profitability of information resources, and to follow like-minded people.

The behaviour semantic approach will lead to the further development and consolidation of a tightly coupled paradigm of spatial, semantic and social navigation. This novel approach offers a methodology that can be used to combine technologies of information visualization and user behavioural modelling. Not only can a user's navigation path be vividly replayed on the computer screen, but also a virtual path derived from a group of users with certain characteristics in common.

References

BUSH, V. (1945). As we may think. *The Atlantic Monthly*, **176,** 101–108.
CHALMERS, M. (1992). BEAD: Explorations in information visualisation. *Proceedings of SIGIR '92*, pp. 330–337. Copenhagen, Denmark.

CHALMERS, M., RODDEN, K. & BRODBECK, D. (1998). The order of things: Activity-centred information access. *Proceedings of the 7th International World Wide Web Conference (WWW7)*, pp. 359–367. Brisbane.

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

CHEN, C. (1999*b*). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing and Management*, **35,** 401–420.

CHEN, C. & CARR, L. (1999*a*). Trailblazing the literature of hypertext: author co-citation analysis (1989–1998). *Proceedings of the 10th ACM Conference on Hypertext (Hypertext '99)*, pp. 51–60. Darmstadt, Germany.

CHEN, C. & CARR, L. (1999*b*). Visualizing the evolution of a subject domain: a case study. *Proceedings of IEEE Visualization '99*, pp. 499–502. San Francisco, CA. (plus color plates).

CHEN, C. & DAVIES, J. (1999). Integrating spatial, semantic, and social structures for knowledge management. *Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS '32)*, Hawaii.

CHEN, C., THOMAS, L., COLE, J. & CENNAWASIN, C. (1999). Representing the semantics of virtual spaces. *IEEE Multimedia*, **6** (2)**,** 54–63.

CROSSLEY, M., DAVIES, N. J., McGRATH, A. J. & REJMAN-GREENE, M. A. Z. (1999). The knowledge garden. *BT Technology Journal*, **17,** 76–83.

DIEBERGER, A. (1997). Supporting social navigation on the World-Wide Web. *International Journal of Human-Computer Studies*, **46,** 805–825.

DOURISH, P. & CHALMERS, M. (1994). Running out of space: models of information navigation. *Proceedings of HCI '94*, Glasgow, U.K.

ECKSTROM, R. B., FRENCH, J. W., HARMAN, H. H. & DERMAN, D. (1976). *Kit of Factor-referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

ERICKSON, T. (1993). From interface to interplace: the spatial environment as a medium for interaction. *Proceedings of the European Conference on Spatial Information Theory*, pp. 391–405. Island of Elba, Italy.

FURNAS, G. W. (1997). *Effective View Navigation. Proceedings of CHI '97*, Atlanta, Georgia. March 22–27, 1997.

HILL, W., STEAD, L., ROSENSTEIN, M. & FURNAS, G. (1995). Recommending and evaluating choices in a virtual community of use. *Proceedings of CHI '95*.

HILL, W. C., HOLLAN, J. D., WROBLEWSKI, D. & McCANDLESS, T. (1992). Edit wear and read wear. *Proceedings of CHI '92 Conference on Human Factors in Computing Systems*, pp. 3–9. Monterey, CA.

JEFFREY, P. & MARK, G. (1999). Navigating the virtual landscape: co-ordinating the shared use of space. In A. J. MUNRO, K. HÖÖK & D. BENYON, Eds., *Social Navigation of Information Space*, pp. 112–131. London: Springer-Verlag.

KAMADA, T. & KAWAI, S. (1991). A general framework for visualizing abstract objects and relations. *ACM Transactions on Graphics*, **10,** 1–39.

MUNRO, A. J., HOOK, K. & BENYON, D. (1999). Footprints in the snow. In A. J. MUNRO, K. HOOK & D. BENYON, Eds. *Social Navigation of Information Space*, pp. 1–14. London: Springer-Verlag.

PIROLLI, P. & CARD, S. K. (1995). Information foraging in information access environments. *Proceedings of the Conference on Human Factors in Computing (CHI '95)*, pp. 51–58. Denver, CO.

PIROLLI, P., PITKOW, J. & RAO, R. (1996). Silk from a sow's ear: extracting usable structures from the web. *Proceedings of CHI '96*, pp. 118–125. Vancouver, BC, Canada.

RABINER, L. R. & JUANG, B. H. (1993). *Fundamentals of Speech Recognition*. Englewood Cliffs NJ: Prentice-Hall.

RENNISON, E. (1994). Galaxy of news: An approach to visualizing and understanding expansive news landscapes. *Proceedings of ACM UIST '94*, pp. 3–12. Marina del Ray, California, USA.

SALTON, G. (1975). *A Theory of Indexing*. Philadelphia, PA: The Society for Industrial and Applied Mathematics.

SHNEIDERMAN, B. (1996). The eyes have it: A task by data type taxonomy for information visualisation. *Proceedings of IEEE Conference on Visual Languages*, pp. 336–343. Boulder, CO.

SMALL, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, **50,** 799–813.

WATERWORTH, J. A. (1999). Spaces, places, landscapes and views: experiential design of shared information spaces. In A. J. MUNRO, K. HOOK & D. BENYON, Eds. *Social Navigation of Information Space*, pp. 132–154. London: Springer Verlag.

WEXELBLAT, A. & MAES, P. (1997). Visualizing histories for Web browsing. *Proceedings of RIAO'97: Computer-Assisted Information Retrieval on the Internet*, Montreal.

WISE Jr J. A., THOMAS, J. J., PENNOCK, K., LANTRIP, D., POTTIER, M., SCHUR, A. & CROW, V. (1995). Visualizing the non-visual: Spatial analysis and interaction with information from text documents. *Proceedings of IEEE Symposium on Information Visualization '95*, Atlanta, GA, USA.