Integrating Information Seeking and Information Structuring

Spatial Hypertext as an Interface to the Digital Library

Submitted in partial fulfilment of
the Degree of Doctor of Philosophy
awarded by Middlesex University

George Buchanan

December 2004

Abstract

Information seeking is the task of finding documents that satisfy the information needs of a person or organisation. Digital Libraries are one means of providing documents to meet the information needs of their users – i.e. as a resource to support information seeking. Therefore, research into the activity of information seeking is key to the development and understanding of digital libraries.

Information structuring is the activity of organising documents found in the process of information seeking. Information structuring can be seen as either part of information seeking, or as a separate, complementary activity. It is a task performed by the seeker themselves and targeted by them to support their understanding and the management of later seeking activity. Though information structuring is an important task, it receives sparse support in current digital library systems.

Spatial hypertexts are computer software systems that have been specifically been developed to support information structuring. However, they seldom are connected to systems that support information seeking. Thus to day, the two inter-related activities of information seeking and information structuring have been supported by disjoint computer systems.

However, a variety of research strongly indicates that in physical environments, information seeking and information structuring are closely inter-related activities. Given this connection, this thesis explores whether a similar relationship can be found in electronic information seeking environments. However, given the absence of a software system that supports both activities well, there is an immediate practical problem.

In this thesis, I introduce an integrated information seeking and structuring system, called Garnet, that provides a spatial hypertext interface that also supports information seeking in a digital library. The opportunity of supporting information seeking by the artefacts of information structuring is explored in the Garnet system, drawing on the benefits previously found in supporting one information seeking activity with the artefacts of another.

Garnet and its use are studied in a qualitative user study that results in the comparison of user behaviour in a combined electronic environment with previous studies in physical environments. The response of participants to using Garnet is reported, particularly regarding their perceptions of the combined system and the quality of the interaction. Finally, the potential value of the artefacts of information structuring to support information seeking is also evaluated.

Acknowledgements

I would like to thank my supervisors, Dr. Ann Blandford, Dr. Bob Fields, Dr. Matt Jones and Prof. Harold Thimbleby for their support throughout the writing of this thesis and my doctoral studies.

Thanks are also due to the New Zealand Digital Library Project at the University of Waikato, particularly Prof. Ian Witten and Dr. David Bainbridge,for their ongoing co-operation on digital library research without which the work presented in this thesis would have been considerably more arduous.

This work was supported by Middlesex University through their financial support of my study during my employment their, and many members of the School of Computer Science deserve mention. Particular thanks are due to Prof. Colin Tully, Director of Research, for his continued support of this work during an ongoing illness of this thesis' author; Dr. Richard Butterworth, Dr. Paul Curzon and Dr. Elke Duncker-Gassen, Suzette Keith, Hanna Stelmaszewska and other members of the Interaction Design Centre; and Leonard Miraziz for much timely and generous technical assistance.

In the hypertext community, I must acknowledge the helpful insights of a number of respected researchers in the field who provided unpublished information, pre-prints and many helpful discussions – Dr. Frank Shipman, Texas A&M University; Dr. Cathy Marshall, Microsoft Corporation and Dr. Robert Akscyn of Knowledge Engineering.

Foreword

Much of this thesis has been presented in published papers. The pertinent parts of the thesis and papers are listed below. The full details of the papers may be found in the references.

Chapter 2

[Buchanan et al 2001]

Literature review

Chapter 4

[Buchanan et al 2001, Buchanan et al 2002]

Overview of technical design and interface of system.

[Bainbridge at al 2001]

Digital library protocol review and implementation of the Greenstone Protocol.

Chapters 5-7

[Buchanan et al 2004] –

Design and results of experiment, with a particular focus upon the results pertinent to Spatial Hypertext research.

This paper won the ACM Ted Nelson Award 2004, for best paper written by authors not previously published in a hypertext conference.

[Buchanan et al 2004a]

Design and results of experiment, focussing upon the digital library viewpoint.

Other papers by the author of this thesis pertinent to this work are referred to as normal in the text of the thesis.

With the exception of Bainbridge et al, the above works were written in the course of my Doctoral studies with my supervisors as co-authors in normal academic practice.

Bainbridge et al was primarily co-authored by myself and Dr. David Bainbridge, with the Greenstone protocol and comparative review done by myself, the Dienst, SDLIP and Z39-50 Greenstone interfaces created under the direction of Dr. Bainbridge, further editing and project oversight by Prof. Ian Witten and technical contributions to Dr. Bainbridge's work by the other authors.

Chapter 1: Introduction

Information seeking is the process by which a user searching for information attempts to discover documents that are relevant to their interest. A reader in a library who refers to the card index catalogue is engaged in information seeking, as is a user who reads an online document to identify possible leads for new material in its references. Many years of research into information seeking has revealed that it is a complex process containing many interconnected tasks.

This thesis focuses upon information structuring, one of the many aspects of information seeking. Information structuring is the activity of collecting and organising documents found by a reader in the course of their information seeking. Examples of information structuring would include sorting books into different piles on a desk in a physical library, filing copies of papers of interest into a filing cabinet or drawing a mind-map™ of articles and books found whilst preparing an essay. The structuring activity supports the user's management of their information seeking work. The particular benefits of an individual structuring task – intentional or unintentional – vary widely. For instance, it may support the user's selection of further documents, identification of the most relevant material from the documents already retrieved, identification of the key points to make in writing that builds upon the documents retrieved, or any other task. Information structuring is a behaviour, not a procedure with a single understood outcome.

As will be seen, research into information structuring is at an early stage. User studies have identified its presence in traditional, physical information seeking environments [Kidd 1994][Malone 1983]. Some simple computer software has been created to support information structuring or exploit its concepts [Marshall *et al* 1991][Mander *et al* 1992], the connection between information structuring tools and the broader electronic information environment has been weak, and usually entirely absent.

Research in physical environments has suggested that the interplay between information seeking and structuring is complex [Kidd 1994][O'Day and Jeffries 1993]. The form of the interplay between information seeking and information structuring is influenced by the available means of performing each task. Its form in an electronic environment within which both tasks are supported is not known, as no such environment has been created and evaluated. Therefore, the advantages and problems that emerge in a combined electronic information seeking and structuring environment is an open research question.

Information seeking needs resources from which information can be extracted. One such resource is a digital library. Digital libraries (DLs) are electronic repositories for digital documents. They mirror the function of physical libraries, and often the form of retrieving documents also reflects traditional library forms. For example, documents can be accessed by topic – as with the classification of books under the Dewey Decimal system. Digital libraries are an established topic of research in computer science, with several established conference series and published journals. Support for information structuring in digital libraries has received little attention, and I will use digital libraries as a context within which information structuring can be studied.

In this thesis, I investigate both how information structuring could be supported in a digital library, and what advantages arise from that support.

1.1 Spatial Hypertext

Computer scientists have investigated the support of information structuring activity in a digital environment. In Spatial Hypertext, researchers have designed different systems to support the organisation and structuring of document collections. However, Spatial Hypertext research is a small and specialised field. As I have already noted, spatial hypertext systems have little or no connection with systems that support the more traditional areas of information seeking – i.e. repositories from which documents can be retrieved [Shipman 2001].

Spatial hypertext is a field within the wider topic of hypertext. Hypertexts are systems where documents are connected by active links whereby activating a link in one document takes the user to another. The ubiquitous example of a hypertext is the World Wide Web.

A recognised subject of research in hypertext is what is termed "computation over hypertext" – this is where a computer system processes the links between documents, and the documents themselves, in order to create models across documents that support some other task. The PageRank algorithm [Brin and Page 1998] whereby the number of links to a document on the web are used to evaluate its importance, and subsequently its rank when displayed in a search result list, is an example of the use of computation over hypertext.

In the context of spatial hypertexts, computation over hypertext is at an early stage of research [Shipman 2001]. In this thesis I will introduce and briefly evaluate one example of computation over spatial hypertext – a technique that supports information seeking by exploiting the information structuring carried out by the user within the spatial hypertext.

Another spatial hypertext research question is how to represent the wider information environment in a spatial hypertext. Given the low level of experience in this area, and the intention to connect a digital library and a spatial hypertext, this issue is of clear relevance to this thesis.

1.2 Digital Libraries

Digital libraries have already been briefly introduced as electronic information repositories. The features of a digital library commonly reflect those of a digitally catalogued physical library, with online digital documents replacing printed material. A common additional feature in a digital library is searching across the full text of some or all documents in addition to the catalogued information of each document (e.g. title, author, etc.). Some libraries also support the repeated searching of the library with the same search each time that new books are added to the library.

In their support of information seeking, digital libraries are influenced by existing library practice and by available computer technology and science.

Computer science research has not covered every aspect of information seeking with equal effort or equal success. Much of the research done to date focuses upon the challenge of retrieving documents from the library – i.e. the means by which a reader may identify material relevant to their needs. This area of research is often termed "information retrieval" and it is a classical computer science research area. Commonly, the selection of material is achieved by receiving from the user a description of their need and then comparing the text of that description against the text of each of the documents that the user wishes to search across. Information retrieval is a challenging and widely researched field, and its influence on digital libraries is strong. The fact that electronic library catalogues have facilitated the speedy discovery of documents is of potential interest.

This bias in computer science somewhat reflects the emphasis of technology use in physical libraries – catalogues are maintained at great expense and with substantial technical support, whereas other tasks are often supported via simpler and less costly media. For example, reading desks support information structuring by providing a space for placing books and reader collaboration perhaps through the provision of group study rooms.

These imbalances in computer science research and physical libraries are reflected in digital libraries. However other areas of information seeking, such as information structuring, are still worthy of attention. Though the definition of what is a digital library is open to debate, the wider issues surrounding information seeking are addressed in papers such as Marshall *et al*[1999] and Shipman *et al* [2003]. Information structuring and spatial hypertext research has exerted some influence upon visual interfaces to digital libraries [Furnas and Rauch 1998][Hendry and Harper 1997], and thus this thesis builds upon existing research in digital libraries by focussing more directly upon the role of information structuring.

Information structuring, as part of information seeking, has a clear relevance to digital libraries. This thesis studies the addition of an information structuring environment – a spatial hypertext – to a well-accepted digital library system.

1.3 Evaluation Techniques

Both digital library and spatial hypertext research depend strongly upon other fields of computer science. The most relevant fields of research for this thesis are information retrieval, information and library science, information seeking and human-computer interaction. Information retrieval is a classical computer science research area that studies the accuracy with which documents on a certain topic can be retrieved from a larger set of documents. Information science and library science research is another long-established field that investigates the optimum means of providing readers with effective techniques for finding material within library systems and processing that material once found. Information seeking is a field within computer science that approaches many traditional information science problems with a computer science rather than a library science methodology. Finally, human-computer interaction explores the means by which users can be achieve effective interaction with computer systems, usually in terms of minimising time or effort costs.

In the evaluation of my combined information seeking and structuring system, I investigate its acceptability from a human-computer interaction perspective. The issue of how to represent the digital library in a spatial hypertext is a user-centred concern. The workflow between information seeking and structuring in the combined environment is also of interest, and in identifying patterns of user behaviour, I use an information seeking research viewpoint. Information retrieval techniques are used to identify certain aspects of the user organisation of documents (e.g. consistency) and to support the evaluation of some of the spatial hypertext research questions mentioned in the earlier spatial hypertext section.

1.4 Summary

This thesis investigates the connection of a digital environment for information structuring – a spatial hypertext – with a digital environment for document retrieval – a digital library. The use of this combined environment, called Garnet, is studied to identify any similarities in the role of information structuring in a purely digital environment and the known patterns in physical environments. Similarly, the known patterns in spatial hypertext use are compared against what is observed in the use of Garnet, where general information seeking can also be observed. In addition to the observation of the use of the system, the perceptions of potential users are acquired, particularly the benefits that they perceive information structuring providing in an electronic context.

As well as investigating the behaviour of users in their use of the system, I also study the potential benefit of information structuring in supporting wider information seeking, particularly information retrieval – i.e. the interactive discovery of documents. The artefacts of information structuring – i.e. the organisation of documents created by the user – may yield implicit information on the topics of interest to them. Researchers in spatial hypertext have noted the potential exploitation of the artefacts of information structuring. However, spatial hypertext research has not previously investigated this issue (for practical reasons that will be discussed later), and neither has information retrieval. Information retrieval researchers have used formal human-created topical classifications and other topical structures to support information retrieval. However, the informal structures that are seen in the course of information structuring have not been considered for the same task. Using the individual's topical classification within a spatial hypertext to sort later retrieved documents addresses questions in two research fields.

Chapter 2 – Literature Review

2.1 Introduction

This chapter reviews the current understanding of computer science research regarding information seeking, and introduces some related material that will be discussed more fully and in detail in the following chapter.

It is the intention of my research to create tools to support long-term interaction in a digital library. Where users are engaged in long-term ongoing activity in a digital library, rather than in disjoint searches over a period of time, they are likely to be an "information worker" who produces and processes information as well as being a consumer of it.

This chapter first briefly touches upon some research upon information work that sets some context within which information seeking occurs. Then information seeking is discussed in detail, starting with the introduction of a number of different models of the information seeking process.

The discussion of these models is followed by an outline of the two main methods for performing information seeking in an electronic environment – searching and browsing. Searching is classically defined [Marchionini 1995] as information seeking that exploits 'analytical' methods where the user must identify descriptors for their information need and articulate those in an explicit manner to a retrieval system – e.g. through the selection of search terms input to a retrieval engine. In comparison, browsing exploits the ability of a user to recognise appropriate descriptors for their need, and to locate documents through a series of selections of those descriptors. In each case, a section will study how each method (searching or browsing) can be improved or facilitated by processing performed by the computer.

A brief summary of the information seeking models and methods concludes the review of information seeking.

The remainder of the chapter discusses information structuring in particular, introducing spatial hypertext as an information structuring tool.

Finally the chapter will conclude with a review and summary of the questions and issues which are outstanding, and the current understandings which could shape and support a response to those problems.

2.2 Information Work and Information Workers

A number of researchers have endeavoured to provide an insight into the pattern of the work of information workers. For example, Goh and Leggett [2000], suggest a four-phase framework for this sort of activity:

1) Acquiring – locating useful documents through searching, browsing, etc.

2) Structuring – organising acquired documents to make them more useful.

3) Authoring – the creation of further material.

4) Publishing – the dissemination of authored documents.

Similarly, Ben Shneiderman [2000] suggests a similar four-phase cycle consisting of "Collect", "Relate", "Create" and "Donate". The sequences and roles of each phase are closely related to that suggested by Goh and Leggett.

I am not concerned with the Authoring and Publishing aspects of these processes, which leaves the tasks of Acquiring information and Structuring information. The more common term for what Goh and Leggett term "Acquiring" is "Information Seeking", which I will use for the remainder of this chapter.

It is my hypothesis that these two tasks, information seeking and information structuring, are deeply interconnected, and a tool can be created which will encourage the mutual support of activity in each task.

I will now look in turn at information seeking and information structuring and discern the requirements of users in each task, and relate those to the performance of existing systems which have been created for digital libraries or similar electronic information systems. The chapter will then close with a summary and review of the conclusions drawn and questions outstanding.

2.3 Information Seeking

Information seeking is the task of locating documents or other materials that answer an unfulfilled information need. Information seeking is a complex task that can be achieved by using many different strategies and resources. Many of the existing strategies and resources do not depend upon the use of electronic or computational devices. Others have become heavily dependent upon such systems. Clearly digital libraries, with their potentially large resources of information, can support information seeking by the effective delivery of quality materials to a user. However, the most effective means and manners by which they could do so is a matter of much research.

In this next section, I will review the patterns of human behaviour which have been identified in the study of information seeking, and the methods which have emerged to support this activity.

2.3.1 The Information Seeking Process

If one is to understand the requirements of a user in any task, one needs to understand the task in detail; resources required, the time sequences of the process through which the task is enacted and fulfilled, and the decisions and work which are involved. In this section, I will discuss a number of information seeking models proposed by researchers such as Marchionini, Ellis and Bates.

Each of these models focuses on different aspects of the information seeking process, and uses different types of model. There are three main forms of model: process-oriented, behavioural and cognitive. Marchionini's model takes a process-oriented approach that identifies a set of actions, their sequence and iteration. In comparison, Ellis' behavioural model identifies different 'behaviours' that can be observed in the work of information seekers. The sequence or interdependence of different behaviours is not captured. Behavioural models also avoid reasoning about the user's decision to choose one or other behaviour. Modelling the decisions of users is a key aspect of cognitive models that separates them from both behavioural and process-oriented models.

The models in this discussion take either the process or behavioural approach, though cognitive models will be mentioned later in this chapter. Due to their fine-grained model of the decision making of humans, cognitive models focus on small-scale, localised behaviour. In this section, I will be focussing on larger-scale behaviours that at present are only captured by either process-oriented or behavioural models.

I will now study the most commonly cited model of the information seeking process, by Marchionini; alternative models from Ellis and Kuhlthau and the findings of others that suggest limitations or shortcomings in the classic Marchionini model will follow this to flesh out a broader view of the information seeking process.

2.3.2 Marchionini's Model

A well-accepted model of the Information Seeking Process is that introduced by Gary Marchionini [1995], which describes the process as having eight sequential stages:

1) Recognition and acceptance of an information requirement
2) Definition of the information problem
3) Selection of an appropriate source which might address the problem
4) Formulation of a query
5) Execution of the query
6) Examination of query results
7) Extraction of information from result documents
8) Reflection on the process

The process is highly iterative; for instance, the examination of query results may lead to the query terms being reformulated several times. Shneiderman [1998] , Belkin [1993] and Salton [1989] draw a similar model of the process, differing only in the emphasis of detail given at the beginning of the process and also at the end of the process; the progression from Selection to Extraction are the same.

In addition to observing these separate stages in the process, Marchionini also observes the process to be not simply systematic, as a high-level process description suggests, but also highly opportunistic as the searcher identifies new leads or opportunities.

2.3.3 Ellis' Model

Another, very different, model is a behavioural model from David Ellis [1996] which identifies eight primary information behaviours:

1) Starting: Identifying initial sources of interest
2) Chaining: following and connecting new leads in an initial source.
3) Browsing: Scanning known sources for information of interest
4) Differentiating: Assessing and organising sources in terms of theme, usefulness, etc.
5) Monitoring: keeping up-to-date on an area of interest by regularly following specific, known sources (*e.g.* journals).
6) Extracting: Systematic evaluation of sources for material of interest.
7) Verifying: ensuring that information discovered previously remains accurate and reliable.
8) Ending: Concluding activities, summarisation, etc.

The model has been modified and altered several times. For instance, in Ellis [1997] various items are renamed and refined, and another step was added to the process called "Filtering" where personal objectives and influences are used to increase the quality of interest of the materials being used and sought for.

It may be noted that in Ellis' model, information structuring is explicitly present within the information seeking process itself particularly in the differentiating phase, whereas it is not represented in Marchionini's model, or those many similar models from Shneiderman, etc.

In Marchionini's model, querying is the only form of information discovery, whereas Ellis' model contains a number of discovery activities (monitoring, chaining, browsing). However, certain aspects of the two models show clear similarities. For example, the extracting behaviour in Ellis' model closely corresponds to the extraction stage (7) of Marchionini's model.

2.3.4 Kuhlthau's Model

Carol Kuhlthau [1992] introduced a model that portrays information seeking as a six-stage process, each of which is associated with a set of thoughts, feelings, actions and strategies. For example, the 1st Stage – Task Initiation – is associated with: the thoughts of contemplating the given assignment and considering possible topics; the feelings of apprehension and uncertainty; the actions of discussion with others and browsing a library collection and the strategies of brainstorming and discussion. As the seeking progresses, the task becomes more specific in the information being sought and the work engaged in more detailed, before the outcome becomes clear, and the information worker's experiences resolve into satisfaction or disappointment.

The six stages are:

1) Initiation
2) Topic Selection
3) Prefocus Exploration
4) Focus formulation
5) Information Collection
6) Search Closure

Though it may be tempting to tie each of these to a part of Gary Marchionini's model, the formulation, execution and evaluation of individual searches which forms his model would be repeated several times within a number of Kuhlthau's stages. For instance, in Stage 3 (Prefocus Exploration), searches will be used to identify topics which have an appropriate level of available information and to discard those with poor resourcing. The same stage would also include recording bibliographic information and taking notes. This recording activity would trigger several later iterations of searching and querying. Thus Kuhlthau's model depicts seeking as a complexly threaded, multi-faceted activity through abstracting above the individual queries which are very much the focus of Marchionini's model.

Structuring activities occur across Kuhlthau's model – both in the construction of topics and themes in the early stages, and in the organisation and presentation of final, chosen material in the latter stages of Information Collection and Search Closure.

2.3.5 Alternative Models

There are many other models of the information seeking process, which provide alternative methods of analysing and partitioning the work. A number of these will now be briefly reviewed in relation to the procedural and behavioural models just described.

Marcia Bates [1989] shares Marchionini's observation of information seeking as an opportunistic process. Bates makes two major observations of the information seeking style observed in users. First, discoveries made in the process change the information requirement continually. Secondly, that the solution to the original need is not found in a final set of documents, but rather in the accumulation of choices and incidental information discoveries made during the whole process.

For Bates, the process is not best modelled by a sequential model as suggested by Marchionini, but rather each stage of activity is better described by a behaviour, as suggested by Ellis. Unlike either Ellis' or Marchionini's model, Bates emphasises the artefact or event that triggers a change from one strategy or behaviour to another, rather than the strategies or actions undertaken as a result.

Influences that provoke such changes vary from events within the information system such as the discovery of a document, through to social effects such as discussions with colleagues. The elements which cause a search to be suspended or terminated are seen, however, as less determinable in the abstract, but determinable to some degree in a specific instance.

O'Day and Jeffries [1993] echo Bates' findings, reporting that activities such as review, searching and ordering were complexly interwoven when users were engaged in search tasks. They also observed that regular and continuous note-taking and progress review throughout a set of searches acts as a significant influence upon the future direction of seeking and the effectiveness of the final outcome. They do not suggest a model as such, though their findings strongly echo the emphasis on artefacts and triggers already noted by Bates.

More recent observers such as Hendry and Harper [1997] and Cousins [1997] also challenge the procedural nature of the process. These researchers again emphasise the triggers of activity in the seeking process, and the need to capture the more unstructured activity that occurs within and around the information seeking carried out by users of search systems.

All these alternative viewpoints on the model carry a clear common theme; they emphasise the role of artefacts and communications that occur within the duration of the process in determining the subsequent course of action of the process as it is enacted.

2.3.6 Extended Information Seeking

Much of the work patterns observed above can be used to explain short-term and long-term information seeking. However, there are a number of patterns of information seeking which are unique to people whose work requires long-term information discovery and tracking.

The first examples are routed searches; these are searches which are automatically run either periodically or when the collection over which the search is run changes. This is often used to track a continually changing area of continuing interest, where keeping up to date is important.

Routing is well discussed in an information seeking context by Oard and Marchionini [1996]. There are a number of different implementations of routing; from a normal text search, to complex user profiling (either automatic or manual control) using Machine

Learning or other techniques. The technological aspects are beyond the scope of this work, but there is a clear need for the support of this sort of work in any environment which supports ongoing information seeking. Routing bears a clear relationship to the "monitoring" activity reported by Ellis [1996], though it is realised computationally rather than by repeated human action, though it is poorly represented in the classic procedural model of Marchionini.

Ellis also identifies another behaviour, verifying, which is used to ensure that information previously discovered remains valid. As the common intention of routing and verifying is to ensure that the seeker retains a current and accurate knowledge of their field of interest, verifying will probably also be achieved through what would be described as routing activity, the discovery of new information that invalidates or verifies the old.

Thirdly, researchers such as Bates [1996] and Kidd [1994] have observed knowledge workers who were engaged in ongoing information seeking. A high degree of dependency was demonstrated on a number of forms of note-taking and storage which the workers hoped would guide future work. Similar insights were made in the course of O'Day and Jeffries studies, and Bates' earlier work [1989] mentioned above. Indeed, in comparison Bates [1994] observes a higher dependency upon this method of work support, and instead of simply causing reactions in the shorter term, work artefacts are actively used to direct and manage later efforts.

As with our observations of artefacts in relation to Ellis' and Marchionini's models of information seeking, these artefact-centred influences are not easily included in the process, being neither a behaviour nor a task. Clearly, however, one needs to consider how artefacts can be used to support the information seeking process.

2.3.7 Summary

The information seeking as a process has been described as a sequential set of actions. However, these actions are influenced by and controlled through other aspects of the process such as artefacts created by the seeker and influences from external factors such as colleagues. If the complete process is to be effectively supported, it is necessary to provide some means for facilitating not just the actions such as query execution and formulation, but also techniques for the user to do the sort of tracking and guiding actions observed by Bates, O'Day, and others.

Ellis' work is illuminating because of its different separation of information seeking activities into behaviours. How these relate to traditional information seeking models as typified by Marchionini will be discussed further in later sections; however, some behaviours are explicitly ones conducted in the context of extended information seeking, and so are of particular interest.

At a smaller scale, Ellis, Oard and Marchionini, all observed repetitive cycles – *e.g.* a sequence of reformulations of a query in order to achieve a satisfactory set of results.

These tactical cycles themselves occur in longer-term cycles where users return repeatedly to a subject to identify new documents of interest, and validate the assumptions previous work has led them to. In all this second work, a high degree of task awareness by the user is required, and results in the various form of future planning observed by Kidd, O'Day and Jeffries, with artefacts again playing significant roles in co-ordinating over time.

Therefore, human-generated artefacts are a key element of supporting the information seeking task and a person's management and control of their own information work. This requirement seems to be likely to be more significant for those performing information seeking over an extended period rather than simply performing a quick one-off search.

2.4 Methods of Information Seeking

The Information Seeking task has been the focus of much research as the availability and use of electronic information systems has expanded over the last 20 years. Different methods of discovering information which fulfils a requirement, such as browsing, searching, filtering etc. have emerged in the progress of time. Later experience has indicated that what were once seen as different methods (*e.g.* searching and filtering) are in fact substantially similar [Belkin & Croft 1992]. Therefore, most current authors, *e.g.* Marchionini [1995], Shneiderman, [1998] divide the information seeking task into two primary methods:

1) Analytical seeking (Search): where the user inputs to the system some expression of their information need and the system responds with related documents.
2) Browsing[1]: where the user is presented with a set of documents, usually structured in some hierarchy, which they navigate and read to identify appropriate literature or data for their requirements.

There are significant differences both in the technical infrastructure and user skills required by each approach. In Search, the user must articulate their information need in an abstract form with precision and clarity if they are to obtain useful documents, but if successful, they can subsequently select from a small number of documents which are highly likely to be relevant. Conversely, in Browsing, the browsing structure is there to support the user's selection method by presenting selective themes within which the user can browse. The user is dependent upon the degree and quality of fit between the structure and their information need, but again if successful can hope finally to choose between a small number of relevant documents towards the end of the task. The distinction is ultimately

[1] N.B. browsing in this usage is not the same as in Ellis' behavioural model above; in Ellis' model, it is always a casual, scanning activity. It is often as a information seeking strategy far from casual, and may be used to support virtually all of Ellis' behaviours.

between a demand on user articulation in the former method (searching), and user selection in the latter (browsing).

Naïve or novice users are likely to depend highly upon a supportive tool, and have thus been observed to depend heavily upon a browsing approach, whereas advanced or experienced users are more likely to utilise highly sophisticated techniques within a searching approach. As observed by Marchionini [1995] and Bates [1989], often the strategies combine when a user changes the focus of their information seeking; if an expert in one field needs to investigate material in an unfamiliar area, they will often start with a browsing strategy to orient themselves in the language of the unfamiliar area and later move to exploiting their transferable searching skills to identify documents of specific interest with greater precision.

The information seeking behaviours observed by Ellis and others generally are fulfilled by either or both of these two approaches; for instance if one were performing a monitoring task, search and browse would be useful, whereas the browsing behaviour is naturally much more likely to emphasise the browsing method.

As the two methods differ in a number of ways, I will now discuss each separately in turn.

2.4.1 Searching

Searching is the predominant form of information seeking used in most large information systems. As already stated, the user needs to generate a description of the material they require, and the information system then returns a selection of corresponding material. We have already described the process above, and the "search" method occurs across stages 4 to 8 (formulation to reflection/refinement).

The most significant challenge in searching is the formulation of the search terms for the query. A poor selection can result in any of a large number of unsatisfactory outcomes; no or too few search results due to an overly specific search, too heterogeneous a set of documents as a result of the use of common or ambiguous words, for example.

Shneiderman and Byrd [1998] list a number of aspects of each of the discovery stages described by Marchionini in his model, and they divide the process up much as Marchionini does with the exception of combining review of results and extraction into a single "review" phase. I will now review the different parts of the discovery phase of the Information Seeking process in detail and in turn. Later, I will look at how the different phases have been supported by software tools and how the relationships between these different phases have been exploited to improve the total performance of user and information system.

Query Formulation

Shneiderman and Byrd [1998] list a vast series of difficulties that can be faced as a result of both the technologies being used and the material being searched. Clearly users who use a larger variety of material sources and using a wider selection of search tools will experience more difficulties than those for whom at least one factor is stable. As I am considering users with a substantial dependency upon information systems, they are more likely to have to use many tools on many different sources.

Those who have been creating systems to support continuous activity in digital libraries have addressed this challenge. For example, both the DLITE service [Cousins 1997] and SketchTrieve [Hendry and Harper 1997], automatically normalise the interface for different services and sources.

There are a number of different factors that make up the formulation challenge, which I will now address in turn.

Expression Syntax

Different systems allow the user to control the search by using operators. For example, the use of the '+' operator before a term commonly requires that any document in the result set *must* contain at least one instance of the following word. Other operators may express boolean logic, encapsulate phrases etc. These operators tend to vary between different search engines. As observed by Marchionini, the selection of operators at the *semantic level*, i.e. independently of syntax, is in fact itself a challenging task that only more experienced and sophisticated users can successfully exploit.

Query Terms

The user also needs to select appropriate words to express the goal of their search. Many problems face the information seeker; particularly *polysemy*, where one word has several meanings, and *synonymy*, where many words have the same meaning. Only by combining appropriate words can the user hope to palliate these problems, often in combination with the use of carefully selected control expressions.

Query Sensitivity

A third set of controls is over the sensitivity of word matches; at the most restrictive, the query should match words by case and exact matching; at the most liberal, the query may match regardless of case and with alternative word endings (i.e. stemming). Other options may include matching initials in the case of names, abbreviations and anagrams in the case of technical terms.

Summary

Clearly creating appropriate query requests is a complex task; normalisation across search engines, as described briefly above, provides a degree of simplification over several collections, but clearly does not remove the need to select query terms, sensitivities and expressions with care. The matters of expression syntax and query sensitivity have already been addressed in many digital library systems, *e.g.* DLITE [Cousins 1997], and so will not receive further attention.

Query Execution

The performance of the actual search can be instigated and respond to the user in a number of modes; traditional on-line searches are performed immediately at the explicit instruction of the user. Systems such as that described by Doan *et al* [1996] provide "query previews" – approximate representations of the search results – as the user sets the parameters for the search, responding immediately to each change. Another possible response is for the search to be run at a later date, usually at set intervals or when the collection is updated; this approach is traditionally called "routing".

However, most commonly the query execution is not exposed to the user; the results are given to the user at the end of the performance of the search.

Result Review

When the user receives the results of their search, there is a significant probability, particularly when a search is done against a large information source, of there being a very large number of results. Shneiderman and Byrd [1998] suggest that the results should be capable of being ordered in a number of different ways, *e.g.* sorted alphabetically, by date, relevance, etc.; that the user should be able to select which information about the individual results is displayed, and they point out the power of methods which allow the user to interact dynamically with the results using post-processing tools which are interactive or analytical.

One reported problem is that users find it difficult to identify how the documents returned from a search relate to the query that they gave to the system [Muramatsu and Pratt 2001]. In part this is in fact an extension of the problems of Query Formulation above; there is an uncertainty as to the effect of any given operator in or option selected over a particular query. A number of systems are reported, which I will discuss later, which try to give some better insight to the user of the relationship between the query terms and the query results.

Query Refinement

Once a search has been executed, the user may need to improve it. Shneiderman and Byrd [1998] identify two requirements for users. Firstly, users should be able to recall previous searches easily (to track the terms and combinations used before). Secondly, users perform significantly better when supported by relevance feedback, particularly in the interactive form implemented by Koenemann and Belkin [1996], to which I will return later. Certain elements of refinement can also be supported at the review stage, with indication given for terms which are mis-spelt, or are stop words etc.

Much of the work described in refinement is in fact using the outcome of a search to support the effective reiteration of the formulation activity described above.

Summary

Each of the different stages of the search process presents its own problems to the user. For instance, in query formulation, they need to derive precise descriptions for their information need and in result review, they must interpret both the results of the search as a whole, reason about the outcome and select individual documents. Support can be given for each of the stages, but how this is to be achieved is seldom obvious. Often, the underlying issue concerns the effect of a query term or option upon the search outcome – be it anticipating that effect in advance or judging the actual effect when the search results are received. In some cases, support can be achieved by technical means, and these will be described in the following section.

2.5 Improving Search Performance

My main purpose in the work presented in this thesis is to support a wider variety of the work associated with information seeking. This wider range of information seeking tasks may be able to provide additional contextual information to support the specific activity of search. Over the years, many researchers have endeavoured to assist users in the search task. The existing knowledge of and tactics for supporting search provide the foundation for developing new techniques. Therefore, in this section I review the key techniques for improving search that exist at present. These techniques and the paradigm within which they have developed will later be used to identify potential techniques for improving search that exploit the wider information seeking context that I present later in the thesis.

I have just listed a number of separate problems that are faced by users in some of the tasks within information seeking, particularly those in the search phase of the process. I will next review a number of methods used to improve the performance of searching. Before that is done, though, it is worthwhile identifying what the main relevant questions are.

The main hypothesis at this point is that the user's activity in structuring information could be used to support the total performance of the same user with an information system. I have not yet discussed evidence to suggest how this could be done, nor if there is any evidence that a similar pattern of using the outputs of some earlier work may be a useful means of improving the performance of a system.

Therefore, it is these two questions which will be addressed: firstly how search performance is improved generally, and secondly how organisational information in particular can be used for that purpose.

Before moving on, however, it is worth bearing in mind that the Query Execution phase is generally only altered by the inputs given to it by the Query Formulation stage; or in other words it is of itself a purely deterministic, computational process that is controlled by its inputs. Therefore, it will not be discussed directly, though the construction of its inputs will be.

2.5.1 Improving Search Refinement and Formulation

Having just described the problems present in the searching task, it is worthwhile observing some of the support tools which have been created to improve search performance. Many of these improved search systems rely in some form or other on improving the feedback from the refinement activity to the formulation activity. As the refinement process is a complex one, let us be clear about which aspects are of concern.

Query refinement, as described by Shneiderman and Byrd, actually covers two separate tasks as defined by Marchionini:

1) Extraction of information from the documents selected at the review phase.
2) Reflection, altering an information goal in respons to information discovered in extraction, or observing failures in the query formulation phase.

Two common techniques for improving query formulation are *Query Term Expansion* and *Relevance Feedback.*

Query Term Expansion

In the case of query term expansion (QTE), the user is presented with a list of common terms in the search results which may assist in improving the precision of the search. The user selects from these words as they feel appropriate, and they are added to the query terms. A simple but extreme form exists as Automatic Query Term Expansion (AQTE), where the search system automatically selects a number of words without interaction from the user and re-runs the search immediately, presenting the user only with the results of the modified search.

Query Term Expansion has been proven in a number of search scenarios to improve the precision of searches, with the interactive form generally providing better results than the automatic form [Koenemann & Belkin 1996].

When a user executes a series of queries, the suggested term list is regenerated. The user then again selects those items that they deem appropriate; as the list varies between one query and another, the selections are cleared at the commencement of the new query.

Relevance Feedback

Relevance Feedback (RF) works on a similar basis to QTE, but in this case the user selects *documents* rather than words. An individual document may be given approval or rejected as the user determines, and the search system responds to this by selecting documents which are similar to those approved of and rejecting those which are similar to those which were rejected. Again, this has proven to be beneficial to the performance of searches, with evaluations reporting improved precision in the search, *e.g.* [Salton 1990] [Hancock-Beaulieu 1992].

As with QTE, the selections made are cleared between searches; as the context and purpose of each query is not known to the computer, this is important, as otherwise what would correctly be a document of interest in a new query may be marked as undesired as a result of selection in a previous, unrelated search, or vice versa.

Other Methods

As well as RF and QTE, further systems exist which can permit the user to alter the result set through interacting with the search results. Often, methods use visual graphical manipulation of word or document representations on a display, *e.g.* the system described by Anick *et al* [1990] and Lyberworld [Hemmje 1994]. The manipulations alter the position of documents dynamically, the user being unaware of the nature or form of the computation occurring. In many cases, the algorithms used to place the documents are based upon traditional information retrieval approaches already discussed, with additional modifications. Lyberworld alters the weights of individual words dynamically on a continuous linear scale, which is not usually available when controls are text-based. This approach permits word weights to be varied over a larger number of discrete values than is usually the case with the text-based controls.

Summary

All the techniques described above for improving search formulation are used to support "one off" searches; the effort used in one search is seldom used to support later searches, unless, for example, the user edits the results of the full terms created using QTE. Certainly, the benefits are unlikely to be directly carried from one session to another (though a user may re-select similar terms or documents for a related search at a later date). Clearly, some reduction in effort could possibly be achieved by carrying the benefits of one search into another. However, this must not preclude documents from being falsely labelled as either highly relevant or irrelevant in any given context.

The main feature of these approaches is that each endeavours to support the challenging task of term production. Term production is undertaken by both the user and the system, rather than in the traditional manner of by the user alone.

Aspects such as query syntax alterations to force terms, etc. are beyond the strategies which they support. They also are supporting the transition between iterations of the query formulation to query refinement loop in Marchionini's model, exploiting the outcome of the previous iteration to suggest new terms for the next.

2.5.2 Improving Search Refinement and Review

A problem identified above in the Query Review task was that users find it difficult to map between the query they gave to an information system and the results that were returned.

A number of systems have been created which endeavour to address this problem, and a wide variety of approaches have resulted. The anticipation of the following systems is that where the user is able to detect the relationship between their current search and the resulting documents, they can either validate the query formulation used for it, or alternatively diagnose some anomaly between what they expected to occur and how the information system interpreted the query. Thus, through evaluation and analysis by the user, supported by information exposed by the information system, their query formulation improves.

Dynamic Weighting

LyberWorld [Hemmje 1994] is a visualisation system which endeavours to permit the user to adjust the significance or weighting given to the terms in their search by manipulating visual representations of the search terms. In response, the document representations are attracted towards the terms according to their given significance. This is similar in fact to the use of Relevance Feedback and Query Term Expansion just mentioned above, except that in this case the activity is perceived as being separate from the execution of the search itself (i.e. the total set of results does not change, merely their relative weighting).

Term Feedback

Another approach is to give term feedback (N.B. this feedback is from the system to the user, not vice-versa as in relevance feedback above), indicating to the user how many times a word was found in a document and sometimes as in the case of systems such as TileBars [Hearst 1995] the position of the term within the document is also given. In the latter case, the user can gain some idea of the degree of co-occurrence of words within a document, *e.g.* whether they often occur in the same section as each other, or whether they rarely do. Again, the search result set itself is not directly affected, and in the case of most Term Feedback systems, there is no scope for manipulating the ranking of the result set either; it is foremost a system for improving the scrutability of a set of results.

Result Segmentation

Search result sets can be partitioned into different topical groups. One particular approach is termed 'clustering', and I will present the concept of clustering systems more fully in the next section. Similar effects have been created using other systems, such as self organising maps, in the case of reviewing query results. Once more, to be clear, the search is not re-executed; rather the results are compiled into non-intersecting sets of documents, often presented in a 2-d visualisation. The user can usually focus on a subset of the documents (visually) and obtain more detail about a particular subset, its internal divisions etc. One example of this approach is the SOMLib digital library interface by Rauber and Merkl [1999].

Many alternative visualisations, representations and partitioning systems can be used to create the segmentation of results; for instance, the Hieraxes browsing system described by Shneiderman [2000b] uses segmentation through metadata items and the Cat-A-Cone system [Hearst 1996] exploits existing classification systems; the former system is highly graphical, whereas the latter can be used with a text-based presentation. However, an exhaustive list of tools and approaches is beyond the scope of this work.

The concept behind these sorts of result segmentation systems is that the user can focus upon the descriptors which seem most indicative of the information they are seeking; this is in fact eliding the user from a search process into a localised browsing process. Therefore, it is facilitating the sort of transition of tactics which is suggested by Bates above. The nature of browsing itself I will discuss later. It is interesting to note before moving on, though, that the performance of some systems such as Dynacat [Pratt 1999] has been significantly better than the superficially similar approach of clustering which will be described later.

However, a secondary benefit is that the user is also exposed (in a good system) to the descriptors which make good distinctions between the documents which correspond to their particular interest and the documents which are unrelated to their needs. Therefore, the user may be prompted to redo their search but with a new or alternative set of query terms.

Summary

Compared to the formulation support in the previous section, the primary function of each of these systems is to enable the user to gain a more complete understanding of how the given query relates to the resulting document set; improved query formulation comes through giving the user a better insight into the correspondence between their query terms and the resulting set of documents. The reformulation can either be through explicit actions performed by the user adding new query terms or come through hidden refinements which are produced through a browsing style activity (*e.g.* Dynacat) in which metadata is implicitly used to refine the search precision through selection in a hierarchy.

The former, like analytical seeking generally, will require more effort on the user's part; in the latter a low-cost browsing style activity is used.

2.5.3 Improving Search Review

A number of tools have been created to improve the review stage alone, without any intended direct benefit to other parts of the information seeking process.

The most common form of these is the use of document clustering techniques. In such cases, the user is provided not simply with the list of target documents for a query, but the documents are processed by the query system into sets of textually similar documents.

These sets are either generated independently of the query, in which case the same sets are used in each query; or alternatively they are generated dynamically on the query results alone, in which case the sets usually vary from one query to the next.

Most early evaluations such as [Cutting 1992] in fact used clustering as a browsing method rather than a tool for supporting search; these will be discussed later. However, Hearst *et al* [1996] evaluated the use of clustering to create browsing structures over documents found as the result of a search. Similar studies followed – for instance, by Zamir *et al* [1997] performed the same experiments on a system that used another clustering technique. I will now briefly describe the user experience of using a clustering tool over search results, and summarise the findings of these studies.

When a clustering tool is used on search results, the system will present the user with a set of labels, one for each topical group of documents that it has created. As the user selects one or more groups from this pre-prepared list, their task is, as in the case of browsing, more one of identifying matches between their information requirement and the words used to summarise a set of documents, than one of generating descriptors themselves. Having selected a group of documents, then that group is itself divided – clustered – into further topics, from which the user chooses in turn. By iterating this selection process, the set of documents is gradually reduced to a set that should closely match the interests of the user.

The evaluations of this technique have continued to use the same methodologies as information retrieval – the 'success' of a technique is indicated by its promotion of relevant documents over less relevant documents. The performance of Hearst *et al*'s system indicated that users who used the Scatter/Gather tool retrieved fewer relevant documents than than those users who interacted with a traditional ranked list. Zamir *et al* came to a different conclusion – their Grouper algorithm performed comparably to a ranked list presentation.

Considering only the retrieval of relevant documents, clustering seems to offer little or no advantage to information seekers. However, a secondary and unexpected benefit was observed by Hearst: the users benefited from a better comprehension of the range of material available in each information source they used. Zamir *et al*'s findings gave further evidence to support this hypothesis.

There are, therefore, two effects of these search review tools. One is direct; that users quickly elide from a directed, analytic method of identifying data onto a localised browsing task where they can respond to the stimulus of group labels, which demands a different, lower, degree of cognitive work. The other is indirect; that users seem to gain an impression of the thematic strands of the material to hand, which in turn could improve query formulation and refinement.

2.5.4 Extended Search Support in Digital Library Systems

In the section on the information seeking process above, we were introduced to users performing the same search over a period of time in order to track changes in a corpus of documents. In the same section, I observed at the same time that users often had rapidly changing requirements for information, and that they made notes in a number of different ways to control their future work.

A number of researchers have responded to this need of users to track the dynamic, changing element of information seeking which is not met by the repeated searches provided by routing. Examples of this are systems such as NaviQue [Furnas and Rauch 1998], SketchTrieve [Hendry and Harper 1997] and DLITE [Cousins *et al* 1997]. In each case, every search a user conducts is represented by one or more artefacts to provide some reminder or representation of the search. The artefacts assist the user in reviewing or remembering which activities they have already started or completed, and conversely the absence of artefacts for a task indicates implicitly that it has not yet been performed.

The evidence for the effectiveness of this artefact-centred approach is mixed; in the case of NaviQue, a series of iterative formative evaluations is reported, but their form, procedure and outcomes are not reported individually. In the case of the DLITE interface, only an observation with a small number of users was performed, and no evidence for any impact from the artefact approach is reported. Similarly, the developers of SketchTrieve have only completed a formative evaluation with five users. In this case, users were exposed to a small number of highly diverse tasks in a single session, which significantly reduced the benefit of any longer-term planning. However, even in this context of brief activity, use of what the researchers termed secondary notation, analogous with the use of artefacts to plan and co-ordinate activity was consistently observed. Benefits in performance, and any impact upon the progress of the information seeking process itself, are not reported.

Therefore, there are a small number of similar systems, each of which builds upon the artefact-centred approach suggested by research into information seeking systems. However, the benefits of artefacts upon user activity are not reported, and the systems are not compared in use to traditional information seeking systems. Clearly, the role and impact of artefacts in a library system remains an open question.

2.5.5 Search – Review and Summary

With a number of different concepts and across several separate systems, researchers have been able to use the support of the user's activity in one of our four tasks (formulation, extraction, review and reformulation) to boost performance of the system and user in another of the four tasks. Examples include: improved interaction with the user in the review process enhancing query refinement and formulation (*e.g.* relevance feedback); better support for distinguishing documents under review thereby improving the user's conception of the structures of collections; and review activity supporting the selection of appropriate documents in the extraction phase (*e.g.* clustering and self-organising maps)

In many ways it can be argued that searching is itself one method for improving the effective-ness of browsing, the information seeking strategy I will look at next, by limiting the browsing to a number of documents which are more likely to be of interest to the user than a subset selected at random.

A common theme for improving information seeking has been to leverage the capacities and strengths of one part of an information system to improve the performance of another part, often in concert with a small degree of interaction from the user.

Another common theme is that two strategies suggested by Shneiderman and Byrd have been dominant. In the first, the system's ability to support the user's term production is extended; in the second, the system constructs browsing structures to better facilitate the user's review work (extraction and review in Marchionini's model). In a few examples (e.g. some clustering systems), the two are combined to a degree, though often one is emphasised over the other (e.g. the user would have to manually enter cluster keywords as search terms).

Thirdly, although the effectiveness of artefacts of previous search activity may not have received deep evaluation in the digital library system in which they are supported, Shneiderman and Byrd argued for the provision of these as well, though again without conclusive evidence.

2.6 Browsing

Browsing is most commonly associated with casual investigation of a collection of documents, or with naïve users. However, it is also commonly found as a practice among experienced users too. As noted above, it provides very different demands upon the user, and in general it has often been evaluated to be much less effective than traditional textual searching *e.g.* [Campagnoni 1989].

As with searching, the support of browsing has been an ongoing topic of research. Different systems have been designed and evaluated for their different behaviours, advantages and shortcomings, and researchers such as Marchionini have endeavoured to probe beneath this to discover more persistent factors and influences.

2.6.1 Browsing Processes and Behaviours

If we first return to the information seeking process model of Marchionini, we will find that in fact the process doesn't well describe the behaviour of users who are browsing. The basic mode of browsing is selection rather than articulation, and reformulation is replaced by re-navigation. One product of this difference is that browsing patterns are much more influenced by the available structure which the viewer navigates, and researchers' observations about the user's experience of and activity in the process are usually behavioural in nature.

As just observed, a user doesn't formulate a query in the case of browsing, though they may choose one or more headings which they will navigate in greater detail and/or with higher attention than others. Early work by Furnas [1986] has been a continual influence on the practice of creating browsing structures; Furnas suggested giving the user an ability to focus on particular areas of a browsing structure, revealing more detail, whilst still retaining the ability to see the larger context which contains the area of interest. This came from observations of users, where browsing labels of a more general nature, at higher levels, were used by more effective users to guide the selection of navigational choices within an individual node. This firmly set the model of browsing in a hierarchical framework; the selection of each new grouping is a refinement within a parent collection or set, combining into one gesture query refinement and execution. Indeed, this merging of process steps has also been made by the hypertext community, where links have come to be seen as a specialised form of search [Golovchinsky 1997].

However, simple hierarchies are not the only browsing structures, and observations of them cannot fully explain browsing behaviours. The more chaotic network structures of common hypertext have also been used for browsing, and clearly Furnas' top-down model is inappropriate. An extensive range of papers by Card, Pitkow and Pirolli [Chi *et al* 2000, Pirolli 1997, Pirolli 2000] introduces the concept of *scent following* as a model of user behaviour when browsing in an information space. When the user detects the "scent" of relevant information, they then follow that trail until either they discover an alternative one, or their information need has been sated by any documents or overviews gained en route.

In the normal, hierarchical, browsing structures of libraries, this scent following is often in the first place a top-down affair, focussing on increasing detail. However activities such as Ellis' Chaining [Ellis 1996] may in fact be closer to the hypertext behaviours which inspired Card, Pitkow and Pirolli: in chaining, users follow successive connections between documents just as in a hypertext users activate a series of links to move from document to document.

Clearly, there are some similarities between both the scent-following description of activity and the browsing structures suggested by Furnas. Also, browsing can be through both hypertext-type networks such as references, or in a hierarchical classification such as a subject index.

So, models of browsing are more related to behavioural or cognitive models than the procedural model that we've generally seen for information seeking, and are often closely related to the browsing structures in use. The selection of a sub-category can be compared loosely to query formulation, should a mapping be required. The main impact of these different models, particularly that of Furnas, is upon visual representation rather than the en-action seeking process itself. There is little guidance as to how to create effective support for browsing though artefacts, the creation of classifiers, links or other system properties.

2.6.2 Browsing Tools

Browsing has been observed in its different forms, hierarchical and otherwise, to assess its relative merits as an information seeking method *e.g.* [Campagnoni 1989, Chimera 2000]. The general pattern of the outcomes is that hierarchical browsing provides better effectiveness than hypertext networks, and that in general browsing performs poorly when compared to query-based search retrieval.

Most often, hierarchies are created manually by an information manager or librarian. However, this can be a costly exercise, particularly for large yet homogenised specialist collections.

As a result, many means of creating hierarchies automatically have been created, the most popular of which form the clustering set of algorithms briefly mentioned above in the section on search. In the main, the resulting hierarchies are transient maps over dynamic sets of documents: for example, upon a query recently executed by one user. However, they can often be used with equal effectiveness to create persistent structures that are common to all users.

Personalised Browsing

It has been argued that the user should only be presented with the information that is of interest in order to reduce their cognitive load [Allen 1992][Marchionini 1995]. The logical impact of this on browsing is to create browseable structures which correspond to the user's own perceptions of their information task and how different topics relate to each other. The usual means of generating the data to facilitate the creation of such personalised structures is to create a profile to represent the interests of an individual user or a group of users with a common interest. The user profiles can be generated by two principal means:

1) Implicitly – the information system tracks the user's interests and actions and makes hidden judgements on these.
2) Explicitly – the user gives explicit feedback as to their interests.

It has been found [Manber et al 2000] that users tend to spend little if any effort in setting up any profiling system. This is consistent with the "paradox of the active user", observed by Carroll and Rosson [1987], where users move to immediate use of a system before any preparatory familiarisation or set up. Therefore most recent and current research into profiles is based upon *implicit* techniques. However, it is more difficult to accurately gauge users' interests implicitly than explicitly.

For example, say one uses the heuristic that a user who spends a long time with a particular document on the screen is interested in that document. If a user were distracted from the computer after displaying a document that is not of interest, a false positive indicator of interest would be given. On the other hand, if a user quickly closes a document which is not of current interest, though it may be useful at a later date or for another purpose, a false negative may be received.

So there are substantial problems for the system in correctly interpreting the user actions, and what significance to give them. As a result, often the hoped-for correspondence between the user profile and the user's interests for many implicit profiling systems is not as good as originally hoped.

However, profiles can improve with time; as with most "intelligent" systems, prolonged training tends to produce better results. The largest difficulties can come early in the training cycle, which may result in the user rejecting such technologies early.

Dynamic Browsing

As well as tools which rely on user profiles to create the browsing structures, others exist which are intended to work on any set of documents at the time of use. The most common forms are clustering systems, which group documents based on a textual analysis of their content. Documents with similar words are grouped together. Problematically, such systems tend to operate on a pseudo-random basis, so the same set of documents will often be clustered differently on each occasion that the clustering is performed.

This creates a problem that there is not a learnable structure which the user can eventually start to exploit. Contrariwise, an advantage is that a user can ask for a new attempt to be made to structure the documents if the current organisation seems difficult to comprehend. One solution to the issue of learn-ability is to create an initial structure when the set of documents is first brought together. However, if the set is later changed and the clustering is rerun as a result, the new structure may be very different.

The problem for dynamic browsing in the case of long-term information seeking is that its benefits are short-term and somewhat unpredictable. Evaluations carried out on the use of clustering systems as a means of supporting information seeking, such as [Cutting *et al* 1992] have indicated that whilst there are benefits in terms of comprehension of the overall content of a collection, there are disadvantages in terms of retrieval effectiveness. It may be useful within a particular search as an alternative information seeking tool, but such purposes are beyond the scope of this thesis.

2.6.3 Browsing and Information Seeking

So, it has been seen that there are numerous methods of creating browsing structures. An important question for us is how browsing fits into long-term information seeking. As has already been stated, novice users tend to prefer browsing over searching, and more experienced users vice-versa. Over a shorter timespan, it has been observed [Campagnoni and Ehrlich 1989] that as users retry an individual information seeking task, their preference shifts from browsing to searching; at the first attempt 75 % of users used browsing; by the 4^{th} attempt, 75% used searching, so searching is an important fallback even when browsing is preferred.

Contrarily, as I have observed in the Search section above, browsing often occurs in the review and extraction parts of the information seeking task when search is the primary mode. Both hierarchical and hypertext network browsing occurs in this phase of Marchionini's information seeking model. The different tools seen above, particularly dynamic browsing, tend to appear in the support of this end-game seeking work.

Furthermore, the routing search tasks outlined earlier also tend to be heavily reliant upon the same or similar types of user profiling met in personalised browsing. Clustering systems, implicit and explicit profiling, hierarchical and hypertext structures have all been used in combination with both routing systems and browsing. As already observed, routing corresponds closely to the monitoring type activities reported in Ellis' information seeking model, and also appear in some of the strategies reported by Bates; *e.g.* journal runs where a specific journal of interest is repeatedly used, which can be reiterated and extended with each new volume or issue.

2.6.4 Review

Looking back over the browsing and searching techniques involved in information seeking, it can be seen that what is used to improve or focus browsing is also used in the review phase of searching. Therefore, effective browsing tools may also yield positive benefits when they are also deployed to improve the review stages of search, and vice-versa. Browsing has been shown to be effective in smaller structures, and in the end-game of an effective search, the user should be focussed upon a relatively small and selective set of documents, so this synergy is to be expected.

User profiling is a technique that has clearly proven of more interest to those creating browsing structures than in the case of search support, where this technique is rarely encountered. A problem with profiling is the cost of constructing the profile, and the accuracy of the profile. Automatic methods for profile generation often suffer lower acuity, whereas users are reluctant to commit time to creating and maintaining a manually generated profile.

Artefacts of previous work, which were suggested as a possible support tool for searching, at first seem absent in the case of browsing. However, the trace of a user's activity is in fact an artefact, and one often used to generate automatic profiles. The explicit provision of a user's interests or other deposit to create a "manual" profile may also be compared to an artefact, if a specialised one which otherwise gives no benefit to the user.

This suggests the possibility that other user artefacts may be useful foundations on which to build browsing structures to support information seeking, either in the latter stages of searching, or for browsing.

2.7 Information Seeking – Summary

Classically, search and browse have, as seen above, been seen as two separate methods of information seeking. As has been seen, in fact browsing occurs in most cases in the later phases of information seeking when in search mode. Furthermore, a number of technologies which have been used to facilitate browsing, particularly over dynamic structures, have also been used within the comparable review phase of searching.

However, the range of means for improving browsing is much more limited than the means for improving searching. When using the analytical method of information seeking, there are many ways in which different parts of seeking are combined through the information system to provide superior effectiveness, or alternatively additional feedback or control is given to enable the user to input more precise query terms.

The use of text-processing methods has resulted in a number of effective automatic or semi-automatic methods to make searching more efficient. Some of these work within one part of the information seeking process, but most techniques work across and between two parts of the search. The use of explicit user activity was demonstrated by Koenemann and Belkin [1996] to be more effective than automated processes. However, contrarily, Belkin's later work [2000] observes that users prefer to have to give the computer system no input.

In the case of browsing, personalised information structures created though explicit user instruction have similarly met with a poor rate of adoption. Again, users seem disinclined to provide extra information for a better result. On the other hand, automatically generated profiles can suffer inaccuracies as a result of having to double-guess the user, whilst set structures can provide poor effectiveness when they ill-match the purposes of the user.

Therefore, users clearly resist any requirement of explicit inputs to the computer system in order to provide more effective results for their personal information needs.

The use of artefacts by users to shape the information seeking process has been identified by those who have studied the activity. However, little has been found which enables the user to use comparable means within an information system. Furthermore, nothing has been seen in which the text-processing methods use or exploit any user artefacts. It is also to be remembered that observers have particularly noted the use of artefacts by those performing extended information seeking. This suggests that, if artefacts exist which could be used in some manner with the techniques for improving search that have been discussed above, further improvements may be made.

In the next section I will discuss information structuring. It is worth noting that this activity, which has been described as separate from the information seeking task, has fallen within Ellis' behavioural model of information seeking. As query formulation and the different aspects of post-query review have been used to support each other, one prospect is also that any structuring activity may also provide a framework for supporting the rest of the traditional information seeking process. Again, previous research suggests that the organising of documents is part of the manipulation of artefacts which information workers use to manage their long-term work.

Therefore, this section concludes with two separate, though potentially interlinked, possibilities for facilitating long-term information seeking; one through exploiting some form of artefact or artefacts (not yet defined), and the other through observing any information which may be present in the connected activity of information seeking. However, a caveat must be maintained; users are reluctant to give the computer system any additional input to amplify the performance of their task.

2.8 Information Structuring

As I noted at the beginning of this chapter, users not only obtain and retain documents, but also structure, organise and store them. The structuring of collections of documents by users has generally been less widely studied than the activity of users in seeking information.. Nonetheless, a number of different investigations have been made both into how users structure and organise collections of documents, and how systems can be created to support this activity.

Marshall and Shipman [Marshall 1994, Shipman 1995, Marshall 1997] have written a series of papers investigating how users organise documents, and developing a series of systems to support this activity. Their initial observations investigated the performance of such work by human subjects using paper-based documents, and compared this with others who were using computer-based systems that were intended to provide similar affordances. Their researchsuggested that a number of properties of systems that were important in supporting natural organising practice:

1) Allow for increasing, incremental organisation
2) Permit ambiguity and uncertainty to be expressed; *e.g.* the provisional nature of organisation

3) Use visual cues to suggest the importance and purpose of documents

4) Facilitate free association and annotation

5) Fluid movement from a large-scale view of the organisation to small-scale views.

Earlier examinations into the use of hierarchical filing systems [Mander 1992] had indicated that users did not perform particularly well using hierarchical navigation to recover documents, when compared to directed search access. However, Marshall and Shipman believed that this was because the visual presentation used in such systems did not facilitate overviews and broad scanning well. This argument was similarly made by Dourish *et al* [Dourish 2000] and Baldonado and Shneiderman [Baldonado 1997].

Much of Marshall and Shipman's work was based upon the findings of studies such as Malone [1983], O'Day and Jeffries [1993] and Kidd [1994], who found that knowledge workers and other information seekers used their environment as a form of external memory. Kidd's work particularly identifies the use of spatial layouts and materials. People undertaking knowledge work in her observations used certain areas (*e.g.* their desks) to hold different sorts of information; *e.g.* those pieces which had an as yet uncertain role in their work, others which were the output of their labours, some to remind them of what remained to be explored, and others as a place to organise digested and understood material. As with O'Day and Jeffries comments on information seeking, Kidd discovered much uncertainty and provisional-ity in the work patterns she saw. Malone, in observing the organisation of material upon the desks of knowledge workers, identified the informal organisation of documents as an inherent part of short-term storage and management of documents during information work.

It is notable that those studies that identify and discuss information structuring as an important activity, are also studies that started from wishing to identify information seeking procedures. Both O'Day and Kidd are examples of this, a trend continued later by Bates [1996].

2.8.1 Information Structuring Process

Unlike the information seeking process, no models of the information structuring process exist. Therefore, process models cannot be used to shape the development of any system that provides support for structuring.

2.8.2 Spatial Hypertext

Through their investigations, Marshall and Shipman created a new form of system, which they termed a Spatial Hypertext [Marshall and Shipman 1993]. Other, similar, systems have since been created by other researchers, *e.g.* Pad ++ [Bedersen 1997] and Chaomei Chen's Pathfinder-based system [Chen 1998]. Studies such as those of Kidd [1994] strongly influenced Marshall and Shipman; observing the dependency of information workers upon work artefacts, they wished to create systems that enabled the user to capture the artefacts that influenced their work patterns so much.

The defining characteristic of a spatial hypertext is that documents are represented in a graphical/textual form in a freeform visual space. The visual representation of each document is used to indicate the document's properties. Properties such as colour, position, size, shape and textual content all form different cues which can be exploited *by the user* to communicate the aspects of a document which are important to them and their organising work. For example, the objects which represented similar documents could be displayed in a row or column, in a loosely clustered group, or iby sharing a common colour, to indicate their common nature.

The visual representation of each document can be interacted with to control its appearance and in addition has the property of being a hypertext link to the actual document that it represents.

It is this combination of a visual, spatially arranged objects and hypertext capability which gives rise to the term "Spatial Hypertext". In addition to this use of visual properties as implicit descriptors of items in the hypertext, many systems, including VIKI for example, permit users to add textual descriptors, consisting of a label and one or more values. This latter facility gives an opportunity for more traditional textual searching in addition to visual scanning as a means of locating documents of interest.

A sample illustration of a spatial hypertext is given below:

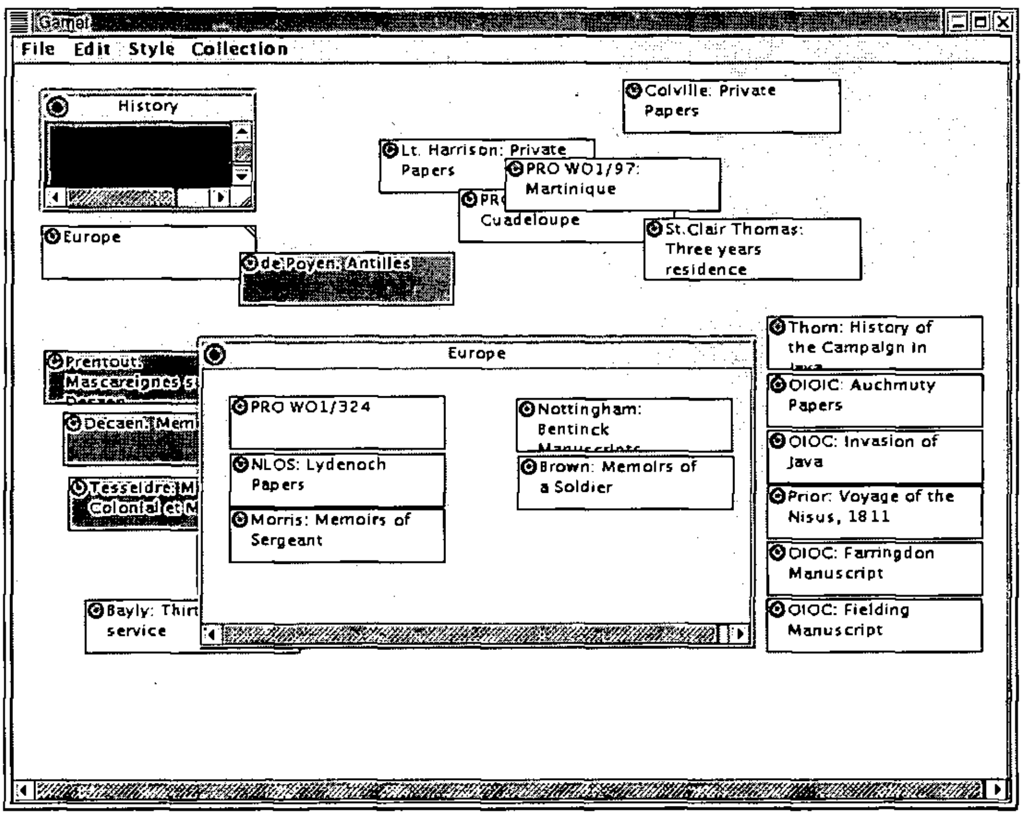


Figure 2.1: A sample spatial hypertext

In this hypertext, a number of common spatial hypertext features and idioms can be seen. In the centre of the hypertext is a window labelled "Europe". This window contains five individual document labels. In spatial hypertext terminology, this window is called a 'collection' – a formal, explicit structure containing a number of documents. Collections may often contain further (sub-) collections within them.

The "Europe" collection's five documents are organised into two columns. Each of these columns is an informal, implicit structure created by the user's use of space. Another example of a column can be seen on the right hand side of the main workspace. At the top of the main workspace one can find a number of documents loosely scattered about. However, four of them (*e.g.* "PRO WO1/97") are grouped on top of each other. This is another informal structure called a 'pile' or 'stack'.

Each structure – implicit or explicit – is created by the user to express their partition of the documents at hand. A group of documents may be related by topic, author or any other property that the user feels is relevant. The degree of certainty with which a document is associated with a group can vary or be ambiguous. For instance, the "Colville Papers" document may or may not be associated with the pile of documents at the top of the workspace.

Colour and other visual properties (*e.g.* the shape of the document label) can be used to express other aspects of the document. For example, in the workspace above a number of documents have had their label shaded. In this case, this represents that the documents are French in origin.

Spatial hypertexts therefore contain a number of means of expressing orderliness, certainty and relationships (*e.g.* by topic or author). The explicit properties are transparent to the computer (*e.g.* which colour a label is) whereas implicit ones are not (*e.g.* informal structures). I will now briefly discuss the problem of identifying these implicit structures.

Identifying Structures in Spatial Hypertexts

An important property of Spatial Hypertext is its ability to express relationships that are ambiguous or certain. For example, if a group appears as a well-structured grid of objects, placing another item nearby out of alignment with the group may suggest that it is similar to, but not certainly a member of, the group. Given this, and the freeform nature of presentation available in VIKI [Marshall 1994], Pad ++ [Bederson 1996] and other spatial hypertexts, it is difficult for the computer to identify aggregate structures [Botafogo and Shneiderman 1991] in a hypertext.

Without a system being able to identify structures and relationships within a hypertext, and their meaning, the hypertext will remain a tool that leverages only the user's tasks. If the user is able to identify groups of documents to the system, then the user can ask the system to perform some task over one or more groups.

This problem of identifying groups can be addressed by giving the user the ability to select groups and present them to the system for processing in some manner. For example, in the NaviQue system, the user can select a number of elements in the hypertext and request that NaviQue searches the documents for a particular term. The downside of this method, however, is that the user has to identify a group each time that they want to perform a task over it; there is no persistent group identity, only the selection currently made by the user. This forms the first type of structures available in a spatial hypertext: short term user selections.

However, this task of identifying groups can be performed by the system as well as the user; in [Shipman 1995] Shipman introduces the concept of a spatial parser which identifies visual patterns and sets, allowing the user to interact with a number of related documents at once. However, as the VIKI system is only loosely connected to an information system (see [Shipman 1997] for the only published example), the benefits of this ability have yet to be exploited fully. Thus, we have a second type of structure available in spatial hypertexts that may be identified and used by a computer system: implicit and continually available structures.

Finally, where structures are explicit and continuous in a spatial hypertext (*e.g.* user-created collections in VIKI), then clearly they can provide a coherent common ground between the user and the system. Clearly, any structures within a given item will need to be identified by one of the three methods outlined so far.

2.8.3 User Filing Systems

Another approach to the task of user structuring of documents[2] is the traditional filing system used on virtually all computer operating systems. Historically, this is seen in the context of users storing and organising files that they or their colleagues use.

However, in [Dourish 1999], Dourish *et al* expand the concept for use in the context of a user-controlled cataloguing or index system for documents. The documents themselves may not in fact be stored in the system. Dourish criticises the traditional model of immobile hierarchies for being inflexible and inefficient when user requirements change (*e.g.* the task of reordering the hierarchy in a system when the order of higher and lower "directories" are reversed).

Dourish *et al*'s response to this impedance is to enable the user to identify or "decorate" each item with labels or descriptors which can then be used to dynamically generate alternative hierarchical orderings of a set of documents as the user requires. Thus, the filing system becomes at heart a catalogue of its contents that can be dynamically queried.

Although Dourish *et al* approach the problem of user organisation of information from a different paradigm to those from the hypertext community, the two approaches can be easily mapped to each other. Firstly, the connection between the document entry in the user's catalogue in Presto (the system created by Dourish *et al*) falls within the accepted definition of a link in the hypertext community. Secondly, the use of dynamic queries to generate a navigable structure, used to create the hierarchy in Presto, also falls within the accepted form of a query as a hypertext link. Thus, although Presto may be an inheritor of traditional user-controlled filing systems, it clearly also belongs to the family of hypertexts. This sort of convergence was predicted some time ago in the seminal projections of Halasz [1988] and later expanded by Nürnberg *et al* [1996].

The annotation system of Presto could be described using a number of hypertext idioms; certainly the annotation capabilities of, *e.g.*, NoteCard and VIKI could be compared, though they are less queriable than Presto. Similarly, Link typing could also be used to describe the mechanism.

Like VIKI and other spatial hypertext systems, Presto, Dourish's system, can contain collections of documents. These can be created via two mechanisms: query, in which the collection is created by executing a query over the total document store; manual, in which an individual document is included or excluded from a collection. The two mechanisms can be combined; so a typical collection may be seeded by a query, yet altered by the addition of some other documents and the exclusion of some which resulted from the search. The manual system of building documents is comparable to *e.g.* Pad++ and VIKI [Bederson 1994, Marshall 1994]; the automatic one has been used in some form by systems inspired by spatial hypertexts, particularly digital library interfaces such as DLITE [Cousins 1997] which indeed Dourish compares his system with in [2000].

The interface is presented visually, as is the case for spatial hypertexts.

Another user filing system approach is outlined in Mander *et al* [1992], where Piles, informal structures of overlapping document icons, are introduced as a helpful short-term support for users when engaged in a creative task including many files. This system is influenced by the observations of Malone [1983] upon the organisation of people's desks when engaged in similar tasks in a physical environment. The comments made above concerning Presto – its visual nature, its similarity to Presto, can also be made of the Piles metaphor, and it can be particularly compared to the "aggregate" spatial hypertext idiom later identified by Marshall and Shipman [1993].

Although Presto is described by its authors as a filing system, and the Piles metaphor by its creators as a feature of a filing system, it seems justified to considering these and similar

---

[2] NB: A filing system is taken to be a system specifically intended for storing and retrieving documents, rather than a system for storing files in general; this distinction is itself made by Dourish et al.

systems as a special case of Spatial Hypertext. Their interactive and presentation styles are similar to spatial hypertext forms, their expressive forms such as annotation are comparable to hypertext facilities, and their construction techniques bear a close relation to hypertext forms.

2.8.4 Summary

We have seen a couple of common information structuring tools, and the shared theories that underpin them. Both provide the means of describing documents in a multi-faceted manner, which organises and classifies items in a series of attributes. In the case of the Presto system, the main means of articulating these classifications is through the use of textual descriptors, which are explicitly entered into the computer. By comparison, in Spatial Hypertext, these descriptors are generally visual (though textual descriptors are also used in a number of spatial hypertext systems), in which case the semantics are implicit, and seeking must be performed visually by the reader.

Filing systems with a hypertext interface can also be seen as a particular form of spatial hypertext themselves, so we again found similarities. Combined with the fact that filing systems have a function subset of the organisational capabilities of spatial hypertexts, it seems fair to treat the former as a specialised subset of the other.

Spatial hypertexts clearly provide an artefact-centred approach for organising information. Perhaps unsurprisingly, a number of the works which inspired an artefact-centred approach for information seeking are also to be found as formative influences upon spatial hypertexts. I shall now finally consider how the information seeking and information structuring processes can support each other inside a computer-based system as they have been reported in traditional information work.

2.9 Discussion

In reviewing the work patterns of information workers, we have seen that structuring is an important part of their interaction with the documents with which they are working. We have also briefly remarked on spatial hypertext systems, which facilitate the organising and classifying goals of the structuring task.

Earlier, we observed that certain synergies exist between different stages of the information retrieval process, particularly in respect to the task of query formulation, either through the user's own efforts, or through the support of the information system itself. The traditional, procedural, information seeking model by Marchionini does not include the information structuring task, but the behavioural model of Ellis does, and frameworks for information work as a whole such as Goh and Leggett's ASAP model do. This suggests that there could be synergies between information structuring and information seeking as there are between different aspects of the information seeking task.

We have seen that spatial hypertexts provide an effective and plastic environment for the information structuring task, and that they permit the existence of informality and uncertainty which researchers have observed in the organising work of information workers. Furthermore, we have seen that structures can be identified by a computer system from the implicit information in freeform visual stores such as Spatial Hypertexts.

Some methods for supporting information seeking through filtering and classification have endeavoured to create user-specific structures through the use of text processing. Automatic methods have particularly suffered from the problem of discerning intent in user actions, and manually programmed systems have suffered from not well-matching the user's interests, particularly as they have changed over time.

Spatial Hypertexts offer an opportunity to provide some support for identifying the interests of a user through those documents they select to place in it and through how those documents are ordered, i.e. through intended actions. Traditionally, the only purpose for Spatial Hypertexts has been to support the information structuring task, but as observed above, this task can be combined with the information seeking task to which it is usually tied.

Using textual extraction techniques over the visual structures in a Spatial Hypertext, one may hope to generate textual structures which correlate to the user's interests, and can be used to support the better performance of the information system in their information seeking activity. These structures will alter as the user adjusts their information needs and changes their perception of the topical structures that encapsulate the field in which they work.

2.10 Conclusion

I have introduced the processes of information seeking and information structuring. The role of information structuring in information seeking has been outlined, and the support of information structuring through Spatial Hypertext has been introduced. The generally low level of connection between spatial hypertexts, supporting information structuring, and information systems supporting information seeking has been noted. One clear opportunity that arises from this knowledge is the connection of a spatial hypertext to an information system such as a digital library, which would permit some insight into the benefits of a combined information seeking and structuring environment.

In reviewing the means of improving the performance of information retrieval systems, I have demonstrated the ties between different stages of information seeking. Information retrieval has exploited the artefacts from one part of information seeking to enhance effectiveness in another. It is thus possible that information structuring, as part of an extended information seeking process, may be similarly exploited.

The grouping of documents that match a query by topic has been introduced as a viable alternative to ranking by relevance, as has the identification of document groups within a spatial hypertext. If users do, as argued by spatial hypertext researchers, organise documents topically, it may be possible to glean information about the user's fields of interest from the organisation of their workspace. In a group of documents that have a common topic, it is possible that a number of phrases or words may be found in each document. These words and phrases may then be useful in finding other related documents. Therefore, a second issue of interest is what benefits may be gleaned from extracting the implicit and explicit structures in a spatial hypertext to assist the discovery of documents relevant to a user's interests.

This thesis will now proceed with a discussion of the technical issues involved in the connection of a Spatial Hypertext and a digital library, the considerations for identifying and using the organisation performed by a user in their workspace, and how that organisation might be used to identify documents which are of similar topics to the groups created by the user.

Chapter 3 – Technical Background

3.1 Introduction

In the previous chapter, the important role of information structuring in the activity of knowledge workers was identified. Information seeking, the discovery of new information sources, was discussed, and a number of information seeking models (e.g. [Ellis 1989]) were found to include information structuring. Information structuring tools – particularly Spatial Hypertexts – were briefly introduced, and it was noted that these lacked a close connection to information seeking tools such as digital libraries. Conversely, information seeking systems were observed to typically include little or no information structuring support, having the more limited view of the information seeking task models such as Marchionini's.

Another important observation was that Information Retrieval research – in attempting to improve the effectiveness of a particular stage of the information seeking process – has often relied upon user activity in one phase of the process to boost effectiveness of the computer system in another. Therefore, in principle, one may be able to improve information seeking performance through capturing data on user behaviour in the information structuring task.

I concluded the chapter by suggesting that a combined information seeking and information structuring system could be of benefit, especially given the interleaved pattern of the two tasks found in physical information seeking environments. Such a system could be formed from the information structuring tools of Spatial Hypertext and the information seeking tools of a Digital Library.

Therefore, I propose to create such a system – which I shall call Garnet – and through it explore the viability of identifying user patterns of organisation in a spatial hypertext. Garnet will give the user access to the features of a digital library through a spatial hypertext interface. Such a system would facilitate the evaluation of the benefits of a combined information seeking and structuring tool, through the observation of the work patterns of users working with it, and the capture of their information structuring activity.

To create such a system, a more detailed investigation of the technical and design aspects is required. This chapter will therefore review more thoroughly Spatial Hypertext, Digital Library and Information Retrieval research that is relevant to the proposed system.

Firstly, the variety of structuring facilities which are available to the user will have a profound effect upon the ability of the computer to capture the user's organisation of documents, and the user's scope for expressing their own perceptions of the organisation over and relationships between documents. Therefore, this aspect of spatial hypertext systems, and the structuring tools in existing visual interfaces to digital libraries will be discussed.

Secondly, the means by which informal structures created by users can be discerned by a spatial hypertext system, spatial parsing, will be described.

Having identified how a computer system can recognise the structures created by a user, the chapter will proceed with a discussion of how that structural data could be used for supporting information retrieval – the obtaining of documents from an information repository. This discussion will include reviewing the benefits of structures in information retrieval, how document groups can be represented for information retrieval purposes, and how those group representations can then be used in performing information retrieval tasks. This will also require the study of some digital library protocols, as these substantially affect the means available to perform information retrieval computation in the context of digital libraries.

The chapter will conclude with a summary of how these different aspects interplay, and the design implications for Garnet.

3.2 Spatial Hypertext

In Chapter 2, Spatial Hypertexts were introduced as effective systems for the progressive organisation of information, and a number of common spatial hypertext features were discussed at an abstract level. As spatial hypertexts systems have existed since the early 1990's, over a decade of development has resulted in the development of a number of different systems, each with particular advantages and design objectives.

This section will discuss the information structuring facilities available in a number of key spatial hypertext systems, and the degree to which the user structures created through these facilities can be identified by a computer system. It will also discuss the known task affordances of the different structuring facilities – e.g. how readily uncertainty or confidence regarding the organisation is represented, relationships between documents articulated, etc.

It is also proposed in Garnet to provide access to digital library facilities within the spatial hypertext environment. How the products of another system could be represented within a Spatial Hypertext workspace will also be discussed.

3.2.1 Expressing Structure and Relationship in Spatial Hypertexts

As discussed in Chapters 1 and 2, each object represents a single document in the spatial hypertext workspace. The positioning, colour and shape of the object can be used to articulate its role, content, relationship to other documents or any other properties that the user wishes to express. The exact tools or visual properties available to the user vary between different Spatial Hypertext systems, and this variation influences the range of expression available to the user, and consequently upon the patterns which are likely to emerge in the workspace. The role and impact of particular structuring tools and cues is, unfortunately, not yet well understood [Shipman 2001a], and therefore design decisions are neither clear nor trivial.

Some distinct forms of structuring can, however, be observed and distinguished from each other in existing spatial hypertext systems. This section will discuss some of these, demonstrating the distinctions with examples, and in each case the ease with which the resulting structures may be detected by the spatial hypertext system will also be briefly discussed.

One main division of structuring tools is between "explicit" and "implicit" organisation.

In explicit structuring, the membership of a group is defined by some object or artefact, whilst in implicit structuring, there is no external reference to define whether two documents are in the same or different groups. To take VIKI as one example, documents may be placed into a hierarchical tree structure, where each node is displayed as an object in the workspace. Those items placed inside the window that represents a node or "collection" are members of the collection. Items not inside the window are not members of the collection. In a physical information environment, an equivalent would be a storage folder or filing cabinet [Malone *et al* 1983].

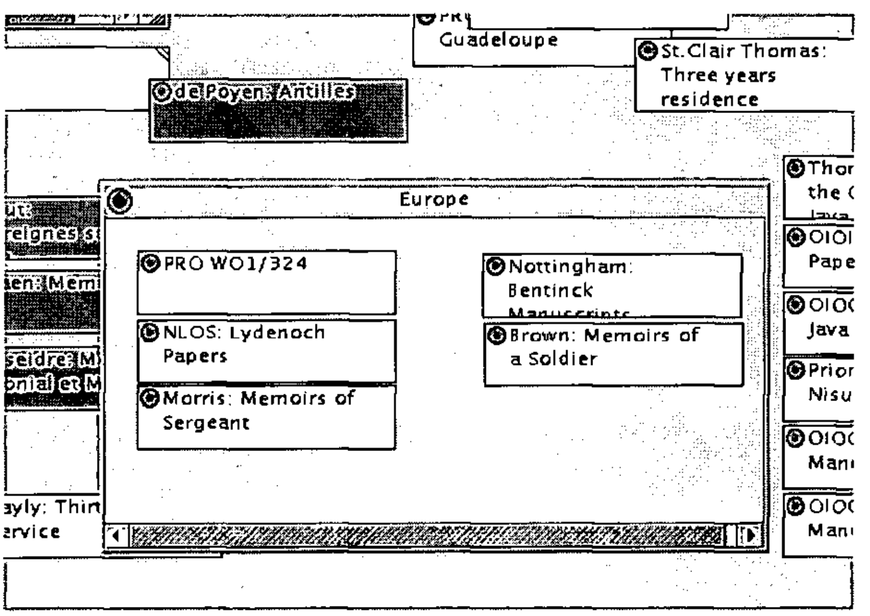


Figure 3.1: A spatial hypertext in use: a 'collection' called "Europe" appears in the centre of the image. Membership of the collection is clear both visually and logically.

However, this facility is not found in the Pad++ system [Bederson *et al* 1994], which possesses implicit structures alone. An example of implicit structuring in Pad++ is where a large label at one level of zoom may be used to indicate the presence of documents at a deeper zoom level – focussing the display upon the label at a closer zoom level will reveal the related documents, which are close to the label. Similarly, in VIKI, similar documents may be placed close to each other, but the membership of a group has to be deduced or intuited by the reader. Again, this has a parallel in physical environments, for instance the pile structure observed by Malone *et al* [1983].

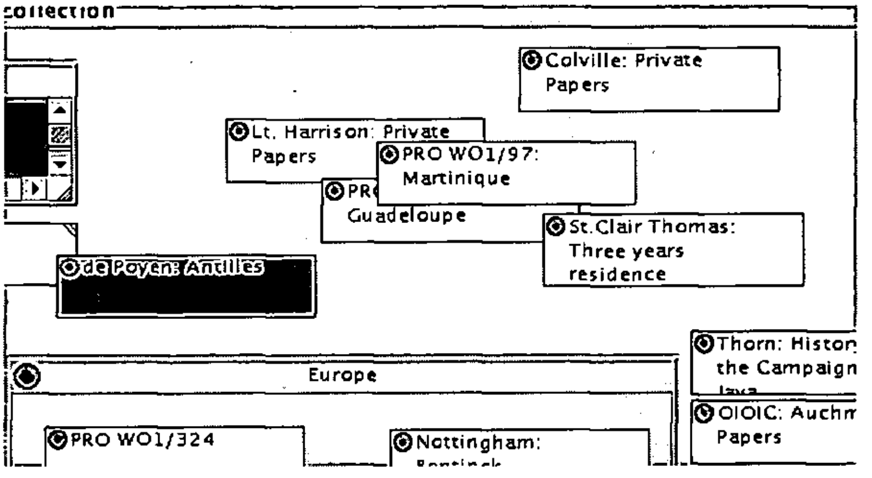


Figure 3.2: A pile is seen centre top of this snapshot. It is not quite clear which documents are in the pile or not. 'PRO WO1/97' in the centre and on top is clearly a member – but what about St. Clair Thomas (right) or Colville: Private Papers (top)?

In both these examples, the explicit or implicit structures relate to the position of the document in the workspace. However, this is not the only manner in which structures – implicit or explicit – can be expressed. In VIKI, for example, documents can be given an explicit 'type' which determines their shape, colour and text style – this is again an example of "explicit" structuring – a document is either assigned a particular type or not. On the other hand, the user could select these properties on a document-by-document basis. In this case, the system would have to detect consistent patterns in appearance across a workspace – again, implicit structuring.

Explicit structuring is easy for the system to detect – membership of a document type, or of a node in a hierarchy are readily determined from the data structures that underpin the hypertext. However, implicit structures are more difficult. For example, if one considers the use of colour, to what degree would a close but not exact match of colour indicate a common property? Similarly, how close would two documents have to be physically to be considered part of a group of neighbouring documents?

Just as membership of a group may be implicit or explicit, so may the meaning or role of a group. Returning to Pad++, for example, the large label which may be used to indicate the existence of a group may be used to indicate the common theme of the documents, and the name of a collection in VIKI may serve the same or similar purpose. Conversely, in either system a group may not be labelled at all. As a result, a reader of the hypertext, or the system supporting it, would again have to resort to inspection of the documents to reach some conception of what the topic of the group may be.

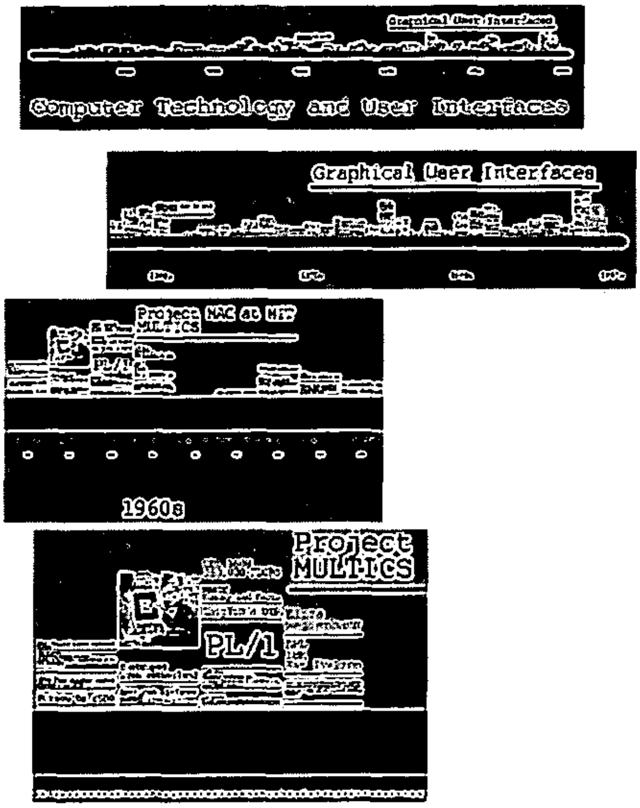


Fig 3.3: Pad ++ in use; at increasing Zoom from top to bottom. Some relationships are clear (e.g. time sequence from left to right), others (of topic) are not.

Marshall [2002] believes that the expression of uncertainty about membership of a group is an important feature of Spatial Hypertext – reflecting the fluid, provisional form of the structuring task. This extends and complicates the question of what constitutes membership of a group. Again, placement may be used (e.g. moving a document slightly further away from a group may indicate a higher level of uncertainty of its belonging to the group), or colour (e.g. using a lighter or darker shade). Uncertainty would seem to be more easily expressed through implicit structuring features, as in the examples just given, than with explicit structures such as collections or types in VIKI, where a document is clearly either a member or not – doubt is not capable of being expressed. (See. Figure 3.2)

No tool which I have seen described as a spatial hypertext system has explicit structuring alone – it would seem that the informal organisation which motivated Marshall and Shipman's early work, and can be found even in the early Notepads system, and which is well represented in informal structuring capabilities, is in some sense fundamental to the role or definition of Spatial Hypertext. However, as I have observed, detecting implicit structures is far from simple.

Though explicit structuring would be simpler to recognise this would lose a considerable amount of structuring expressiveness, and an expressiveness that would seem central to the concept of Spatial Hypertext. It is therefore necessary for a spatial hypertext system to support implicit structuring, even though that necessarily makes evaluation of the structure more complex and prone to error.

3.2.2 Navigation

The navigation of the user through a large hypertext is also open to different means of facilitation. When working in a digital library environment, where a user may recall many documents over a long period of time, and additional material may appear in the workspace besides the documents themselves, large workspaces are very likely, and thus the navigation across the workspace becomes an important design consideration.

The Pad++ system [Bederson *et al* 1996] relies purely upon zooming, with "high level" objects indicating topical areas at a low level of zoom, and as the user focuses upon a particular area of interest, it fills and then extends beyond the visible display, and "low level" documents gradually become visible. VIKI also supports zooming, but it also uses a traditional, structural navigation akin to what is found in a filing system on most contemporary GUI environments. However, the visibility of documents also suffers under this approach as, although documents in the current node in the hierarchy are generally highly visible, documents in other nodes are outside the current field of vision.

Neither of these approaches, therefore, represents a complete solution to the problem of visibility. This is a problem endemic to the viewing of large document spaces [Li *et al* 1998][Card *et al* 1999].

3.2.3 Comprehension

Cathy Marshall [Marshall 2001] voices concern over the ability of authors of spatial hypertexts to comprehend their own implicit structuring of documents after a period of time focussed upon other tasks and/or workspaces. Similarly, in [Marshall *et al* 1999], she observed that when a reader annotated documents electronically or on paper, these annotations often seemed to be less comprehensible to the same person when they returned to the documents after some interval. In each case, implicit work is deeply connected to the context in which it was performed, and when that context is degraded, either by interruption or through its presentation to a person other than its creator, meaning is difficult to recover accurately. Therefore, some support may be required to assist the recovery of context, e.g. identifying the role or purpose of groups.

3.2.4 Representing Information Sources in a Spatial Hypertext

It has already been observed that spatial hypertext systems have seldom been connected to information sources. There are a small number of exceptions to this, and this section will review these to identify how an information system is represented inside a spatial hypertext.

At the simplest level VIKI and its successor VKB both allow the reader to link from an item in a spatial hypertext to a web document in a manner similar to a traditional web link – clicking on the shape which represents the document then opens the document itself in a web browser. When a document is created, the link may be added using the operating system cut-and-paste operation. Similarly, links could be dragged into the spatial hypertext and a shape would automatically be created to represent it. VKB extended this to permit a link to any file accessible via the user's usual filing system as well. Pad++, another popular spatial hypertext system, also allows for the inclusion of graphics, web links and files. In all these examples, all that is a provided is a link to an external document. When the link is activated to view the corresponding document, no additional content is supplied to the spatial hypertext. For the purposes of viewing a document, this is useful. However, for connecting to any system that would produce content into the workspace, there is little to be learnt.

An extension of VIKI, used for authoring "Walden's Paths" presentations [Shipman *et al* 1997] permitted web search results to be presented inside a spatial hypertext. A dialogue was presented to the user to enter their search criteria. When the dialogue was completed, the top ten search results were presented in a collection of two columns each of five document objects. Unfortunately, the usability of this presentation was not assessed. How well the users of the system were able to discern between their own material and that produced and organised by the search system was not studied, for example.

NaviQue [Furnas *et al* 1998] and KidPad [Druin *et al* 2001] are both Digital Library systems that are strongly influenced by Spatial Hypertext research. In the case of KidPad, based upon the Pad++ spatial hypertext, the user can browse the library as one would the hierarchy of collections in a hypertext such as VIKI. However, KidPad does not contain facilities for the user to add their own documents to the workspace. Therefore, the potential conflict between objects placed by the user and those placed by the system does not arise. Nonetheless, one possibility that may be used for browsing of a digital library within a spatial hypertext is to portray the structure and content of a library as a hierarchical workspace that the user cannot change.

In NaviQue, the library is again presented as the content of a pre-authored Spatial Hypertext. However, in this case, the user can perform searches that result in the creation of sets of search results which appear within the same space. The search result sets can be used much as any other container of documents in the virtual space. Thus, as with the Walden's Paths support in VIKI mentioned above, search result sets are placed into the main workspace.

The common theme across these different circumstances is to present the output or content of a system in a similar manner to the other documents in the hypertext. However, the colouring and positioning of system objects is a potential problem. The distinction between which elements were placed by a user, and which by the system may be important, and so some visible indicator of this should be available. However, there is a limit to the visual cues available. Any cue used by the system could not be used, unambiguously, by the user and vice-versa. Therefore, the wider the system's expressive capability, the lower the user's free scope for expression.

3.3 Visual Interfaces to Digital Libraries

A number of visual interfaces specifically designed to support extended work in digital libraries already exist. Three such interfaces are DLITE [Cousins *et al* 1997], NaviQue [Furnas *et al* 1998] and SketchTrieve [Hendry and Harper 1997]. All three systems are intended to give coherent access to a number of information services (e.g. content search, author indexes etc) and sources (different collections from a number of libraries), and represent separate searches as discrete objects in a 2-dimensional workspace. Each of these systems has been created to facilitate information workers in extended tasks, and so support some degree of task organisation. This section will briefly examine these systems from a Spatial Hypertext perspective, to discover what differences may exist between a Spatial Hypertext interface to a Digital Library and the features of these individual systems, all of which have been influenced by Spatial Hypertext research.

From the perspective of Spatial Hypertext, a major question would be the range of expressiveness that these systems give in the organisation, relating and annotation, implicit and explicit, of documents. Therefore, I will now consider how much control the user has over the appearance of objects in their workspace, the relative significance of different objects in the workspace, and (considering our immediate interest) in what manner they can use the objects in their workspace to perform further work.

The representation of individual documents varies considerably. DLITE represents documents as small graphical icons whose appearance is set by the system (with the exception of colour), whereas SketchTrieve's representations are larger, including the entire document text, with the visual properties such as colour and size being controlled by the user. In the case of NaviQue, the basic representation is somewhat similar to VIKI, but NaviQue emphasises zooming as a method of browsing large-scale areas (a property it inherits from the Pad++ spatial hypertext of Bederson *et al* [1997]). The consequence of this emphasis is that when a wide area is visible, the impact of individual documents is very small – often just as single points of colour. The degree of expressiveness in regard to single documents therefore ranges from the extremely limited to the highly flexible.

Structuring facilities across documents is much more fundamental to Spatial Hypertext. In DLITE the ordering of documents within sets is system-, not user- controlled. Sets also cannot be structured into a hierarchy, merely as a number of peers, and all sets of documents have explicit structure. NaviQue, on the other hand, lacks formal document groups – the user identifies sets of documents in which they are interested by directly selecting each one. Thus, document groups are explicit but temporary. SketchTrieve falls between these two positions; document sets do exist, but only as the results of a search – sets cannot therefore be used for explicit ordering within the workspace. Overall, therefore, any structures in the Visual Interfaces to DLs are explicit ones, and are generally not created by the user. Hierarchical organisation is also not found in any of these systems.

In comparison to these Digital Library interfaces many spatial hypertext systems, e.g. VIKI, have formal sets that can be organised hierarchically. Documents can be added to a set at the user's discretion, and the internal organisation of a set is fully within the user's control. The exception to this is in the case of sets created by the system – usually search result sets – where the organisation and membership of the set is determined by the system. In addition, as was observed above, implicit structures are important aspects of Spatial Hypertext, allowing for the provisional and tentative nature of the information structuring process. Thus, visual interfaces to DLs are very weak when compared to the information structuring capacity of spatial hypertext systems.

SketchTrieve and DLITE are notable for one means of relating objects to each other. Objects can be connected together to relate them. In the case of SketchTrieve the relationship is expressed by lines that are drawn between objects. In the case of DLITE, objects can be docked together if they are 'compatible'. However, this expressiveness is not used to express the sorts of semantic relationship that I have noted elsewhere. Rather, these are used to connect documents to the search list in which they occur (in the case of SketchTrieve) or search terms to search engines (in DLITE). These sort of connections are not found in spatial hypertexts, which have not been connected to similar information repositories and therefore have not had a need to express relationships between searches and their inputs or outputs. However, this does extend the understanding of how visual cues can be used to express relationships – not only between documents, but also between all sorts of objects in a digital library.

The key-pin of these visual DL interfaces has been the connection of an information workspace to a digital library system, not the organisation of chosen documents by users. It is worthwhile, then, identifying the facilities available. SketchTrieve contains basic searching facilities alone, whereas DLITE is very extensible and could, in principle, access a wide variety of DL facilities. However, most of the facilities reported in DLITE are traditional DL functions (e.g. browsing), and the core remains interactive search. NaviQue possesses a 'similarity engine' that permits the comparisons between documents and sets of documents, somewhat similar to the ability to match documents against the document groups created by users that I propose in Garnet. However, in NaviQue the principal benefit is navigational assistance – highlighting similar documents in their current position – rather than support for organisation and structuring, or for improved information seeking, which is the goal I am addressing.

Overall, therefore, these interfaces facilitate traditional digital library actions and exploit some of the idiomatic characteristics of spatial hypertext to assist long-term work. When compared to traditional spatial hypertext systems, the scope for emphasising, organising and structuring documents via visual cues is weak, and sometimes virtually absent. It is therefore unsurprising that the user's use of space, colour etc. has not been used by any of the systems to identify user structuring activity, which is what Garnet does.

These interfaces all intend to support long-term information seeking. Through their expressive range, as seen from the perspective of Spatial Hypertext, each may have valuable insights into how long-term work may be supported.

One common set of features, argued for by the creators of each of the systems, is the provision of facilities to track the user's search history and co-ordinate across searches. Multiple searches can appear within the one workspace, and the user may switch freely between these to view, compare and contrast the results of each search. Each search is represented by a single object in the workspace, which can be expanded to view its results, either within the workspace (e.g. SketchTrieve) or in a separate view (NaviQue). In DLITE, each set of search criteria is represented by a single object in the workspace, and these can be combined with the objects representing individual search engines to trigger the execution of searches, whereas in the other systems these inputs are entered via a dialogue, and no search engine objects are present in the workspace. All the systems support a search history facility, to allow the ready recall and inspection of earlier search criteria and search results. Hendry and Harper [1997] argue that the arrangement of search result sets and other artefacts of information seeking by the user represents an important secondary notation – the term that they use to identify the use of placement to impart meaning and significance. However, across these systems the user's control over the representation and appearance of search artefacts is limited, as was the case in the representation of documents and document groups. These artefacts are presented to readily allow their visual distinction from other components of the workspace such as documents, search result lists or search histories.

One further aspect that is briefly discussed by Hendry and Harper is how the occurrence of an object in more than one context – e.g. a document in two separate search lists – can be expressed in the workspace.

This section has demonstrated the relatively limited scope for user control of the appearance of document representations and document organisation when compared to spatial hypertext systems. Spatial Parsing, which I will discuss next, has not been exploited in any of these systems, despite the fact f the influence of Spatial Hypertext upon their creators. In addition to representing documents in their workspaces, these systems also represent search artefacts (result lists, histories) there. These representations are readily distinguished from documents, though this again may limit the range of expressions that the user can use without risking ambiguity between the system's use of appearance and their own. A new issue that has emerged is the problem of representing the recurrence of documents in different contexts.

3.4 Spatial Parsing – Recognising Visual Patterns in Spatial Hypertext

In the previous section, two forms of structuring were observed in Spatial Hypertext – implicit and explicit. The membership of explicit structures is readily identified. However, the challenge of how to identify the membership of implicit structures has not yet been addressed.

Spatial Parsing identifies patterns and groupings of objects in a graphical space. The VIKI and VKB spatial hypertexts have exploited Spatial Parsing to support the identification of informal groups of documents so that they can be readily selected and manipulated as a block, and this is outlined in both [Marshall *et al* 93] and [Shipman *et al* 95].

The underlying principles have emerged from the use of visual programming languages in papers such as [Lakin 87], and applied in other areas such as graphical editing. The current research corpus is generally rather small, and is heavily skewed towards the use of Spatial Parsing in hypermedia as a whole and Spatial Hypertext in particular.

The construction of a visual parser requires there to be some defined set of "visual expressions" which are to be matched. Each expression is encoded in a recogniser function, and for unrecognised objects each is called in turn until a recogniser returns a positive result, in which case the corresponding graphic context is eliminated and the process restarted until no more material can be matched.

The coding of the visual expression recognisers is somewhat heuristic, and no theoretical framework has evolved for these.

In Spatial Hypertext, Frank Shipman has developed a number of recognisers for the Spatial Parsing component of VIKI and VKB. Those which have been described in the available literature, e.g. [Marshall *et al* 1993], are:

> Row (or List): a horizontal line of objects of a common type, non-overlapped
>
> Column (or List): a vertical line of objects of a common type, non-overlapped
>
> Aggregate: a group of mutually overlapping objects
>
> Taxonomic set: a group of neighbouring documents of a common type.
>
> Composite: a group of neighbouring documents of heterogeneous types.

The recognisers for these patterns are applied in turn, and any identified group can itself form an object that can be matched by later application of the recognisers. Thus, for example, a grid of documents might be recognised first as a set of rows of document objects, and those rows then used as the elements in a column of 'row' objects. Colour, shape etc. can all be used as the basis of an 'object' type. Thus, a column of three green objects at the top, three yellow in the middle, and three red at the bottom would be evaluated as three columns of one colour, joined into a column of column objects. This recursive, layered approach is repeated until no further patterns can be found.

However, Shipman's algorithms are not described in detail in the available literature, nor is the source code of either the VIKI or VKB system available.

In the case of any spatial hypertext used to access a digital library, a distinction will need to be made between the objects placed by the system, the structure of which is already known, and those placed by the user. Evaluation of the known structures would only add computational cost for little or no benefit. In the case of any documents temporarily placed into the user's workspace by the system, these may affect the performance of the Spatial Parser, causing document groups to be mismatched or ignored. Such issues have not affected VIKI or VKB, as the system has not had any active role.

The correspondence between the use of particular visual patterns, such as a column, and any semantic distinctions is not known, nor is the value or otherwise of a multi-layered, hierarchical evaluation of visual patterns known. Therefore, it is possible that understanding the internal structure of groups by visual pattern may assist the better characterisation of it, but this is neither certain nor understood.

Computational efficiency is clearly important in any interactive function of a system, and the spatial parser may be involved in the user's selection of document groups, and work done in re-organising the workspace. Reinert *et al* [1999] introduced in their implementation an *incremental* spatial parser – only re-parsing those parts of the workspace that had changed at any given point. Shipman's implementation may or may not be incremental – this is not known. Synchronisation concerns are also salient, as clearly the behaviour of any facet of the system relying upon the Spatial Parser needs to reflect the organisation of the workspace at that point in time, and so any re-parsing of the workspace must be complete before any dependent computation is performed.

In summary, then, little is known about spatial parsing, but a small number of informal structures have been identified as common idioms in the layout of spatial hypertext workspaces. These simple structures have been reliably identified in a number of systems, the details of which have not been disclosed. The semantic significance of these structures is another unknown factor at present. A little is known of the design of spatial parsers – for example, their computational efficiency has been improved by using incremental approaches, responding only to changes rather than re-parsing the entire workspace when alterations occur.

3.5 Exploiting Spatial Parsing

In the previous section, I described the state of spatial parsing research. In this section, I discuss how spatial parsing may be exploited in a spatial hypertext. First, I identify how I intend to exploit the spatial parser in Garnet to support information seeking. I then turn to how to represent that use of the spatial parser within Garnet's interface.

In §2.10, I stated that I would investigate how to use information structuring artefacts to support information seeking. Spatial parsing, described above, identified one technique for analysing the primary artefact of a spatial hypertext – the organisation of the documents within it. Spatial hypertexts often include the ability to organise documents in a formal, explicit hierarchy. All spatial hypertexts permit the informal creation of document groups. These informal groups may be considered implicit sub-structures within the parent workspace. Thus, the entire hypertext may be seen as a tree.

Researchers have previously evaluated both automatically generated [Hearst and Pederson 1996] and expert-generated topic hierarchies [Chen and Dumais 2000] for organising search results. In each case, the outcome revealed that this approach was comparable or superior to the traditional ranked list presentation. Thus the topic hierarchy in a spatial hypertext workspace for the same task seems viable.

However, these previous tools support searching only, not both searching and structuring. Furthermore, spatial hypertexts do not present themselves as an outline-style hierarchy, as is the case with these previous systems. Instead, the common form of spatial hypertext is a two-dimensional freeform space:

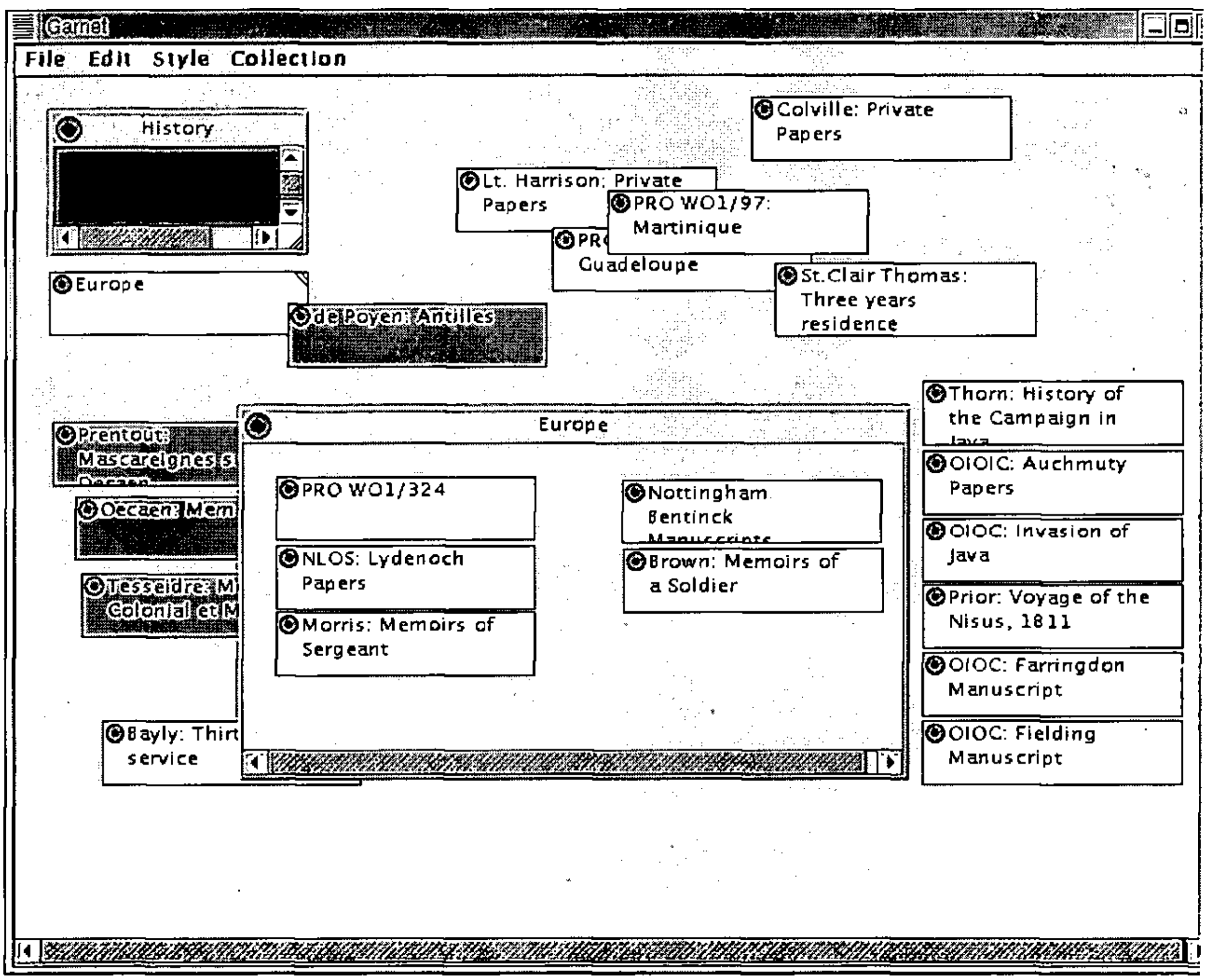


Figure 3.4: A 2-dimensional spatial hypertext.

When a user uses their workspace to organise the results of a search, the documents in the search result list are probably not already found in the workspace. Thus, the approach of Furnas and Rauch [1998] mentioned in §3.3, where matching documents in the workspace are highlighted, does not translate readily. The search presentation systems just described use a task-specific presentation totally dissimilar to the format of a spatial hypertext. Thus, existing approaches cannot be simply adopted.

Garnet instead adds the documents to the workspace, next to the document groups most relevant to them, as determined by text matching described later in this chapter. The spatial parser computes the visual location of each document. This placing will be discussed next.

3.5.1 *Automatically Placing Documents into a Spatial Hypertext*

Since the initial paper on Garnet [Buchanan *et al* 2001], Shipman [2002] has produced an extension to his Visual Knowledge Builder system (the successor to VIKI) which produces suggestions to assist the user, placing document representations into the user's workspace. His method of evaluation and his results are, therefore, of interest in the context of how a system may add objects to the workspace or suggest placement.

VKB introduced a wide range of different types of suggestions, many of which are unrelated to the context of using the user's organisation of space to support information seeking. For example, Shipman identifies potential visual structures to be used to organise a set of documents – this feature supports information structuring, but it does not amplify information retrieval, as the semantics of particular structures (be they implicit or explicit) are not well understood [Shipman 2001a].

A key question in considering how information structuring may support information seeking is how to express any relationship between the structures created by the user and documents in the information repository being searched. The natural means of demonstrating topical similarity in a spatial hypertext, as has already been discussed, is to use visual cues such as placement and colour.

What is of interest, then, is how Shipman's system suggests the placement and colouring of documents to the user.

Shipman perceives such suggestions as a means of assisting the user to organise documents, not as a means of information filtering, and they are performed on a document-by-document basis, rather than upon several documents at once. This is reflected in the goals of his evaluation, which seeks to detect the intrusiveness of the presentation of suggestions, not to evaluate the underlying value of the suggestions as a means of improving information seeking. Shipman's evaluation, therefore, gives us little insight into the utility of spatial hypertext structures for improving information retrieval.

Shipman presents his visual property suggestions through a textual presentation – a pop-up dialogue describing what should be done to the selected document. Therefore, the suggestion of placement is not done within the hypertext workspace, but rather in a separate, textual view. This would seem to be somewhat at odds with the highly visual idioms of Spatial Hypertext.

Shipman's system contains a number of controls to limit the intrusiveness of the giving of suggestions. Once a suggestion is repeated three times and rejected each time, similar suggestions are not given again without the explicit request of the user. Suggestions can be presented either in a pop-up dialogue, with a high degree of intrusiveness, or in a display below the hypertext, giving a much lower level of disruption to the user's task flow. Individual types of suggestion can be explicitly switched on or off.

The details of the implementation of the suggestions system are also of interest. As with Shipman's spatial parsing algorithm, there is no published data on the method or algorithm used to provide the suggestions, and so no lessons can be drawn there.

3.6 Spatial Hypertext – Summary

Spatial hypertext systems support a range of structuring facilities. All systems support implicit structuring, and explicit structuring is a common supplement to the freeform organisation of implicit structures. Explicit structures are readily identified to the computer system and implicit structures are much less tractable. A small number of idiomatic implicit structures have been identified and corresponding Spatial Parsers written to recognise them. Little is known, however, of the design and construction of these parsers.

Navigation is achieved through a number of means, typically either through explicit structures or through zooming. Each of these approaches suffers in large hypertexts, as many documents cannot be clearly seen and visibility of individual elements of the hypertext consequently falls.

The comprehension of a hypertext when its author returns to read it at a later date, or when it is placed before another person to study, is a known problem. This is believed to be a particular problem with implicit structures. Hence, some system support is required to assist in either the later recovery of the original context or the inscription of that context into the hypertext when it is written.

The representation of information sources in a spatial hypertext can be approached in two ways. Firstly, navigable structures such as classification hierarchies can be presented as part of the workspace. Secondly, systems that require some interaction, such as the input of terms and criteria to a search engine, can be represented through an interactive dialogue, with the resulting documents then being presented in the workspace. However, difficulties arise with providing visual indicators as to which elements in the combined hypertext are tied to the information system and which the user has placed and can control. Other problems emerge in expressing relationships between different instances of the same document in different contexts.

Finally, the automatic suggestion of relationship between documents has been partially addressed by recent developments in VKB. However, these are presented in a separate view to the main hypertext, which seems inconsistent with Spatial Hypertext idioms, and they have not been evaluated from a perspective of information seeking effectiveness.

3.7 Improving Information Retrieval

In the previous part of this chapter, I have discussed the various structuring facilities that can be found in a spatial hypertext system. How implicit structures can be identified in a Spatial Hypertext has been introduced – through the use of a Spatial Parser.

How, though, can a spatial hypertext be used to support information seeking, or improve information retrieval? The structure of the organisation of documents performed by a user may provide an insight into the user's topics of interest and conception of the themes within their sphere of work.

In information retrieval, a number of approaches have been taken to attempt to use topical organisation of search result lists to improve the selection of documents of interest –in terms of either speed or accuracy. The different approaches to presenting or altering search result lists that have been taken will be discussed in detail in the Information Retrieval and Clustering section later in this chapter. The overall performance of the classification of search result documents has had variable benefits, as again will be discussed later, but in comparison to a typical ranked list approach, comparisons have demonstrated that outcomes are at least broadly similar, and in some studies notably superior.

There are, briefly, three key approaches to the use of classification in search results:

1) Filter documents against a fixed, standard, classification of topics provided by a third party or authority – a document may fall within no, one, or several classifications.

2) Filter documents against a set of topic descriptions given by the user – again, a document may fall within zero or more classifications.

3) Cluster documents – the computer splits all results into a number of automatically-generated sets of documents, and each document falls within at least one of those sets.

If one were to build upon the organisation of a spatial hypertext, one could automatically generate classification filters based upon the organisation of the hypertext by the user (Option 2 above), giving the benefit of user-created filters without the explicit extra cost of creating them. In the next section of this chapter, how the spatial organisation of a spatial hypertext could be automatically identified – Spatial Parsing – will be described as a first step in the more detailed description of how such a filtering scheme could be implemented.

3.8 Information Retrieval and Clustering

As described above, given a spatial hypertext, visual groups of documents can be identified using a spatial parser. However, the visual identification alone does not provide a means of matching between those documents and others.

The matching of documents against a classification or grouping requires three elements: an identified set of groups against which the target documents are to be grouped; a set of target documents; and thirdly a means of comparing a candidate document against each group. In the context of using the user's organisation of documents as the classification system, the first phase is addressed through the use of Spatial Parsing. In this section, how the groups of documents identified through Spatial Parsing could be represented and matched with individual documents will be described.

Once the visual organisation of the documents already placed in the spatial hypertext has been identified, the resulting groups will then be used to classify other documents in one or more digital libraries. Which target documents are matched against the user's groups is another aspect of the matching process which can vary. Using large sets of target documents would be computationally expensive. Initially I will consider each target set to be the result of a query conducted in a digital library, and only a selection from the head of the result list of a query will be considered. Providing matches against a whole library would require one of two approaches. First, a sample query representing the document group could be submitted to the library - this could be simply achieved. Secondly, a list of all the documents could be retrieved and each individually evaluated by the spatial hypertext client, which would be very computationally expensive in terms of both time and storage. Therefore, this latter option has been discarded as unworkable.

The actual matching of documents can be approaches in a number of different ways. The first division is on the basis of document representation – whether documents are matched on some selected features of its text or descriptors, or on its full textual content. Given a particular representation, different forms of matching can be used. Some matching algorithms can be used on a variety of different document representations, others are highly dependent upon one particular representation. The area of matching documents to other documents or particular words is the field of information retrieval. For the purposes of this thesis, I only consider the matching of textual documents, and therefore only describe textual information retrieval methods.

The next sections will address the options for document matching, dividing the methods available on the basis of document representations: full text or metadata.

3.8.1 Document Text Approaches

Document text approaches all stem from the initial use of the document's original text, rather than descriptive data created separately from the document itself. From this root, the document can either be retained in its original form ("full text"), or distilled to a more compact representation that is smaller, and therefore in most cases more amenable to rapid computations of matches.

Full Text Matching

One simple approach is to take all the words in a given document set, and match that concordance against the full text of other documents. This approach has been researched for a considerable period of time, and the most common approach is to use the "log rule" which weights each word that occurs in a document (or, in the case of document sets, a set of documents) by the number of documents it occurs in within a closed, larger corpus. The more documents a word occurs in, the lower its weighting, and vice-versa.

The log rule is widely used for matching query terms against documents and therefore this approach is well understood, and considerable work has been done to improve and refine the basic method. Such extended methods include, for example, weighting words more heavily if they occur towards the beginning of a document, or altering the means of weighting words from the basic method.

Given the mature state of the log rule and its derivatives, good quality results can be achieved. However, a downside is the requirement of a closed, or balanced sample, corpus that can be used to generate the word weights – this is a problem to which I shall return. Furthermore, the use of full documents also may scale poorly – resulting in very large representations if a group has a large number of constituent works.

Alternative matching systems exist for the case of document-to-document matching. The classic use of these is to automatically gather a large collection of documents into a set of document groups, where the documents in each group share a topic or theme. Many of these methods rely on the same log-rule basis as query-to-document matching. However, others rely on evaluating the combination of words rather than taking words individually. Some of the word-combination approaches still rely on the log rule, others however do not rely on the knowledge of word frequencies that lies at the heart of the log rule, and so can be used without knowledge of the entire document corpus. An example of this would be the Grouper algorithm of Zamir *et al* [1997].

Selected Keywords

Another approach is to select a few key words from the document and use them as the input to some form of information retrieval system or textual mapping algorithm. Again, there are two parts to the process: identifying key words in existing documents, and secondly matching those key terms against other documents.

One method for obtaining the keywords is by *frequency analysis* – the most common words in a document being used first. However, a problem here is that, clearly, common words such as 'the', 'and' or 'it' would appear regularly in the sets of words for a document, and also provide poor distinction between topics. Three solutions exist to this difficulty:

1) Eliminate known common words, often termed *stop words*.
2) Use knowledge of the frequency of words in the text corpus to eliminate common words
3) Use words that are, by some heuristic measure, not the most common, but more common than the 'average' – e.g. the second quartile of words by frequency.

The problem of matching the selected words to other documents can also be approached in a number of ways. One is to simply use the selected words as inputs to a search engine – the exact matching method may not then be known, but this has the advantage of not relying on knowledge of the target corpus (though, clearly, knowledge of the controls of each search engine is required). Another is to match the selected words against the set of words obtained from the other documents by the same extraction process.

Matching techniques in the case of using a search system are usually going to be based on the log rule for information retrieval [Witten *et al* 1999], which we have already met in pure full-text approaches. This uses word weighting to increase the significance of rare terms and to reduce the value of common terms. The advantage of such an approach in terms of selectivity is that log-rule based retrieval systems are extensively researched and most systems yield results of good quality. As already stated, no knowledge of the exact matching algorithm is required, and no implementation of a matching algorithm needs to be provided. However, the quality of the terms supplied is important. As we shall see, term quality can be a problematic challenge.

Alternative systems for performing document to document matching come in a variety of implementations. As well as the log-rule method, which can also be used with large documents, statistical and word-pairing techniques are viable matching systems. Word-pairing techniques are commonly found in clustering systems (described later in this chapter) and rely on maximising the number of words found in each document or document set. Statistical methods often rely on frequency weighting as with the log rule, but with alternative weighting values and with additional Bayesian tests for the co-occurrence of words.

Where my system would have to rely on performing the matches itself, it would need to generate and prepare representations for all possible matches – i.e. for every possible document. For large target corpuses of documents, such a task would clearly be time-consuming and result in long delays in interaction. Therefore, using keyword representations could only be done effectively on small, selective, target document sets.

Selected Phrases

Related to the use of key words, an extension of this approach is to select phrases rather than words. Phrases are longer and possibly more selective than words. Some of the problems observed with selected words also exist with phrases, however. For example, 'it is' may well occur regularly both within a collection of documents and within a given document. Similar approaches can be taken as with the three methods outlined above in keywords. However, it is more common for keyphrase extraction tools to use machine learning or similar methods that rely on training. Often, the training method eliminates the identification of common, often meaningless, phrases [Witten and Frank 1999].

After extraction, the matching challenge can be met by using a similar range of techniques to those used with keywords above.

Intelligent Classifiers

An alternative approach that could be used with all the representations above would be to use some form of "intelligent" or "learning" matching system which is trained by the set of documents in the spatial hypertext, and then used to classify the new documents found elsewhere. As this thesis already requires expertise in a number of different fields, I dispensed with this approach.

Citations and References

In the context of an academic digital library, the use of citations as a basis for representing a group of documents may be highly effective. However, digital library protocols do not generally support access to the citations of documents as separate, organised objects [Bainbridge, Buchanan *et al* 2001]. Though methods exist for the automatic extraction of citations from documents [Giles *et al* 1998], they are necessarily not of totally reliable quality.

However, were such information available, research has indicated, e.g. CiteSeer [Lawrence *et al* 1999] that there can be a very positive impact upon the quality of retrieval achieved. Like the PageRank algorithm [Brin and Page 1998] used by the Google search engine, documents which are the target of citations from other matching documents, are considered of high quality, where as those which are not referred to by other works are considered to be of less relevance.

Summary of Full Text Approaches

We have met a number of different means of representing documents that are based on their original text. The approaches which use a reduced representation – citations or keywords for example – have the advantage of requiring considerably less storage to represent a document, and the lower level of computation required in comparing documents also reduces comparison times. However, the creation of these reduced forms can be computationally intensive, and – as in the case of citation extraction – can still be open research topics in themselves. One potential approach to reducing this cost, obtaining these representations from the libraries themselves, is hindered by the low level of availability.

Keyword and keyphrase representations are also based on extraction, but the methods are more mature. Such representations can also be used in combination with traditional search tools, giving a higher degree of flexibility in the choice of matching process. However, the general advantages and disadvantages of reduced representations remain.

'Pure' full-text approaches have the benefit that the original library text is often easily obtained, and they also require no further processing to generate a representation. However, merging representations and comparing them can be more time consuming, given the larger structures involved, and unprocessed they cannot be used as inputs to the search mechanisms of libraries.

3.8.2 Metadata Approaches

In addition to matching systems based on the text of documents, in a digital library, matching can also be achieved via the metadata stored on documents. For instance, documents can be described and retrieved on the basis of author, title or date of publication. Abstracts, categories and descriptors can also be used to discover documents of interest.

Some of these approaches – particularly in the case of abstracts – rely on the same matching mechanisms as are used in document-text approaches. However, metadata and abstracts are much more compact than the full text of documents, resulting in lower processing and downloading costs.

Where alternative matching methods are met, the most common are based upon Boolean string or numeric comparison. For example, date matching is usually expressed as a range, sometimes implicitly as in the use of a year, which is either matched or not. Though Boolean methods are very poor in the context of full text searching, they remain effective for the smaller, vocabulary-controlled, word strings generally found in metadata descriptors.

In the context of cross-collection retrieval, however, metadata approaches are problematic. Where the metadata fields or their titles differ between collections, matching cannot be performed across documents in one query. In the case of representing sets of documents from various sources, even representing the group can be problematic for the same reason. Standard metadata formats such as the MARC description format (see http://www.loc.gov/marc/) exist, and where used consistently can facilitate cross-collection searching. Common standards such as unqualified Dublin Core (see http://www.dublincore.org/) seem to offer one means of achieving cross-collection searching. However, a common problem is that the semantics of descriptors becomes confused and inconsistent. For instance, the Dublin Core standard includes a "date" field – but does not specify what the date's relationship with the document is. The date could be that of a conference, the date of publication or the date of accession to a library (or, indeed any other date whatsoever); similarly, the standard also has a "creator" field, which may mean the author, but could equally be the author's employer.

However, even the use of a standard format does not solve the problem – standard vocabularies also need to be used, so though one facet of the problems of cross-collection work is addressed by standard formats, even that is not enough. Beyond formats, standard vocabularies and descriptors are also available, for example the Library of Congress In-Publication data.

Finally, the availability of metadata on different digital library protocols varies widely. Some, such as the Greenstone protocol [McNab *et al* 1998], are metadata format neutral, whereas others require conformance to one or more metadata standard (e.g. Marc, RFC1807, Dublin Core).

Relying on both a standard format and standard vocabulary will result in a much-reduced selection of digital libraries with which one can work, either as a result of the format the library uses, or the protocols with which one has to work. In the case of the system at hand, the effort required in mapping searches across formats and vocabularies is a significant piece of work in its own right. Therefore, attractive as such a mechanism would be, it would only serve as a distraction from the work at hand and I will not use metadata approaches for matching documents.

3.8.3 Clustering

Clustering algorithms were introduced in Chapter 2 as a means of providing browsing over search result sets or entire document collections. One part of a clustering is the matching algorithm, and the previous sections of this chapter have discussed the available techniques at some length. However, in addition to matching documents against groups of documents, clustering algorithms must first generate the groups of documents. This is generally done through iteratively merging documents and document groups into progressively larger 'clusters'. The decision whether to merge a particular set of documents into a single document group, or cluster, is taken using a "coherence measure". The candidate merger that scores the highest coherence measure is performed and then the algorithm again computes the possible mergers until some point at which the best possible merge option scores less than some fixed threshold. Example clustering algorithms include those of Zamir *et al* [1997] and Cutting, Karger *et al* [1992], which have already been briefly mentioned. A high score for a document group should indicate a set of documents with a high level of thematic, textual consistency, whereas a low score would conversely suggest that a group is topically diffuse and heterogeneous.

The coherence measure of candidate document groups and how a decision to halt the further merger of groups is taken represent two key differences between the quality of the clusters created by individual algorithms. Zamir *et al*'s unusual algorithm yields quality results when compared against a number of proven approaches, whilst Scatter/Gather is a mature and well-respected technique.

These coherence measures and halting conditions of these algorithms provide one means of assessing the quality of the organisation of a set of documents. In the context of this thesis, such techniques provide a means of evaluating the consistency of a user's organisation of a set of documents.

3.8.4 Implementation Concerns

Group Representation and Recomputation

Clearly, once groups are identified, the representative text for each group needs to be generated. In the section above, we have described various means by which that representative text can be generated. However, in an interactive tool in which groups could change rapidly and repeatedly, one significant issue is to decide when the groups are both identified, and when the group representations are regenerated.

When the generation of the groups and their representation requires significant computation to be done, that computation will have to be controlled so that it does not impede the feedback delay between the user taking an action and the system responding. Therefore, one needs to manage the evaluation of changes to minimise the impact on system responsiveness. The basic approach of 'eager' evaluation, where re-computation is performed whenever a change is made is clearly inappropriate. 'Lazy' evaluation can be performed instead, in order to minimise the total amount of re-computation required. Work can be done when the system is otherwise idle, or only when a representation of a group is required for matching.

Cross-Collection Weighting

As we have seen above, most document matching algorithms use word weighting, to emphasise the use of rare words and reduce the significance of common terms. This requires knowledge of the relative frequency of different words, in terms of the proportion of documents in which they occur.

This could, in principle, either be achieved by processing all the documents in a library, or by obtaining the information from a library directly. However, in the case of a spatial hypertext system that is to work across several library systems and many documents, retrieving all documents is unlikely to be practicable in terms of communication capacity and computational effort.

Similarly, we are unlikely to be able to obtain data on the relative frequency of the words from the libraries themselves. Though protocols exist for searching and retrieving documents from libraries, they do not provide facilities for extracting term frequency data. Thus, we cannot achieve much information on words in documents which are *not* returned by a given search or browsing activity.

If the word weighting were available for a given collection, then further problems arise when document matches are conducted on different collections. A word may have significant weight in one collection, because it is rare, and in another have little impact on the result, because it is commonplace.

Words which are common in the documents selected by the user to be of interest may be either universally common, or many simply be particular to the interests of the user. Given that the user's workspace is likely to be a selective, biased sample of the documents available generally, the distinction between generally common terms and terms common in the user's chosen documents cannot be deduced from the specialised sample to hand. Similarly, rare words in the workspace documents may well be common in other collections of material.

Thus, obtaining global information is problematic, and time and space are costly. Furthermore, results of matches will vary, even for the same document, across collections. On the other hand, the local information available to a spatial hypertext is very likely to be an unrepresentative sample. Consequently, relying on those matching systems that rely upon word weights is going to be problematic.

Fortunately, not all document-matching systems rely on word frequency. For example, Zamir *et al*'s clustering algorithm [1997] builds groups of documents where the documents within each group match each other well and other document groups poorly. The means by which this is achieved does not rely on weighting individual words by frequency, but rather on maximising the number of words common to *all* the documents of a group, or cluster. In terms of effectiveness, Zamir *et al*'s algorithm is of very high quality, outperforming many well-accepted and refined systems.

Thus, an adaptation of Zamir *et al*'s algorithm for manually created groups may yield good results also.

3.9 DL Protocols and Communication

DL Protocols permit remote, client applications to access the facilities of a digital library. The limitations of the information that can be obtained from a digital library using a given protocol will have a significant impact upon what facilities can be offered and the interactions available. The common features and differences between different digital library protocols also affect the generalisability and portability of any system that uses them.

This section briefly reviews the different digital library protocols that are available, reporting their common features and differences, and the impact these had upon Garnet and has on DL client interfaces in general.

The proposed spatial hypertext system must either access or generate representations of documents for the purpose of building representations of individual documents and groups of documents. Thus, the information available through DL protocols will have consequences of upon the choice and implementation of document representation and matching systems.

3.9.1 Established Protocols

There are four main protocols that have been the subject of published works. First, the rich and extensive Z39.50 protocol [ANSI 1995] is commonly supported by large university and governmental libraries; the SDLIP [Paepke *et al* 1999] and Dienst [Lagoze and Fielding 1998] Protocols were developed by Stanford and Cornell universities respectively, and have been used by a number of research digital libraries; finally, the Greenstone protocol [McNab *et al* 1998] produced at the University of Waikato has been used in a number of systems operated by the United Nations and a number of non-Governmental Organisations, as well as by a variety of academic research groups. In addition, there are a number of more specialised protocols such as the Open Archives Initiative (OAI) protocol [Lagoze *et al* 2003], which provide a more limited breadth of functions.

3.9.2 Common Features

Most protocols support searching, browsing, and the retrieval of both the metadata and text of individual documents. A few, such as the Open Archives Initiative (OAI) protocol, only support a subset of these features – in the case of OAI, browsing and the retrieval of metadata – however, these protocols are not intended as general DL protocols. That the different protocols can be successfully mapped onto each other across these standard features in practice can be seen from demonstration systems such as the Z39.50 to Greenstone and Greenstone to SDLIP translators [Bainbridge, Buchanan *et al* 2001], Dienst to SDLIP converter [Paepke *et al* 2000] and a number of similar projects.

A more detailed analysis of the common and differing features of the protocols can be found in [Bainbridge, Buchanan *et al* 2001].

3.9.3 Web Search Engine Protocols

So far, I have briefly discussed Digital Library protocols. However, in contemporary electronic information environments, as with physical environments, information seekers are prone to use a variety of sources. The Web is a very large repository of information, and some of the facilities commonly found in digital libraries are now available across the Web also. For example, search engine companies such as Google index substantial portions of the Web. Google and other search engine providers provide protocols across which client applications can obtain results of searches. The Web readily permits the retrieval of individual documents using hypertext links, and thus also provides document access. However, on the Web metadata is seldom available for a given document, and few search engine access protocols provide browsing structures. Therefore compared to the four common features of DL protocols: search, browsing, document retrieval and metadata access, Web search protocols provide a subset of the features of the DL protocols, generally omitting the browsing and metadata facilities.

3.9.4 Protocol Modularity

A trend in DL protocol development is away from the original, monolithic approaches where every server was expected to provide every single defined service. In the case of the large Z39.50 protocol, many extensions and options are available. However, these are still defined specifically in relation to the original service – i.e. they are specifically extensions rather than supplementary modules that could be used independently or with another DL protocol.

A recent paper by Suleman and Fox [2002] proposed a purely modular approach, where the separate facilities of a digital library were divided into individual components, each with its own transactions and communications. These component modules could be used in combination as was appropriate to the services offered by a given server or library. For example, a 'search' module would index documents, but would not contain any means of recovering the documents themselves – instead, a separate document storage module would provide that facility. All that is common between modules is some consistent scheme of identifying individual documents.

These two themes were introduced in the original Greenstone protocol paper [McNab *et al* 1998] where future facilities were seen as supplementary and generalisable to being adopted with other protocols, or alternatively as extensions of the existing protocol which were specific to it.

In the case of Garnet, the best practice would be to provide any additional communications as modular supplements to the range of existing DL protocols rather than as an extension to one specific protocol. This exploits the modular approach of Suleman and Fox, yet provides backwards compatibility with existing protocols.

3.9.5 Protocols, Profiles and Clustering

As already mentioned, the proposed spatial hypertext system must be able to represent the topic of documents and groups of documents in the workspace. These will be used to relate new, unseen documents to those already selected and organised in the workspace.

In the case of cluster-based representation and matching algorithms, document profiles need to be accessible. If pre-prepared profiles are not available, then it will be necessary to create those profiles from the information that can be obtained from particular libraries. Such a feature is certainly not found amongst the common features of protocols I presented above and in [Bainbridge and Buchanan], nor in protocols published subsequent to that paper. Therefore, the profiles must either be provided by some new, extended DL protocol, or alternatively generated by accessing features of a given document through the protocol over which it is accessed and then processing those discovered features (e.g. the full text of the document) into a representative profile.

Information on word weights is not available from any DL protocol. The impact of this limitation on the matching of documents and the forming of document groups has been described above in the section on cross-collection weighting.

3.10 Conclusion

Throughout this chapter, a number of technical issues related to the construction of a combined Spatial Hypertext and Digital Library have been discussed. I have reviewed the current understanding of the use and identification of structures within spatial hypertext, the technology aspects of Information Retrieval which are relevant to the exploitation of a user's organisation of documents in a Spatial Hypertext workspace, and the limitations of DL protocol systems. This chapter will conclude by summarising the key impacts of these considerations upon the construction of a combined system.

In §3.2, it was observed that the merging of an information repository and a spatial hypertext provided a number of interface challenges within the spatial hypertext. Two example problems were how to disambiguate the visual cues used for objects controlled by the system or by the user and how to represent information repository features such as search engines and browsing structures within a Spatial Hypertext idiom.

The later examination of visual interfaces to digital libraries, §3.3, provided a number of helpful insights to such interface problems. For example, when performing searches over time, visual interfaces to digital libraries have often supported the long-term storage of search information in the form of search histories and placing searches as objects into a common workspace. Such elements represent entirely different types of object to the documents usually found in spatial hypertext workspaces and need to be visually differentiated from documents and each other. Visual interfaces to digital libraries used distinct visual cues to distinguish between different elements of the workspace. A common approach, seen in DLITE [Cousins 1997], is the use of different shapes to distinguish between search services, search result sets, documents, etc. Though the use of shape to communicate the type of an object could reduce the range of expression for altering the appearance of documents, it is worthwhile observing that the successor to VIKI, VKB [Shipman *et al* 2001], in fact eliminated shape controls for document presentation. Thus, using shape of objects in the workspace to communicate their role and type, whilst eliminating the option of controlling document shape, would seem to be consistent with both spatial hypertext and visual digital library interface practice.

Returning to Spatial Hypertext concerns, implicit structures were found to be key elements of spatial hypertexts that, as was seen in Chapter 2, are important features of information structuring and long-term information work. Unlike the explicit structures that emerge over longer periods of time as documents of long-term interest are archived, implicit structures are more difficult for a computer to identify. However, they are often salient to the immediate information needs of information workers, and therefore may be more appropriate in supporting near-term information seeking activity. It is therefore important to be able to identify implicit structures.

In §3.4, the operation of Spatial Parsers and the current state of research in this field was briefly reviewed. This clarified that it should be possible to identify implicit structuring of groups of documents in a spatial hypertext using a Spatial Parser. As has been seen in §3.7, document structures may used to organise search results, and structures of documents can be represented textually in a number of manners. Thus the next issue of concern is how the visually identified groups could be represented and used to organise search results. This falls within the remit of information retrieval research, discussed in §3.8.

The matching of documents based upon their body text is a technique that can be used without the complication of mapping across different metadata schemes and DL protocols [§3.9]. Therefore, document text matching schemes avoid the problems that may emerge when working across different digital library collections. In addition, algorithms for comparison exist which do not rely upon knowledge of the frequency of a corpus, and this avoids numerous problems in obtaining and using such information. Zamir *et al*'s clustering system utilises a document matching algorithm which only requires document text and does not require knowledge of the frequency of terms across a closed corpus. Such an approach is therefore a clear candidate for matching documents in the context at hand.

Once a document is matched against the informal structures found in the spatial hypertext workspace, that connection needs to be communicated to the user. In §3.5, some relevant spatial hypertext practice was outlined. However, there are problems with the methods described there, where suggestions are given textually in dialogs separate from the hypertext workspace – a means somewhat at odds with the visual, interactive idiom of spatial hypertext. In NaviQue, the similarity engine would use coloured highlighting of documents in their current position to indicate any matches found. This alternative approach requires extensive browsing to find the documents that are similar to a given group, and for the document to already appear in the workspace. Thus, existing approaches cannot be used. Instead, I use the Spatial Parser to choose the location of the matched document and then introduce a label to represent it next to the matching group.

Finally, another consideration is the means of accessing the digital library that is accessed through the spatial hypertext interface. DL protocols [§3.9] contain a number of common features that include the standard methods of information seeking – searching and browsing – and these features can be successfully mapped between different protocols. Therefore, from the point of view of presenting an interface to a remote DL, any DL protocol can be used, as it would be functionally equivalent to others in these standard tasks. In addition, any extended information required may be generalised across DL protocols by taking the modular approach recommended by existing DL research, and for the purposes of efficiency, it would be wise to obtain the profiles of documents through such a mechanism.

The creation of a spatial hypertext interface to a Digital Library which extracts structural information from the organisation of the workspace and exploits this computationally would provide a basis for investigating multi-actor co-ordination and computation over Spatial Hypertext, both acknowledged research questions in Spatial Hypertext research [Shipman 2001a]. Therefore, the proposed system would facilitate research into both the fields of Digital Libraries and Spatial Hypertext.

In the next chapter, the operation of the implemented system will be described, and a number of the finer details of design and implementation left unaddressed in this chapter will be discussed.

Chapter 4: Garnet Implementation

4.1 Introduction

In Chapter 2, the interwoven nature of information seeking and information structuring was introduced. It was observed that although the two processes were interconnected in physical media, digital environments inclined to support one process whilst neglecting the other. In Chapter 3, a combined system of a digital library, supporting information seeking, and a spatial hypertext, supporting information structuring, was introduced as a system that could be used to discern the benefits of a combined digital environment.

This chapter describes the operation and implementation of the combined spatial hypertext and digital library system I have created, called Garnet. Garnet not only contains an integrated implementation of the two previously separate systems, but also includes novel features which are intended to explore the means by which the artefacts of information structuring – the organisation of documents formed by a user – can be used to support later information seeking. The digital library facilities of Garnet are supplied by the Greenstone Digital Library software, whilst the spatial hypertext and the interface between spatial hypertext and digital library have been created from scratch. The evaluation of this system is described in the succeeding chapters.

This chapter proceeds in two parts.

First, the use of Garnet is described and demonstrated with a description of the software in use, highlighting the facilities that support information structuring and the means by which information structuring and seeking are combined. Much of this part of the thesis has already been presented in [Buchanan *et al* 2001, 2002].

Secondly, I will discuss the internal construction and operation of Garnet. The implementation and selection of the digital library and spatial hypertext sub-components will be examined in turn, before discussing the connection of the two as a single system.

4.2 An Illustrative Scenario

I now demonstrate Garnet in use, starting from a 'bare' workspace. Before commencing the construction of a new hypertext, I will first illustrate the sort of hypertext, or workspace, which a user may arrive at, to provide some context within which the construction of a hypertext can be understood.

Overview – A Simple Hypertext

In Figure 4.1, we see a 'typical' Garnet user session in progress; a number of 'windows' appear inside the main browser window. Each of these is a 'collection' of materials that the user has recorded in the current, or a previous, session. The content to the main browser window is termed the 'root' collection, and the window that appears inside it is a 'child' collections. The behaviour of a collection window is similar to those in a contemporary filing system. Within the collections, individual documents are represented by a rectangle containing some text, as indicated in the diagram, which I term a 'document label' for simplicity.

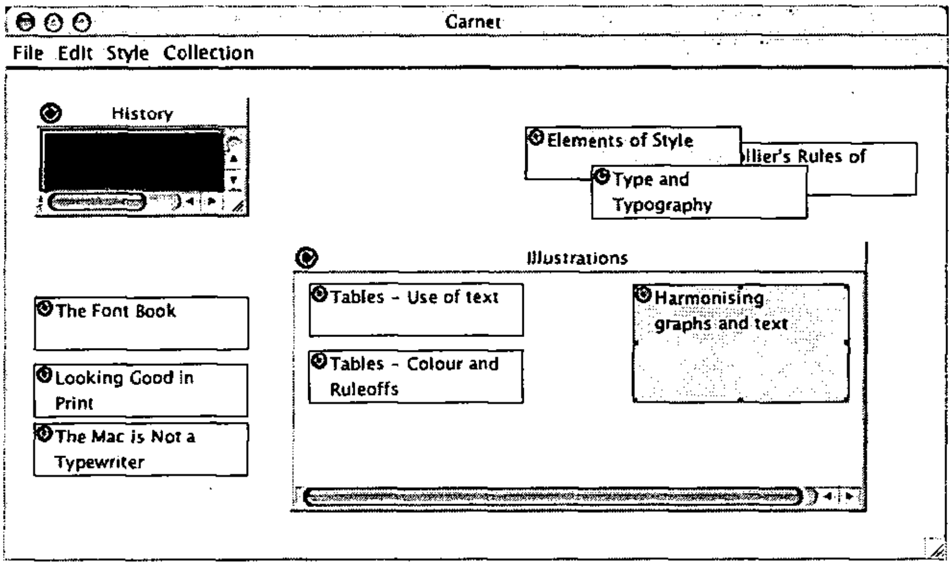


Figure 4.1. A Garnet client in use

Within a collection, the user is free to place, size and colour each document label as they see fit – the space is entirely freeform. Labels can be moved and/or copied between collections in the usual drag-and-drop style of similar direct manipulation environments. Document labels can be added to the workspace in two ways: either explicitly by the user creating or indirectly through interaction with a digital library's facilities – e.g. after the user requests a search.

Using the drag-and-drop manipulation of document labels the user is free to form labels into implicit, freeform structures of their own making inside collections, and in a more formal organisation by using the explicit hierarchical forms of a set of document collections. Taking the example above, we have a collection called "Illustrations", which has a column of documents on the left-hand side, and a single document on the right. The column is a structure created by the user's exploitation of space – it is not a feature enforced by the system. The column idiom [Marshall 1993] can also be seen in the root collection – again on the left-hand side. Spatial hypertext can use cues other than position to suggest organisational structure or to relate documents visually. For instance, some use of colour can be seen here, but the relationship between colour and meaning is not clear to us as readers, though it may well be meaningful to the author [Marshall 2001].

4.2.1 Scenario

For the purposes of this example scenario, say that the user needs to investigate the practicalities of snail farming, and wishes to confirm that the requirements of that form of agriculture are compatible with the available resources. They have chosen to consult the Humanity Development Library of the United Nations, one of the widely available examples of a Greenstone library collection, which consists of several thousand pages of agricultural information.

4.2.2 Performing a search

I will now trace a simple sequence of interactions, starting with an example search. With Garnet loaded, the user starts a new search in the Greenstone digital library system (Garnet also supports web searching), and enters the simple query "snail". In Figure 4.2, a simple collection window is seen with a number of document labels appearing one beneath the other, similar to a typical web-based results list.

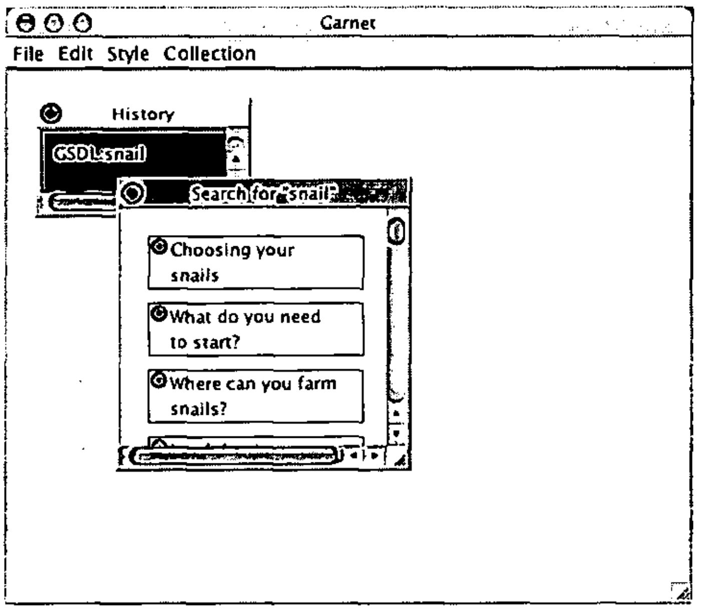


Figure 4.2. A simple search for "snail" appears on an empty Garnet workspace

On reading the first two documents (achieved by a simple double-click on the corresponding document labels), the user decides that they would like to keep the second document ("What do you need to start?"). To do this, they move the document label onto their root workspace window by dragging the document from the "Search for 'snail'" window onto the main Garnet window.

The first document ("Choosing your snails"), however, seems too advanced, and the user deletes it from the list by clicking on the small red 'circle' on its top left corner. As a result of this, the later documents move upwards. Should the user wish to return to the search results at a later date, by default these changes would be retained. Alternatively the original, unedited, form can be restored at the user's discretion. The workspace as it would now appear is shown below in Figure 4.3.

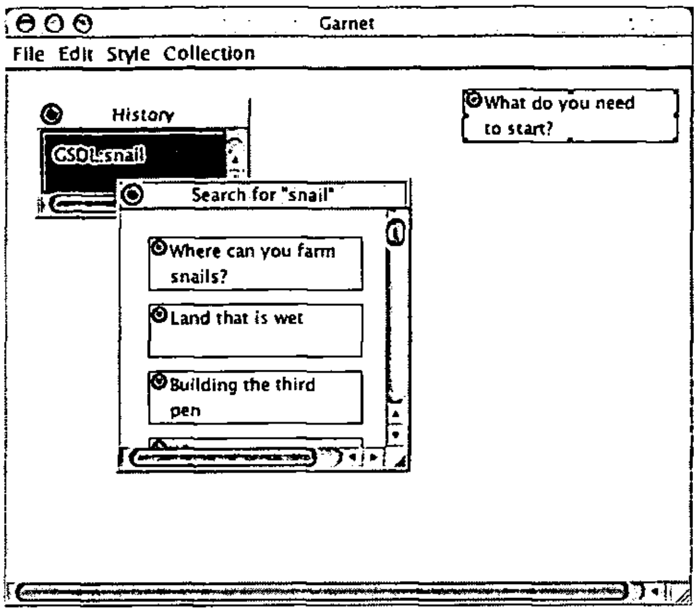


Figure 4.3. The workspace after alterations; one document has been moved onto the workspace from the search result list and another has been deleted.

4.2.3 Exploiting the User's Information Structuring

In the previous section, the user performed a plain search and stored some documents in their workspace. Garnet can exploit the organisation of these documents in a novel manner. A set of documents (including search result lists) can be "scattered" over the existing layout of documents in the workspace. "Scattering" places the search documents near to groups of existing documents with which they have a strong similarity. This provides a filtering service over the documents in the search result list similar to clustering, but placing new suggested documents next to the document groups to which they bear a strong similarity. The use of this filtering facility will now be demonstrated. Continuing the previous example, the user has now selected a few more useful documents, but let us suppose that a couple of questions remain unanswered.

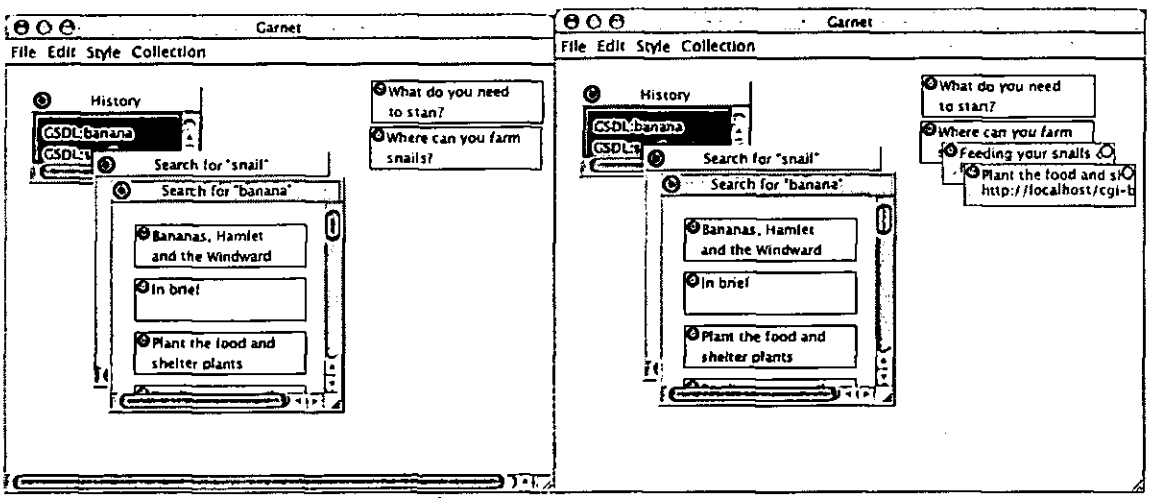


Figure 4.4. Before (left) and after (right) a"scatter" – note the shaded document labels added on the left-hand image.

A plentiful supply of bananas is available, which the user would like to use to feed the snails, but they are not sure whether this would be viable. If they were to do a naïve search, on "banana", the initial results will not match their particular interest well (Figure 4.4 - left).

In fact, documents that relate to their interest can be found in both the 'snail' and 'banana' searches. However, these documents may not appear in the visible part of either list. Normally, the user would have to re-work their query to make it more targeted. In the case of Garnet, they can use the 'scattering' feature to discover any material similar to documents that they have already selected onto the hypertext workspace. Or, in other words, Garnet can generate additional search terms or perform filtering (§2.4) to represent our user's interests, based on the workspace layout they have already created.

Viewing Figure 4.4 again, if the user selects the "banana" search results window, and then does a 'scatter', (right), a subset of the "banana" search results appears on the main collection. This small subset, which appears in a light grey below, has been found by Garnet to be a close match to the existing pair of documents, which appear in white. Suggestions are always displayed in this grey colour, and below and to the right of the group of documents that they are believed to be similar to.

Note the third item from the top of the search result list (for clarity a visible document has been chosen) "Plant the food and shelter plants". This item, certainly more clearly of likely relevance than the two above it, is one of the two suggestions given by Garnet. The other suggestion – "Feeding your snails other food" – actually appears lower down the list, outside of the visible area of the search result list. Indeed this document lies beyond the top twenty items shown in the list by default. Thus, information that may be of low relevance when scored simply by its relevance to the user's query text can be more precisely identified using the user's organisation of documents.

The user can now investigate the two suggested documents that are similar to the previously selected pair, double-clicking on the suggestions to read them as for a 'normal' document label. As it happens, these documents would confirm that ripe bananas could indeed be used to feed snails. If the reader wanted to permanently add one or other suggestion to the workspace, they can click on the 'circle' which appears on the top right corner of each of the suggestions.

The current suggestions are cleared when the user clears the suggestions by choosing the relevant menu item, or when another set of documents is scattered.

The user may now continue their task, perhaps choosing a colour for some documents through the display controls available in Garnet. Below, the suggested document that the user selected above and the earlier document concerning food have both been coloured, and the user has also added their own label to describe the documents that they have found. This latter facility supports the user's recall of the basis upon which they performed their organisation, when viewing the workspace at a later date – a problem observed by Cathy Marshall [2001] and already noted in Chapter 3 [§3.2.3].

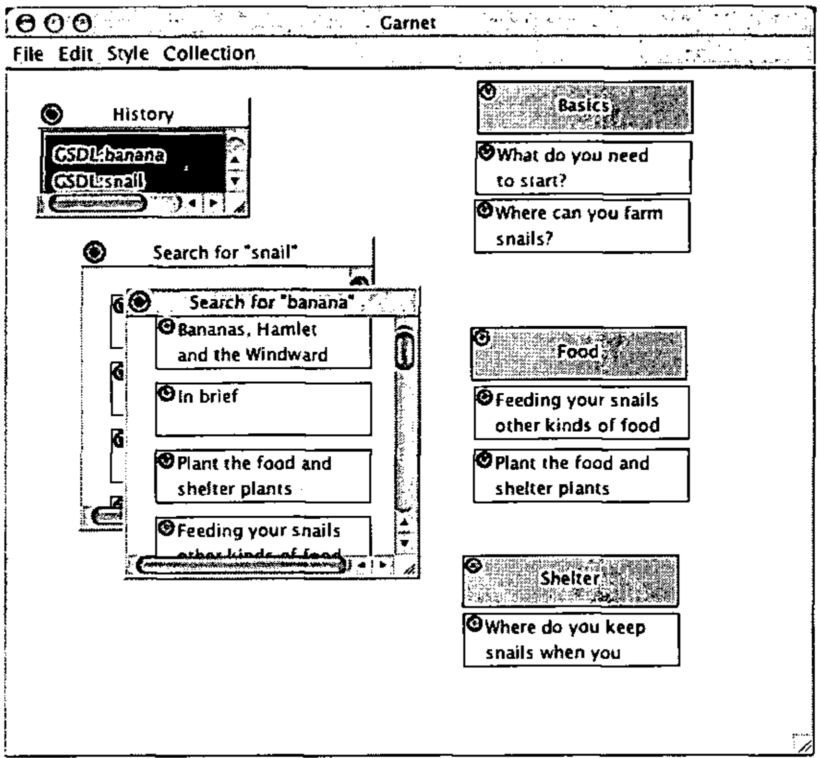


Fig. 4.6 After colouring and labelling; three labels have been added to the workspace, and two documents have been coloured.

In addition to being used in 'scattering', each document group can be used as a basis for a search of a library collection, to find similar documents using the group alone. To do this, the user first selects the document group. Then they select the "Find similar" option from Garnet's menu.

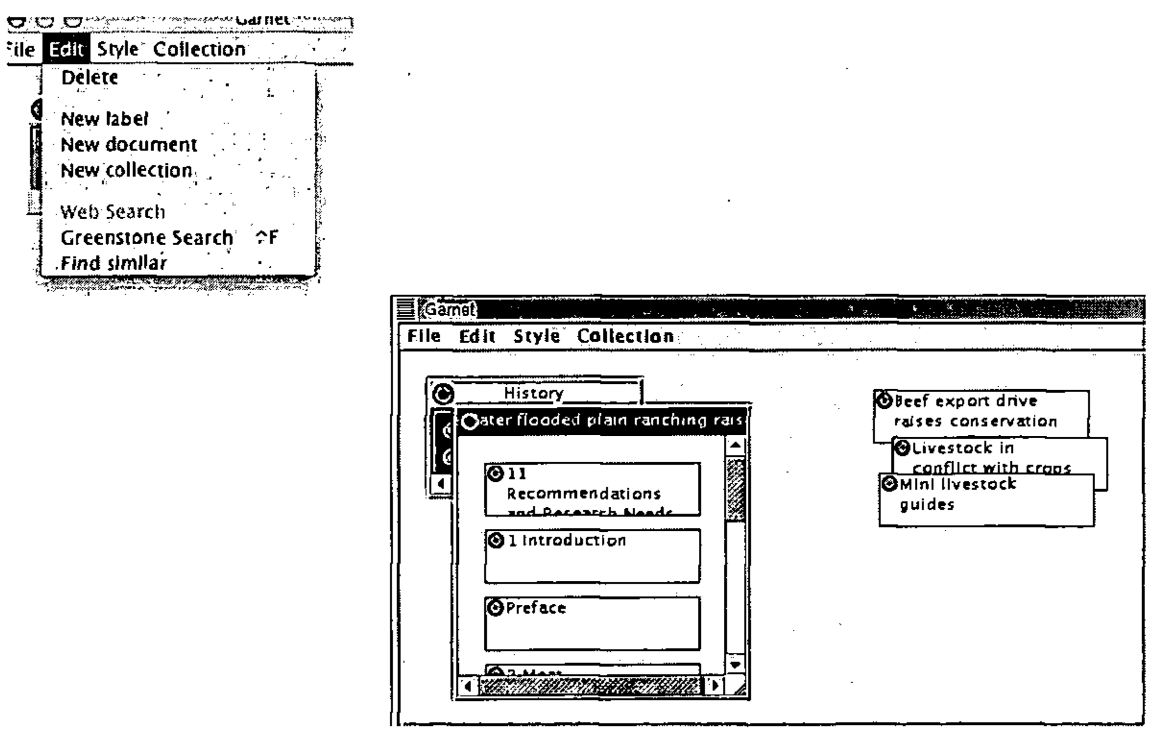


Fig. 4.7: (Left) "Find similar"from the Edit menu can be used to discovering documents similar to a given

group. (Right) – The three documents in the pile of documents on the right-hand side of the workspace seeded the search results displayed on the left.

This produces a new search result list set, which uses the words taken from the group of documents that they had selected. In Fig. 4.7, the three items on the right were used to create the search list seen on the left (when the search list is displayed, the group becomes unselected).

4.2.4 Summary

In this section, the basic operation of Garnet has been illustrated in a simple example, including the selection of documents onto the workspace, the searching and browsing of a digital library and the changing of the appearance of a given document. The use of informal document groups as both a means of filtering search results lists and a basis for searching the library has been demonstrated. The next section will commence the discussion of the technical infrastructure that facilitates this interaction.

4.3 Garnet System Design

The operation of Garnet from the perspective of a user has been described in the previous section. The following sections will describe the construction of Garnet as a system. As Garnet combines Digital Library and Spatial Hypertext features, it contains a series of features that can be particularly associated with each of these separate systems. For instance, the positioning of document labels is strongly related to Spatial Hypertext and the retrieval of documents themselves to Digital Libraries. However, in addition to these separate elements, Garnet provides connections between them, as I have shown above. Thus, after the user has requested a search, it is performed in a digital library system before the matching documents are displayed in the spatial hypertext interface.

This section describes the connection of the digital library and spatial hypertext at a system level, describing the internal construction of Garnet in terms of its components. Then the separate elements of digital library and spatial hypertext will be discussed in the following two sections. Finally, the lower-level details of some of the interconnecting elements will be explained before the chapter turns to reflecting upon Garnet's contribution and novel features as a system.

4.3.1 System Architecture

Garnet combines spatial hypertext and digital library systems and elements of each can be found within it. From Spatial Hypertext Garnet inherits a visual workspace and its associated structures: the spatial parser which identifies document groups, and the document groups themselves as distinct elements of the system. A simple diagram of the parts of a Spatial Hypertext system that are of interest is illustrated below:

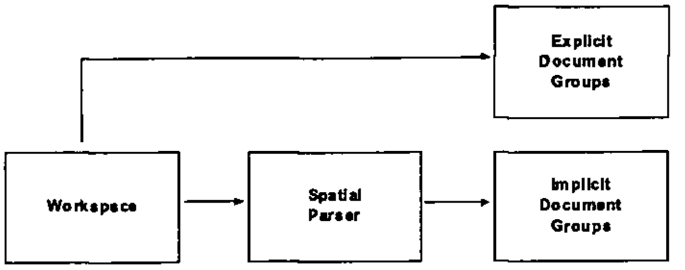


Figure 4.8: Simple schematic of a Spatial Hypertext.

This model is necessarily grossly simplified. The 'Workspace' in the above diagram includes the visual and logical representation of each document, including colour, title etc. whereas the groupings of documents are shown as separate items, as these are of particular interest.

With Garnet, a number of extensions are made to this basic model:

1. A connection to a digital library server for selecting documents through searching, browsing, etc. and for retrieving documents from the library.
2. A repository of textual profiles. A textual profile must be generated for each of the implicit and explicit document groups, and the profiles of individual documents retained to allow group profiles to be altered when documents are moved.
3. A similarity engine to extract information from the document group profiles and match individual documents against the group profiles.

The overall organisation of the expanded Garnet schematic can be seen in Figure 4.9 below. A number of new elements, for instance the group profiles repository and the similarity engine, can be seen in the diagram. In addition to additional components in the Spatial Hypertext area, separate elements comprising the digital library subsystem are also highlighted. Then there are further elements that 'combine' the digital library and spatial hypertext parts and these form the core of the unique functionality of Garnet. In the diagram below, the extensions to the basic spatial hypertext system are highlighted in grey and the corresponding number for its function from the list above.

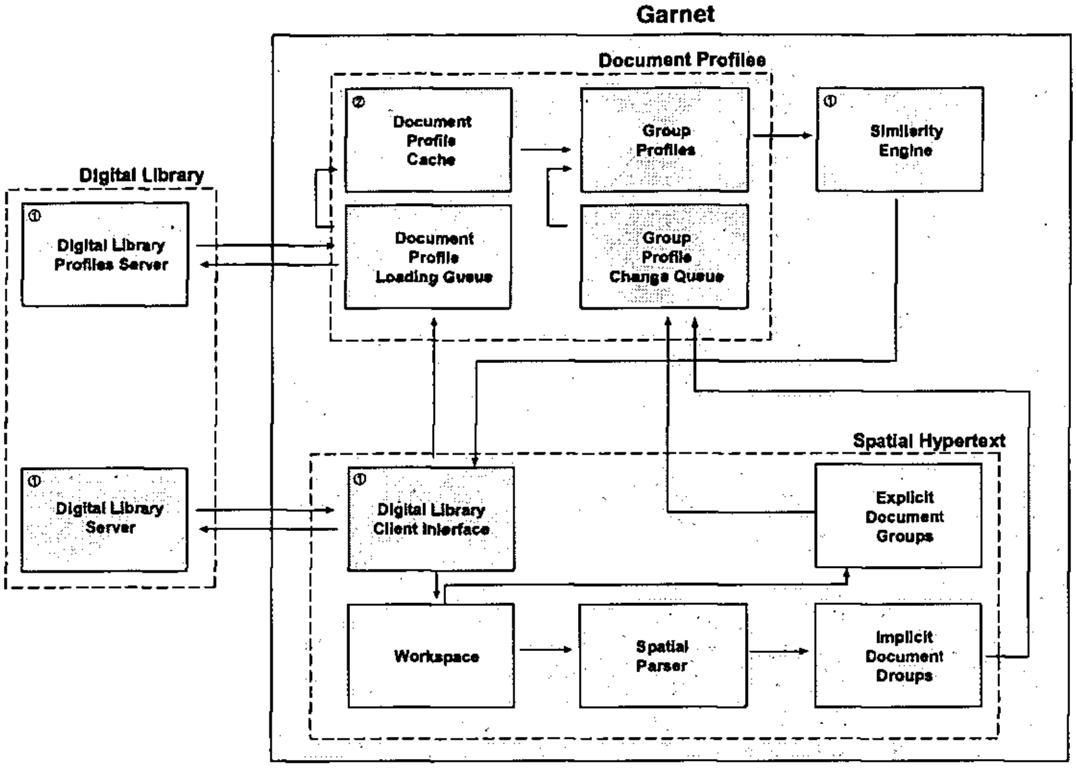


Figure 4.9: Garnet System Schematic

The function of these 'combining' components will be described in context throughout the remainder of this section. Those components that perform the traditional digital library and spatial hypertext systems will be described in less detail, as their function has already been described in the context of existing systems in Chapter 3. The diagram seen in Figure 4.9 will be used to illustrate the communication between the different components of Garnet throughout the rest of this section, within the context of action that will commonly be performed within the use of Garnet as described in §4.2 above.

However, before proceeding, it is worthwhile giving a brief description of the function of a number of these elements.

The digital library client interface in the Spatial Hypertext connects the digital library server to the hypertext. It sends requests to the server and stores the responses, passing information to the workspace regarding the items that should consequently appear there. In addition, it informs the profile repository of documents that it should be caching.

The profile repository holds the textual representation of documents and groups, and controls when each is created, loaded, recalculated or discarded. The group membership details are provided to it by the spatial hypertext system, aided by the spatial parser.

Finally, the similarity engine computes the similarity of individual documents to document groups, and also can provide a further processed representation of a document group for searching the collection. When invoked, it communicates with the digital library client interface and the workspace to ensure that all documents are considered, and that any matches that are found appear in an appropriate location and style within the visual workspace of Garnet.

4.3.2 Retrieving Documents

The user can retrieve documents from the digital library through either the search or browse facilities of any standard DL protocol. Each of these retrieval mechanisms will identify individual documents through their unique document identifier. A representation of each document is displayed in the spatial hypertext workspace. For instance, within a search result list window as seen in Figure 4.2 above [§4.2.2].

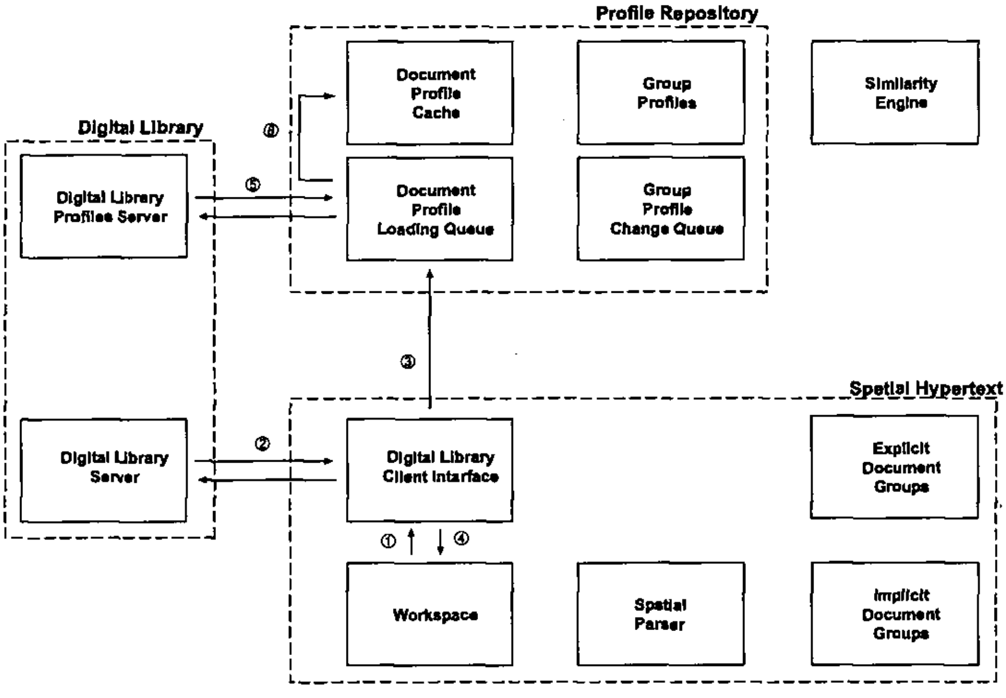


Figure 4.10: The retrieval of documents from the digital library

When a search for documents is initiated by the user in the workspace, the search instructions are sent to the digital library client – see Figure 4.10 ①. The DL client then communicates with the DL server to retrieve a list of relevant documents ②. For each document retrieved, its identifier is then given to the profile loading queue ③ whilst it is also represented in a new object sent to the visual workspace ④. Subsequently, the profile loading queue will use the document identifier to send a request to the document profile server ⑤, and the retrieved profile is added to the cache of document profiles maintained within Garnet ⑥. Thus, only the profiles of documents actually present in the Garnet workspace are ever loaded into the Garnet system. These document profiles are not immediately useful, but will be used in later interaction to perform document-to-document and document-to-group matching, as will be seen shortly.

Some further interactions occur which would only result in communication with the DL and spatial hypertext components of Garnet, not using the document profile information or the user's grouping of documents. For example, When the user double-clicks on a document label, the DL protocol is used to retrieve the document for display. A window is opened in which the document is displayed – the precise details of this will follow in the DL section of this chapter. The action is not influenced by its position on the workspace, nor by the contents of its profile.

4.3.3 Changes in the Workspace

When documents initially appear in the workspace, their profile is currently irrelevant – the search result lists and browsing structures are not products of the user's organisation, so 'scattering' is not relevant to them and the profiles of their documents are therefore not required. Changes to these structures, such as deleting documents from a search result list, will therefore not result in any changes to document groups used in matching.

On the other hand changes in user-created document groups, be those groups either implicit or explicit, are relevant to any modelling of the user's interest and to the 'scatter' facility. A change in the position of a document label may result in it belonging to another or a new group of documents. See figure 4.11 for a diagram of the subsequent communication between the components of Garnet. When a document is moved, the spatial parser must be run over the workspace region that the document has left and that to which it has moved ①, in order to establish any changes in document groups. So that the user experiences a fluid direct-manipulation interaction, the execution of the spatial parser is run on a queue basis, with individual regions of the workspace being re-parsed during pauses in the user's interaction.

Once the spatial parser has identified changes in the membership of document groups ② the textual model of those document groups must be correspondingly altered. The profile of each document will already have been loaded into the document profile cache. What remains is to alter the document groups' textual profiles by removing or adding documents as required. Each change in a document's allocation is placed in a queue, to be processed in turn. As with the spatial parser, the changes of group profiles are processed when the system is in an idle state. Changes to individual groups are agglomerated, and should a user rapidly move a document out of a group and then back into it, the group's profile will often not be recalculated at all in order to minimise computational costs.

Another event that can occur is the deletion of a document from a group. In this case, the document profile cache is informed of the removal of the document, and should it not appear elsewhere in the workspace, its profile will be deleted from the cache.

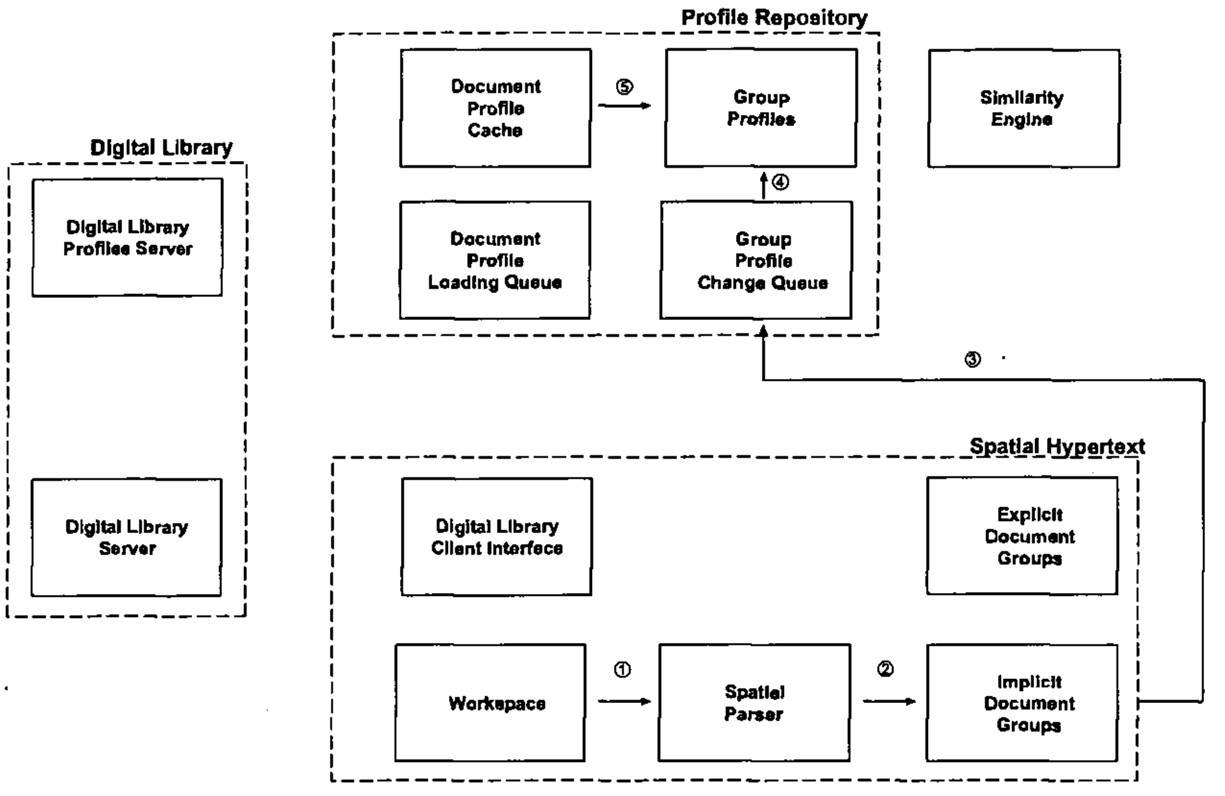


Figure 4.11: Movement of a document from one group to another in Garnet

Figure 4.11, above, demonstrates the flow of actions when a document is dragged from a search result list to join an existing document group in the workspace. The deletion of a document is merely a subset of the same consequences.

4.3.4 Processing Suggestions

When the user requests a 'scatter' to be done, the document group profiles described in §4.3.3 above are used to match documents against each group. All the individual document profiles for the search result set being scattered are matched against the profile of each group of documents. When a match is made between the profile of a document (e.g. one in a search list) and the profile of a document group in the workspace, a suggestion label is placed next to the document group (see §4.2.3). As the profile loading queue for the profile cache, and the queue of document group changes may not be empty when the scatter is requested, all the activity scheduled in each queue must be completed before the scatter is actually performed.

The sequence of actions is:

1. An extended list of documents in the search is acquired and placed in the digital library client.
2. The profile for each document in the extended search result list is obtained from the profile server and temporarily placed in the profiles cache.
3. The similarity engine compares each document profile in the extended search result list in turn against the group profiles in the profiles cache.

4. When a match is found, the similarity engine passes the details of the document and the matching document group to the spatial hypertext workspace.
5. Using information from each document group (originally generated by the spatial parser), the spatial hypertext workspace identifies where to place the representation of the matching document. Additional information on the document may be obtained from the digital library client.

The matching of documents in a "scatter" is performed solely within Garnet using the document profiles it has loaded. The digital library component system – providing search, browsing and document retrieval facilities – is not used at all. If the user, on the other hand, selects a group of documents and requests that Garnet finds similar documents, a different approach is taken. A group representation is generated from the individual documents selected and the keywords that would be used to represent the group (using Zamir *et al*'s clustering algorithm) are sent to the digital library system's search facility through the active digital library protocol. The digital library will return the usual list of matching documents, and these will be presented in a search result list in Garnet's workspace as with a normal search.

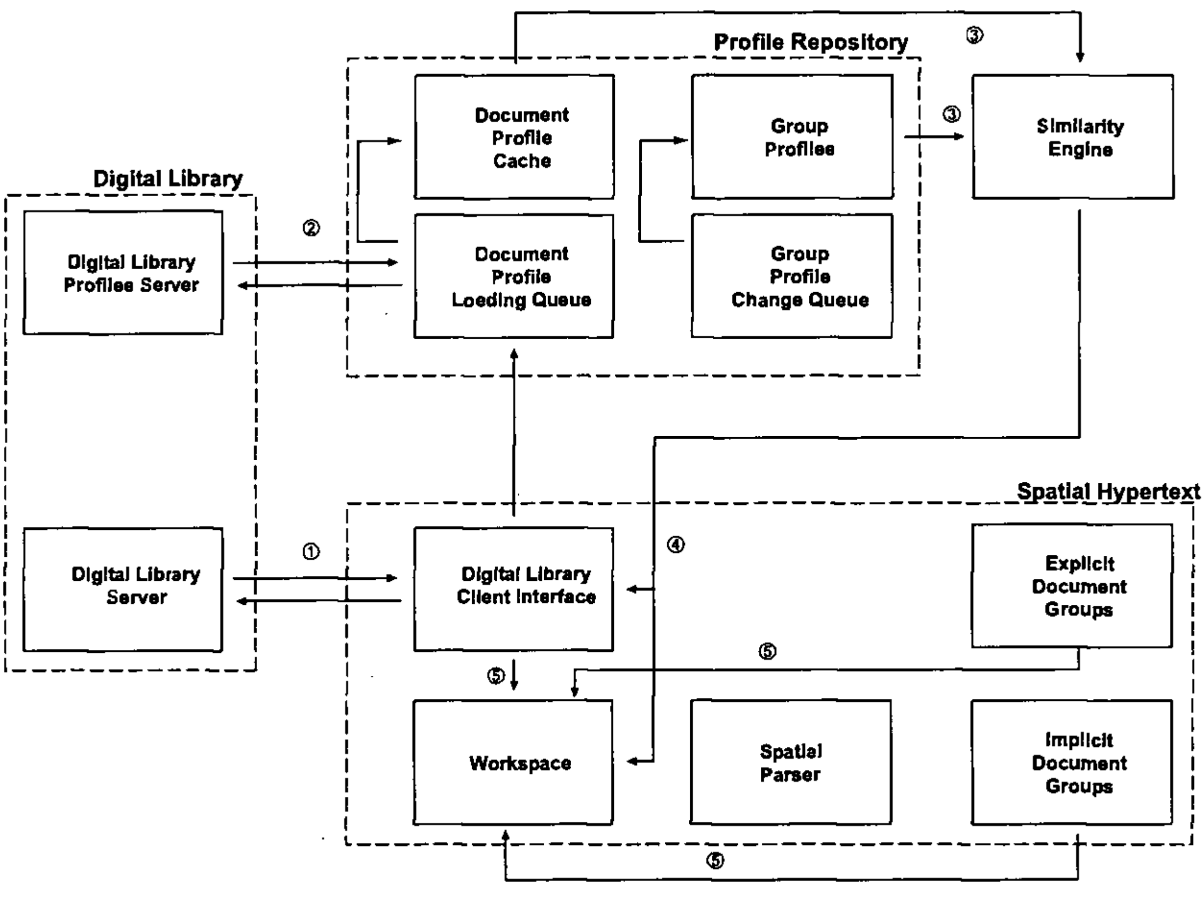


Figure 4.12: Garnet architecture and processing a 'scatter' request

4.3.5 Summary

The preceding section has described the manner in which the digital library and spatial hypertext components of Garnet are joined, particularly in terms of computing representations of groups of documents in the spatial hypertext workspace to match and retrieve documents in the digital library system. In the remaining sections of this chapter, the function of the digital library and spatial hypertext elements themselves will be described in greater detail.

4.4 Digital Library Implementation

The creation of a complete digital library system is a major undertaking. Fortunately, as discussed in Chapter 3, a number of digital library systems provide protocols to connect client systems to their indexes. In the case of my implemented system, Garnet, it was clearly wise to exploit such facilities, for pragmatic and theoretical reasons. Practically, using a DL protocol connection to an existing DL system dramatically reduces the time and cost of creating a working combined system. Theoretically, using a DL protocol generalises the systems over which Garnet can be used.

Furthermore, if a remote protocol connection is used to access a target digital library, in fact the system then may be used with any library using that protocol. Given that the original motivation for Garnet springs from the extended use of digital libraries, it is necessary for the system to support the access of multiple libraries, as any one library is unlikely to contain all the documents pertinent to a given piece of research.

However, if the selected DL system or the chosen protocol is atypical of DL systems as a whole, then the generalisation is in fact illusory. Therefore, the selection of a specific system and protocol is significant.

4.4.1 Digital Library Protocol

In Chapter 3, the various DL protocols were discussed. It was observed that the four major protocols in use: Z39.50, Dienst, SDLIP and Greenstone, can all be mapped to each other in terms of document retrieval, index searches and category browsing. Key differences occur in the areas of access control, authentication, state keeping and document versioning. In the case of the common operations, all the protocols can be successfully mapped to each other.

For purposes of generalisation, therefore, the areas of difference (authentication, versioning and state-fulness) must be avoided. However, for the common operations (search, browse, document retrieval), any of the four protocols can be used successfully without compromising generalisability.

The considerations of connecting to Web-based retrieval services were also addressed in Chapter 3. Again, it is worth re-capitulating those here. Most Web-based document retrieval services, such as Google™, have an XML-based access. The facilities of such systems tend not to include browsing (Google being one such example), but otherwise search and document retrieval are easily mapped to the facilities available in DL protocols.

The key difference is that the document retrieval and index facilities tend to be on different servers – e.g. the index on Google, the document on (say) the BBC web site.

Regarding differences between DL protocols and those used to access web search engines, the web-protocols tend to be 'agnostic' in matters of document access by the user – for instance, Google will index documents unavailable to end-users, leaving authentication etc. to the document server. Thus, documents that are in fact not accessible to a user will be reported to them, potentially resulting in later disappointment if the page cannot be retrieved.

So, in summary then, any DL protocol could be chosen, and Web-access protocols may be viewed as a restricted subset of the DL protocols. In the case of Garnet, the Greenstone DL protocol was chosen, mostly as a result of selecting the Greenstone DL system, which will be discussed next.

However, I also wished to be able to generate document representations in order to facilitate the matching of documents to each other, particularly for matching individual documents against document groups, as in the "scattering" action described above [see §4.2.3]. For performance reasons, generating these representations within Garnet at run-time, which would involve both accessing and downloading the available document information from the library server and post-processing that information to create some standardised representation, was not desirable. Adding additional processing work to Garnet would inevitably slow down the responsiveness experienced by the user, and thus I decided to access the profiles directly from the server.

To achieve this, an additional, supplementary protocol module to the standard Greenstone protocol was created, which facilitates the transfer of document profiles. As has already been noted [see 3.8.4], protocols generally have similar features and providing facilities as protocol modules, complementary to existing protocols, is advantageous. The supplementary protocol created can be readily used in principle with any DL protocol – in each protocol, a document has a unique identifier, and this is the only piece of data used to identify a document to the DL server within the document profile protocol module. The document identifier is simply given in the form originally obtained from the server through standard protocol actions such as search. Thus, the only feature used to tie the supplementary profile protocol to the standard protocols is in fact common to all the DL protocols that I have examined. This is illustrated below:

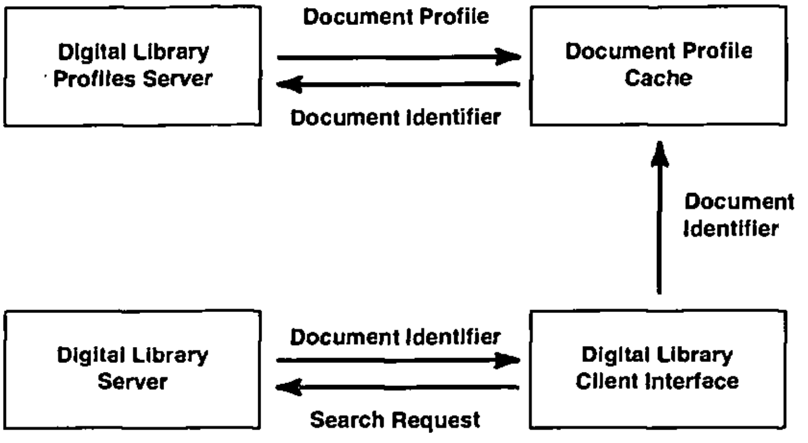


Figure 4.13: DL Protocol and Garnet Profile Extension

Where document profiles are not available via a digital library protocol connection, the profiles can be generated by Garnet independently, though at a cost in time. The creation of these profiles is performed as a background task in order to minimise the impact of the generation of the profiles upon the interactive response of Garnet to the user. However, this may result in some time delay, were the user to perform an action within Garnet that triggered document-to-document matching system behind "scattering", until such time as all the required profiles had been generated.

4.4.2 Digital Library System

Given that DL protocols may be mapped unto each other, it would appear that the DL system itself would not impact on an implementation of Garnet. However, certain issues are affected by the underlying system. For instance, Z39.50 and other DL protocols do not specify the default ordering of documents to be used over the results of a search. Other, anticipated impacts may occur, and so it is proper to expose the particular system used with Garnet.

Garnet has been implemented as a complement to the existing Greenstone digital library software, using the Greenstone DL protocol to connect with the digital library. This means that the digital library itself, the search facilities, index, documents etc. are (usually) served from a separate machine to the one on which Garnet is running.

Greenstone was chosen for a number of reasons. Firstly, I have extensive knowledge of its features, structures and capabilities having worked upon Greenstone for some four years. Secondly, the fact that Greenstone is an open-source package made extending its capabilities straightforward, especially given a sound knowledge of its internal operation, and improves the inspectability and reproducability of Garnet. Greenstone and its features are also extensively documented in academic papers [McNab *et al* 1998, Witten *et al* 2000], books [Witten & Bainbridge 2002] and reference and user manuals (http://www.greenstone.org/english/docs.html). Other open-source alternatives, such as Cheshire II [Larson 1996] are less well documented, or, as is the case with Cheshire II, at an incomplete and unstable stage of development. As just stated, Greenstone's protocol can also be mapped to the other major DL protocols, and its facilities are comparable to any contemporary research or commercial digital library system.

4.5 Spatial Hypertext Implementation

The provision of a Spatial Hypertext component was a much thornier issue. None of the mainstream spatial hypertexts such as VIKI/VKB [Marshall 1994, etc.], Pad++ [Bederson 1998] or WebSquirrel (http://www.eastgate.com/websquirrel/) were available to be developed from when the research commenced. Others, such as CAOS [Reinert 1999] were partially completed, and often focussed on 3d presentation, which is particularly noted by Shipman [2001a] as being a poorly understood environment. Being aware of the many pitfalls of 3d representation, such as occlusion [Li *et al* 1998], I particularly wanted to avoid confounding a partially understood and potentially problematic representation.

Therefore, there was no option but to create the substantial part of a spatial hypertext system from scratch. The various choices for representation etc. in spatial hypertext systems have already been discussed in Chapter 3, and are only summarised here. As the 2d-representations are better understood, and the representational framework of VIKI and its successors the most widely researched and investigated, the basic model of a 2-dimensional, unlabelled hypertext system was chosen.

Clearly, the construction of a comprehensive spatial hypertext system would represent a major undertaking in itself and a distraction from research considerations. Therefore, only the basic set of spatial hypertext facilities was implemented – advanced features such as editing histories [Shipman *et al* 2001] and non-linear zooming [Shipman et al 1999] were completely omitted, due to their presence in only a small number of spatial hypertext systems, and the complexity of their creation. Similarly, original features of the VIKI system that were dropped by later implementations – such as the use of non-rectangular labels – were also omitted from the Garnet implementation.

4.5.1 Organisational Facilities

In Chapter 3, the distinction between Implicit and Explicit organising tools in spatial hypertexts was given. Garnet offers both explicit and implicit organisation in its structuring tools. Explicit organising can be done using the "collections" system found in VIKI and its successors – the equivalent of folders in a filing system. Implicit organisation can be performed by positioning, colouring etc. As with the VKB system, the successor to VIKI, the shape of document representations was limited to rectangles, rather than the extensive range of circles, triangles, etc. which was present in the original VIKI system. This choice was primarily made to minimise the amount of programming required – the fitting of text to non-rectangular spaces being often non-trivial. The fact that VIKI's successor has made the same decision to some extent validates this decision.

There is a secondary benefit of this decision. As with the visual interfaces to digital libraries reviewed in Chapter 3 [§3.3], different types of object can be represented by different shapes. For example in browsing structures, nodes appear as a different shape to the usual document label, reducing the scope for ambiguity.

The selection of documents and document groups is another area of variation between spatial hypertext systems. Here, Garnet again offers only basic capabilities. A combination of the "Control" key and a mouse double-click on a document will select the informal group of near-neighbour documents identified by Garnet's spatial parser (for details of which see later), rather than the complex incremental selection of higher-level structures which is supported by VIKI or VKB [Marshall 1994]. Again, this decision is in part a consequence of the limited amount of programming time available, as the spatial parser of Garnet would have had to be substantially extended to support the identification of these more complex structures.

A final tool for supporting structuring activity is the User Label. In addition to the labels that represent documents in the digital library system, the user can add their own labels to the workspace. These labels do not represent documents, but are simply used to annotate individual documents or document groups. This in turn allows for the user to recover some context regarding how they organised the workspace at an earlier date or time, hopefully to palliate some of the difficulties observed in the extended use of a spatial hypertext [Marshall 2001].

4.5.2 Spatial Parser

Garnet's spatial parser is, like the rest of the spatial hypertext component, implemented from scratch. As was discussed in the previous chapter, spatial parsing is a little researched area and even the algorithms or heuristics used in existing systems are not available.

In the case of Garnet, I decided to implement the simplest spatial parser that would suffice to detect document groups. The parser is a simple proximity parser: documents within a certain distance of each other, edge-to-edge, are identified as being in the same group. The pattern of internal organisation or visual structure of the group – row, composite, etc. – is not presently identified. It is not presently understood [Shipman 2001a] what semantic significance particular forms such rows, columns or composites possess – if any. Thus, there was no certain value in extending the spatial parser in terms of improved quality in the textual classifier produced for the group or the computation performed with it.

In addition, Garnet notes the position of the bottom-rightmost document in the group; this is used as an 'anchor point' from which the placing of suggestions is determined – the suggestions being positioned as a 'stack' moving down and to the right:

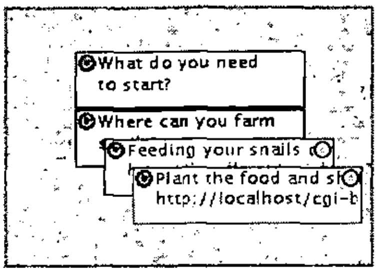


Figure 4.14: Suggestions placed next to their parent group

4.5.3 Summary

Thus, Garnet in fact represents a minimal Spatial Hypertext system in the mould of VIKI and VKB. It supports both implicit groupings and explicit hierarchical collections, and its spatial parser can detect simple implicit document groups that are identified on the basis of proximity. In addition to positional controls, there are basic controls for the colour and text font of document labels, and users are also free to add labels for purposes such as annotating their informal structures.

4.6 Connecting DL and Spatial Hypertext Components

The separate DL and spatial hypertext components from which Garnet has been constructed have now been described separately. However, extra elements have been introduced which exploit the connection of the two parts. This section will describe these additional elements which are unique to Garnet.

Starting from the spatial hypertext element, the spatial parser needed to be connected to a corresponding text representation and matching system that will be described first. Secondly, the connection of digital library facilities such as search and browse to the spatial hypertext is described.

4.6.1 Textual Representation – Individual Documents

For the purposes of similarity matching, each document is represented by a document profile – a list of the words in the document, each being listed with its rate of occurrence within the document. This representation is currently obtained directly from the DL through the Document Profile Protocol Module, described above in the DL section of this chapter. The use of the full text to represent the document was chosen in Chapter 3 as the most readily map-able document element across digital libraries. Where this is not obtained, a composite of the available document fragments is used instead.

In order to avoid the presence of overly commonplace words in the set of words common to a set of documents, a list of such 'stop-words' was removed from every document profile.

The document profiles are retrieved when a document is placed into Garnet's workspace or appears as an item in a search result list. As with a number of potentially resource-intensive aspects of Garnet's operation, this is done as a background task to minimise interference with the speed of response to the user's interaction.

4.6.2 Textual Representation – Document Groups

The visual identification of document groups has already been described above – in both this chapter and the previous one [§3.4, §4.5.2]. This section will describe the specific means by which the textual representations of each group are generated. The underlying method that I have chosen – Zamir *et al*'s – was discussed briefly in Chapter 3. The full details of the original algorithm are available in [Zamir *et al* 1997], but I will first discuss the pertinent specifics to this implementation, and the modifications made for the particular circumstances of the implementation in Garnet.

Zamir *et al*'s algorithm has been used with either phrases or keywords. The use of phrases in their studies slightly improved performance at the expense of a higher computational cost. I implemented the simpler keywords version as a starting point and, for reasons that will be discussed shortly, did not advance to using phrases.

One reason for selecting the algorithm was that it was reported as maintaining high-quality clustering even on small documents and document fragments. Some of the collections in Greenstone are built in such a way that a single 'real' document would be represented in several, related, collection documents, e.g. one electronic document per chapter or section. In addition, bibliographic collections with abstracts, a common digital library form of material, don't have full texts to evaluate, having only the document metadata and abstract to build upon. Whichever element is used to represent the document, one is working with short texts to represent it in bibliographic collections, so a matching algorithm of proven quality over short texts was important. In the absence of the full document text being present, Garnet will first recourse to the abstract, then to the available metadata.

In implementing and testing the system, however, another property of the clustering algorithm proved problematic in use. In the original implementation of Zamir *et al*'s algorithm, each cluster is represented by the words that all documents in the cluster share in common. However, testing revealed that in many cases this would result in a group representation of very few words indeed. One frequent problem was the presence of synonyms for the same concept, e.g. Human-Computer interaction being associated with both the acronyms 'CHI' (in North America) and 'HCI' (elsewhere). Similarly, spelling errors in the original text and the addition of a single document could dramatically affect the quality of the group representation.

One response could have been to move towards a document-weighting scheme as used by the classic TF/IDF of information retrieval. However, as discussed in Chapter 3, this information is rarely available for a given collection of documents, given the limitations of DL protocols, and cross-collection weighting is also problematic due to the lack of word weight information. The solution chosen was to slightly weaken the strict requirements of the original algorithm – words were weighted by the number of documents in which they occurred within the group, down to a cut-off:

$$D\sum_{w=1}^{w=W}(d(w)/D)^2$$

Figure 4.15: Log rule for information retrieval D = number of documents in a group, d(w) = number of documents in the group containing word w.

Similarly, phrases were found to be too rare to be used as document descriptors, and relying upon them would have resulted in few or no descriptors for most documents. Therefore, the keyphrase implementation of Zamir *et al*'s algorithm was not developed.

A second group representation was created for all document groups for the purposes of the evaluation described in the following chapters. This second form was simply the agglomeration of all the words in all the documents in the group, regardless of how many documents within the group each word occurred in.

Stop-words relate to another problem related to word-frequency. Stop-words are intended to avoid the use of very common terms – e.g. the word 'the' would classically appear in all or almost all documents written in English. Including such words in every cluster would result in spurious matches occurring. As described above, such words are removed from document profiles. An alternative, language-independent and corpus-sensitive approach would be to use word-frequency (across all the documents in a corpus) to determine which words would be deleted from the document profiles. However, this would rely upon the representation being generated on the digital library server rather than by Garnet itself, which would limit the flexibility of Garnet's design, or the frequency information of words in a DL collection being available to Garnet, again limiting Garnet's scope of use. Therefore, this automatic approach was not implemented.

Finally, document group representations will change as documents are added or removed from a group, new groups are created and old ones merged or destroyed. This can result in significant amounts of computation, and Garnet has been designed to avoid this interfering with the flow of the user's interaction. When a change occurs, it is placed in a queue, and the resulting alterations to document profiles are performed as a background task.

4.6.3 Text Matching of Groups

Once the document group representation was generated, two approaches to text-matching were implemented. Firstly, the list of words common across the group's documents (the main group representation) could be used as the input to the existing search facility of the DL protocol. Such an approach would permit a user to select a group and search for similar documents to the group within a digital library, receiving a traditional search result list in return.

Secondly, an individual document could be matched either against the group's common words, using Zamir *et al*'s measure [1997], or against the more traditional group profile of all words, using the matching technique of Scatter/Gather [Cutting *et al* 1992]. In the current implementation of Garnet, I use Zamir *et al*'s algorithm to determine similarity.

If a document is matched against a group of documents, the group anchor point identified by the spatial parser (see §4.5.2 above) is used to position a suggestion representation of the document near to the group.

4.7 DL Facilities – Search and Browse

As discussed in Chapter 3 and above in §4.4.1, Garnet uses the Greenstone DL protocol, a typical protocol which can be mapped onto the other standard protocols, and onto which the common features of the other protocols can be mapped. Searching and Browsing are both found in many DL protocols and are both supported by Garnet.

Searching is controlled by a pop-up dialogue that can be accessed either through Garnet's menu system or via a keystroke. This one of the approaches found in the integration of information repository features into spatial hypertexts seen in §3.3. This particular representation was chosen as search tool controls are not themselves documents.

A number of information sources can be searched, and for digital library sources, these are detected through the 'ListCollections' action of the Greenstone protocol. In addition to DL collections, web searches can be done via an Internet search tool. At present, a connection to the web search engine which I implemented for the WebTwig [Jones, M., *et al* 2000] and WebTree [Buchanan *et al* 1999] browsers is used, but an XML connection to the Google™ search engine or similar could be straightforwardly added. The list of sources thus obtained is presented for the user to select from in the search dialogue:

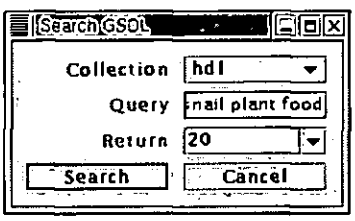


Figure 4.15: Search dialogue form

Once Garnet has made the connection to the chosen information source, its index is queried through the 'Filter' action of Greenstone (which provides both search and browsing facilities) or the 'Search' action of the Web search engine as is appropriate. This will result in the return of a list of the corresponding documents. This is parsed, and Garnet generates a window containing the documents in a vertical column, typically ordered by relevance:

Figure 4.16: Search result list in Garnet's workspace

In addition, the textual profiles of the documents need to be obtained for the creation of document group representations and the matching of documents to groups. Garnet identifies the documents which it does not already have the profiles for, and queues these for loading. This is done as a background task so as not to interfere with the response time of the system to the user's interaction. The profiles are loaded through the Garnet DL protocol module when available – otherwise, they are generated by Garnet itself.

For browsing, Garnet presents a folder icon on the main workspace that can be double-clicked to open the available browse-able DL collections. These collections are discovered using the 'ListCollections' action of the DL protocol, as described above for searching. The availability of Browsing is confirmed using the Greenstone Protocol's 'Filter' action, which with the correct parameters will list the available browsing structures for a given collection. Collections with no browsing structures or classifications do not appear in the browse-able collection list. Currently, there is no direct access to a Web directory service such as Yahoo! This is a different approach to representing the DL's facilities to that used in the case of search. In the case of browsing, the example of KidPad [Druin *et al* 2001] and Navique [Rauch and Furnas 1998] has been followed, as the structure is directly navigable within a hypertext rather than relying upon additional input from the user as with search. Unlike the dynamic behaviour of search result lists, from which documents may be deleted, the permanent nature of browsing structures, set in the architecture of the library design itself, do not permit the user to delete documents or nodes from them.

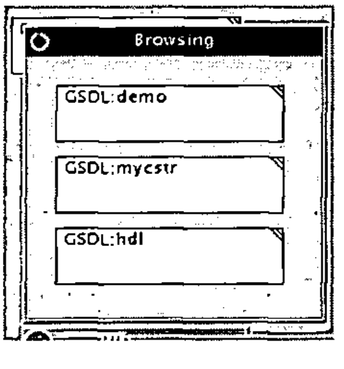


Figure 4.17: Browsing node

If a user double-clicks on a collection name, the list of its browsing classifications is presented in a manner similar to the search result list, a single column, with sub-classifications listed first, followed in alphabetic order by documents in the classification. This structure can be subsequently "drilled down" as the user requires. Each opening of a classification results in a call to the 'Filter' action to acquire a list of sub-classifications and any documents that are immediate children of the category.

4.8 DL Facilities: Reading Documents

Once a user has located a document through searching or browsing, they will wish to read the document. Garnet could access the original document text through the 'GetDocument' action of the Greenstone protocol, but instead initialises reading with a browser of the main DL library through a web browser. This was done to minimise development time by providing a complete reading interface for DL documents, and to provide as familiar as possible a reading environment to the user. As documents are provided to any client, such as Garnet, in HTML, producing a reading interface would have meant implementing a complete HTML renderer, and though a standard Java API exists to provide this facility, the quality of the resulting presentation is very poor. In comparison, with Greenstone, it is straightforward to compute the URL for a document given the information on it returned by searching or browsing, so presenting the document through a browser can be achieved easily.

The approach I have used matches that of Marshall and Shipman in the VIKI and VKB systems discussed in Chapter 3.

4.9 Implementation

The spatial hypertext component of Garnet was implemented in Java 1.3 and tested on Linux, Mac OS X and Windows systems. The Greenstone DL protocol server runs on the same platforms, but for the purposes of development was run on Linux systems alone.

4.10 Discussion

Garnet provides a workspace that encompasses both information structuring, through its spatial hypertext facilities, and information seeking, through its digital library facilities. For information seeking, Garnet uses a standard digital library protocol to provide searching and browsing activities within a spatial hypertext environment. A supplementary DL protocol module has been implemented and used in parallel with the Greenstone DL protocol to obtain document representations that are used for matching documents and document groups.

Merely by its implementation, Garnet can be a contribution to research. The following section will discuss this contribution, viewing Garnet from two different research perspectives: Spatial Hypertext and Digital Libraries. In the case of Spatial Hypertext research, the contemporary research agenda is relatively well defined, as the community is small and its views consistent. Frank Shipman [2001a] has presented a list of seven directions for Spatial Hypertext research, and Garnet as implemented relates strongly to a number of the issues given by Shipman. On the other hand, Digital Library research is a diverse field, and is strongly influenced by the expertise of the individual researcher, be they from Information Science, Information Retrieval, Human-Computer interaction or another discipline.

4.10.1 Spatial Hypertext

From the perspective of Spatial Hypertext Garnet represents a simple implemented system closely related to VIKI and VKB. Garnet supports both implicit and explicit structuring, and it can detect implicit structures created by the user through a simple proximity-based Spatial Parser.

However, unlike previous spatial hypertext systems, Garnet emphasises the role of Spatial Hypertext in a wider information environment and is connected directly to a source of documents – a digital library. Frank Shipman's list of seven directions in spatial hypertext research [2001a] included the item "Integrating Spatial Hypertext into the Information Environment" – which included both reflecting the information environment within the spatial hypertext, and exploiting information in the spatial hypertext to support other work. Garnet not only facilitates a broad range of information seeking and information structuring activity, but it can also use the spatial hypertext structures created by the user to filter documents in the digital library, an experimental means of using the user's organisation to support their later work. Compared to the simple ability to link to web pages in VIKI [Shipman *et al* 1997], the presence of search and browsing facilities in Garnet is much richer. Shipman also states that few options of how to represent other parts of the information environment within a spatial hypertext have been explored, and Garnet also explores some of the identifiable options for presentation (e.g. browsing structures, search engine, etc.). Therefore, Garnet addresses both aspects of the integration question raised by Shipman.

A second issue raised by Shipman was "Computation In and Over Hypertext" – how computational processes could be presented within a spatial hypertext environment ("Computation in hypertext") and how computational work done across the structures in the hypertext could be usefully exploited ("computation over hypertext").

The latter issue has already been in part addressed by spatial parsing in systems such as VIKI, where the hypertext has identified implicit groups and allowed these groups to be readily selected by the user. However, this particular feature falls somewhat short of the original concept of "computation over hypertext" [Halasz 1988] which Shipman is expanding upon: the selection of documents is more an activity required by the hypertext itself, not so much a new capability for processing documents.

In comparison to merely selecting documents, Garnet's computation over hypertext creates a model of the user's organisation of documents and uses this to generate corresponding textual models of the identified document groups. These textual models can then be used filter documents within the digital library. Previous spatial hypertext systems have not used computation over their workspace in a similar manner. In fact the principle of how to do computation over spatial hypertext has been little researched. In providing computation over spatial hypertext, Garnet gives an example of what may be achieved in principle, and permits the benefits of computation over spatial hypertext to be better explored.

The issue of "computation in" a hypertext is represented in Garnet by the presence of the search facility of the digital library. Thus Garnet strongly addresses the issue of computation over spatial hypertexts, and also addresses the issue of computation within spatial hypertext.

Shipman also raises the question of "Synchronous collaboration", which he sees as addressing a variety of technical and representational elements. In Garnet, both the user and the system can place content into the workspace, and thus there are two active 'agents' working in the same workspace. Some of the difficulties that Shipman anticipates include time synchronisation, ownership and control of documents, and visibility of action.

Taking these expected problems in turn, time synchronisation is in many ways addressed by having Garnet only engage in placing when the user triggers its activity (e.g. performs a "scatter") and it only returns control to the user when its task is completed. Therefore, a simple turn-taking approach is used to resolve this issue. Garnet may, however, have to complete some computational work (e.g. updating the groups identified by its spatial parser) before it is able to do its work. Also, as the user and the textual processing element of Garnet are in many ways competing for processor time, a lot of hidden synchronisation is taking place – e.g. building group profiles – to maintain the interactive response to the user. In the case of ownership and control, Garnet permits the user to exercise some control of over certain objects not created by them (e.g. items may be deleted from a search results list) whilst debarring other actions (e.g. moving documents freely around a results lists). Garnet also gives relatively subtle cues as to the ownership or role of a collection – e.g. a search results list is titled "Search for 'snail'" for example – rather than attempting to use bold visual cues. Document behaviours are controlled by which collection they are in (i.e. in a search list, browsing structure or a user's own collection) and are not represented by the appearance of the document itself. This approach has been taken to present documents in as neutral a mode as possible to the user, allowing emphasis to be focused upon the user's task rather than the system's state. Whether this approach is effective is one of the issues that an evaluation may address. Finally, visibility of action is an issue that becomes particularly acute in larger hypertexts. An example issue that could emerge would be the ease with which a user could view all the suggestions placed in a large hypertext – many of which could well be outside the visible workspace area when a "scatter" was performed. At present, this issue has not been addressed in Garnet, and the impact of a lack of this can also be observed during any evaluation. Thus, Garnet has to address a number of aspects of the "synchronous collaboration" issues identified by Shipman, and explores some, though not all, of these as implemented.

Garnet also has relevance to other parts of Shipman's directions, but the three given above are the most directly applicable. Given the relevance of Garnet to synchronous collaboration, computation over hypertext and integration into the information environment, its implementation provides a useful platform through which several contemporary research questions in Spatial Hypertext can be addressed.

4.10.2 Digital Libraries

Garnet's contribution to the well-defined field of Spatial Hypertext is readily identified. However, it can also be viewed from the perspective of Digital Library research, and this section will discuss its contribution there. Two areas of Digital Library research are particularly of relevance to Garnet: information seeking models and visual interfaces.

Firstly, it was seen in Chapters 2 and 3 that digital libraries provide little support for information structuring; therefore, just as providing information seeking support within a spatial hypertext is an issue, so is providing information structuring support in a digital library. The most immediate contribution Garnet makes to digital libraries is allowing the exploration of the benefits and consequences of providing information seeking and structuring support in one system. Its implementation also demonstrates that a combined system can be achieved.

The relevance of spatial hypertext to Digital Library research is further validated by the influence of spatial hypertext systems and research upon the visual interfaces to digital libraries seen in [§3.3]. However, despite that influence, the information structuring facilities in visual digital library interfaces prove weak when compared to those of traditional spatial hypertexts.

The context of visual interfaces of digital libraries is clearly an area to which Garnet is relevant. Issues such as how the user's organisation of their workspace may prove useful for supporting retrieval activities are as relevant here as to Spatial Hypertext. Similarly, the acceptability of Garnet as a particular system, and the benefits and problems perceived by users are worth identifying and comparing against the existing visual interfaces. The relevance of Garnet to visual interfaces to Digital Libraries is further underlined by the fact that Garnet was presented at the First International Workshop on Visual Interfaces for Digital Libraries [Buchanan *et al* 2001].

Another field of research that strongly influences Digital Libraries is Information Retrieval. Garnet uses the visual groupings of documents identified by its Spatial Parser for information retrieval purposes. For each group of documents, a textual model is built from the text of the constituent documents. This is then used by the "scattering" facility [§4.2] to match the documents found by a search to the informal organisation made by the user.

IR has studied the use of both automatically generated structures such as phrase hierarchies and document clusters and of pre-existing structures such as classification systems to organise sets of documents, particularly of search results sets. However, the use of the (sometimes informal) organisations created by end-users has not been studied. Garnet thus provides a platform for assessing the merits of this approach. As it also has access to the organisational facilities of a digital library, the use of such formal structures could be compared against that of the informal organisation created by end users within an IR context.

Another aspect that IR can identify is the degree to which the user's explicit and implicit structures compare in terms of internal consistency when compared against either pre-existing human-generated classifications prepared by information scientists and librarians or the automatically generated organisations of clustering systems. These different approaches can also be assessed from the perspective of Human-Computer Interaction, another field of research that is found within Digital Libraries.

Finally, Digital Library research has been heavily influenced by the work of Information Scientists, and the study of the information seeking process. As introduced in Chapter 2 §2.3.4, some information seeking models include elements of information structuring, and those performing information seeking in physical environments have been observed to use numerous information structuring tasks in support of their information work. Garnet presents an opportunity to place a digital environment containing support for both tasks before users, and to study and observe any interplay between the two tasks. This would subsequently permit a comparison of physical and digital information seeking behaviours in regard to the use of information structuring as a supporting task.

4.11 Conclusion

This chapter has introduced Garnet, a combined information structuring and information seeking tool comprising a spatial hypertext with an embedded connection to the Greenstone Digital Library. Garnet's novelty as such a combined system has been discussed, and I have identified the various aspects of the system which are of interest as a matter of research. The research questions that arise out of Garnet from four different and overlapping fields – Spatial Hypertext, Digital Libraries, Information Retrieval and Human-Computer Interaction – have been identified and briefly discussed.

From Spatial Hypertext, Garnet's connection to a digital library raises a number of questions concerning the integration of a spatial hypertext system to a wider information environment, including issues of representation and interaction. The role of information structuring in support of information seeking in a digital environment is of interest to Digital Library researchers, and the potential role of information structuring artefacts in information seeking to those studying Information Retrieval. Across all these areas, the interactive behaviour of Garnet as a system and the workflows that it supports are of interest from the perspective of Human-Computer Interaction.

The next chapter will study how Garnet may be evaluated from these different perspectives and lead to the design of a user-centred study upon Garnet which allows some of the research questions pertinent to it to be addressed.

Chapter 5: Evaluation Methodology

In designing an experiment to evaluate Garnet, it is important to review the methodologies used by previous researchers. There are four clear areas from which examples can be drawn: Spatial Hypertext, Digital Libraries, Information Retrieval and Human-Computer Interaction. This chapter will briefly review each of these areas in turn, before summarising the lessons drawn, defining the basis upon which Garnet should be evaluated, and reporting the design of the experiment actually performed.

Before commencing, however, it is worthwhile stating which are the elements of Garnet that I wish to evaluate.

Firstly, Garnet provides a new style of interface to digital libraries, and correspondingly embeds an information system within a spatial hypertext. The most basic questions arise here as to the acceptability of the overall system when compared to its constituent parts (Digital Library, Spatial Hypertext), and whether users perceive a benefit in the combination of information seeking and information structuring tools? At smaller scales more detailed questions of usability arise regarding each of the interface components – search result lists, document labels, etc.

Secondly, having combined the two components, how will users behave when they use the integrated system? The observations of Kidd [1994], and O'Day and Jeffries [1993] in studying physical information processes suggest that the two processes of information seeking and information structuring will be finely interwoven. Will this in fact occur? Will the users behave as those studied by Marshall and Shipman in their study of spatial hypertexts in use, creating a mixture of informal and formal structures?

Thirdly, it is known that users can utilise the organisation capabilities of spatial hypertexts: however, are users' groupings of documents sufficiently distinct and consistent for an information retrieval system to be expected to achieve credible matches? Finally, will the matching that Garnet performs between the subject's organisation of their chosen documents and other documents in the library prove acceptable or convincing to its users?

This chapter will now proceed to study the separate research domains relevant to Garnet to ascertain the evaluation procedures and techniques used in each.

5.1 Spatial Hypertext Methodologies

Given the relatively small size of the Spatial Hypertext research community, there are not many user-centred evaluations of spatial hypertext systems. Furthermore, some established systems have received partial or no such evaluation. For instance, the Pad++ 'zoom-able' spatial hypertext by Bederson *et al* has received little evaluation as a whole system, though components such as its zooming technology have been evaluated e.g. [Combs and Bederson 2000].

As will have been clear from Chapters 2 and 3, Cathy Marshall and Frank Shipman have done much of the key work in Spatial Hypertext, and they have produced a series of papers that include a variety of user evaluations. These studies are usually qualitative and observational in nature, capturing the artefacts of the users' activity, supplemented by feedback through interviews and questionnaires. The evaluation is essentially textual and qualitative, as Marshall is primarily interested in differences in users' work patterns and the adoption and effect of features rather than in easily quantifiable items (e.g. the number of documents in the hypertext). This methodological bias and emphasis is also clear in her recent work on digital libraries and reading and annotation tools [Marshall 1999, 2001].

For example, in [Marshall *et al* 1997], Marshall and Shipman had three different working environments – two electronic and one physical – they wished to compare, in order to identify common and differing behaviours. Each of the three environments was used by five subjects. Each subject was permitted 45 minutes to perform a given task – the same task for all subjects. Those subjects using a computer-based system were given a ten to fifteen minute supervised tutorial in using VIKI before embarking on their main task. After their task was performed, the spatial arrangement produced by each subject was recorded (by saving the hypertext in the case of the computer systems, and by photographs in the case of the manual system). Every subject completed a post-experimental questionnaire that obtained background information on the subjects, and elicited the degree of confidence they had over their success in the experimental task. The results were evaluated qualitatively, describing and comparing the product of the users' organisational work, and the comments made by the subjects when reflecting upon their activity.

In the case of their initial work on spatial hypertexts with Aquanet [Marshall *et al*1991], the early implementation of the VIKI [Marshall *et al* 1994], VITE [Hsieh *et al* 2000] and VKB [Shipman *et al* 2001] systems, the predominant goal has been to articulate the functionality of each system and relate its operation to particular work flows and tasks. In each case the system been used by the development group and their peers, and the experiences gleaned from this use had been informally studied to determine the development of the system.

The most recent work [Shipman *et al* 2002] represents a different approach – relying more upon questionnaire feedback, and engaging in a limited comparison. In this case, two separate studies were undertaken. As Shipman was evaluating the provision of a set of suggestion facilities in a spatial hypertext, which bears some similarity to the "scatter" facility of Garnet, this particular study is worthy of particular attention.

In the first study, seven users participated in an undisclosed task or set of tasks and were asked to rate several suggestions made by VKB through an 8-point questionnaire. This rating was compared against the subject's ratings of the Microsoft Office Assistant on the same questions. The ratings of both systems were compared statistically, and p scores were reported for the result of the comparison.

In the second study four users were asked to use the system and to comment upon the differing usefulness of different suggestion types whilst engaged in a brief 15 minute task. In addition to the feedback gleaned from the users during this initial task, each subject was then shown a scripted scenario of a third party using a spatial hypertext. Each participant was then asked to provide a suggestion of a particular type at given points in the interaction. The user's suggestion was then compared to what the system actually suggested at that point in the scripted interaction.

In this last experiment, therefore, limited numerical, quantitative analysis was used, quite unlike the previous experiments. However, the comparison made, with Microsoft Office's Assistant, does not allow one to draw any conclusions relevant to Garnet, and given Shipman's objective of judging the intrusiveness of his suggestions, one is unable to draw any conclusions about the underlying usefulness of suggestions generally.

Marshall and Shipman's recording of users' organisation of space has already been mentioned. In the case of Garnet, the underlying quality of users' organisations of information space is more critical and sensitive than was the case with Marshall and Shipman's systems, as it results in a textual representation. Thus in addition to visual layout I am interested in the semantic content of the documents in each group. However, the means of capture of the positioning data remains of interest.

As has been mentioned, in the case of Pad++, another spatial hypertext system, no overall system evaluation has been performed. However, evaluations of specific features have been done, comparing them against existing tools - e.g. between different history systems [Hightower 1998], navigation performance [Bederson 1998]. In these cases, the method of experimentation has been the traditional, quantitative approach to feature-comparison, with two alternative approaches to supporting the task treated as different conditions in the experiment. Such approaches will be described and discussed in further detail in the human-computer interaction section later in this chapter.

As has been seen, the history of user study evaluations of spatial hypertexts is varied in practice, and few experiments can be readily compared to others. Unless specific features are being compared, qualitative approaches predominate, which operate over observational and elicited data, rather than information achieved through measurement. This is particularly common where there is an endeavour to elicit the patterns of use of a new system and the activity of the user within it. Marshall has particularly relied on studying the artefacts of tasks, the hypertexts produced, and relating the observable visual patterns to previous organisational literature such as that of Kidd [1994] or O'Day and Jeffries [1993], and the comments made by users reflecting on their use of the system.

Experiments have used small numbers of subjects – typically between five and fifteen participants. The use of observation and elicitation techniques, which tend to derive substantial quantities of data per subject, may be a contributory factor.

All these evaluations have also been laboratory, task-based, studies rather than ethnographic or naturalistic observations commonly found in the underlying Information Seeking research of Kidd, etc. This applies regardless of whether the subsequent analysis was either quantitative or qualitative.

5.1.1 Summary

In the case of Garnet, I am extending a spatial hypertext to provide support for the information seeking activity that both precedes and follows on from information structuring work. What would be of interest would be to gain some insight into the flow of work that occurs between the two activities, and how users interact with each. I am also interested in discerning the users' appreciation of the combined system, and to identify immediate difficulties experienced in interactions. These can be obtained through the observation and elicitation patterns of study used by Marshall.

The capture of user patterns of layout and behaviour can be obtained through recording the workspace, including its changes over time. Again, Marshall's practice is appropriate. However, how one would judge the quality of the structuring performed by the users is not an issue which has been investigated by Spatial Hypertext researchers, and thus other fields of research must be examined to discern how this may be done.

I will now address Digital Library and HCI research in turn before turning to Information Retrieval research and its evaluation methods.

5.2 Digital Library Methodologies

Digital Library research has been approached by academics from a number of disciplines – e.g. Information Retrieval, Human-Computer Interaction, Information Science – and unlike the specialised field of Spatial Hypertext, there is a much larger corpus of information and research. As each of these fields emphasises different criteria in evaluation, practice varies correspondingly. Given the graphical interface that Garnet has adopted, I will limit the examination of DL evaluation techniques to that research which has focussed upon visual interfaces to digital libraries.

In Chapter 2, I introduced a number of these systems (DLite, NaviQue, SketchTrieve and Daffodil), and the evaluation of these systems is clearly of interest. Each of these systems has been developed as a proof-of-concept prototype, and in each case only a small amount of evaluative work has been published.

George Furnas' NaviQue system [Furnas & Rauch 1998] provides a useful starting point. The evaluation of NaviQue reported in that paper took the form of pilot experiments primarily aimed at reducing the number of immediate problems in the interface, such as performance issues and control problems. These exposed some difficulties, such as the lack of take up by users of certain features or combinations of features. There is little information given of the precise methods used for either information capture or analysis, which leaves the impression that the experiments were informal and small-scale. What is reported, though, is that the development and experiments were interwoven and incremental – each trial leading to further changes to the interface and system.

Cousins, in his study of DLite [1997], took a more explicitly structured approach to evaluation. His pilot study was of six users engaged in a bibliography creation task, which took approximately 30 minutes, and like the NaviQue study this uncovered a number of user interface blemishes. Again, little of the methodology is described, but the reported material suggests that an observational, laboratory experiment was conducted. A follow-up study was promised, but this never resulted in a publication. Like Marshall's evaluations of the VIKI spatial hypertext, Cousins identified common patterns in the work performed by the subjects, in his case the patterns being of sequences of actions taken rather than of the visual layout of artefacts generated.

Hendry and Harper's [1997] evaluation of SketchTrieve was another small study, including five user subjects who were each engaged in a series of five different sets of tasks. Elements of both the think-aloud protocol [Ericsson & Simon 1984] and interview techniques were used to elicit information on the motivation of the user's choices or their use of particular features. The time taken to perform tasks was measured – unusually, these were compared with similar activity in a physical, printed, dictionary rather than with a comparable electronic system. They also observed spatial arrangements of the artefacts of the tasks undertaken, as with Cousins' and Marshall's work, and compared the visual layout of objects used by their subjects to the visual patterns observed by Marshall and Shipman [1993].

The Daffodil system [Fuhr 2002] was evaluated using two separate studies: a heuristic evaluation involving eight subjects and a satisfaction questionnaire from twenty subjects. The former evaluation was used to identify problematic properties of the interface, the latter to elicit problems arising out of general library skill deficits. The results are generally presented in a simple quantitative manner (only mean scores are given), together with illustrative cases.

VQuery [Jones 1999] is a graphical Boolean query interface to support access to digital library documents through the graphical manipulation of search terms. Unlike the other systems described in this section, VQuery was suitable for comparison with an existing and (relatively) well-understood activity in the digital library. The study was performed in a laboratory-based study using eighteen participants. The evaluation of the Venn-diagram centred approach of VQuery was two-fold. Firstly, users were studied to determine the general ability of subjects to draw appropriate Venn diagrams of given Boolean formulations, or given a Venn diagram to create the corresponding Boolean form. Secondly, subjects were tested in a like-for-like comparison between their performance of

Boolean queries in a traditional interface and the same task through the visual interface of VQuery.

The separation of the identification and quantification of the underlying skills required for VQuery – mapping Venn diagrams to Boolean logic and vice-versa, is an unusual feature. Unlike the purely qualitative evaluations of Marshall and Shipman of spatial organising behaviours, Jones takes a more quantitative approach, noting the number of tasks successfully done in each mode (i.e. Boolean to Venn and vice-versa), and providing statistical information such as confidence values.

5.2.1 Summary

As was the case with the Spatial Hypertext evaluations, many of the studies examined above are essentially qualitative, with small numbers of subjects who were observed in a laboratory environment. Another similarity to Spatial Hypertext is that a number of the studies of novel systems rely upon some evaluation of the artefacts left by subjects' interaction with the system.

Analytical methods often seem partly or wholly informal, and there is not a mass of consensus. However, the level of methodological detail reported is often not sufficient to gain certain comprehension of either data capture or analytical techniques.

The studies, in so far as their methodologies can be discerned from the available literature, are clearly laboratory-, task-based experiments rather than studies in everyday use. It is generally unclear how much and in which manner the experimenters interacted with their subjects – though interviews would seem to have been used in most cases, and two experimenters use limited questionnaires.

In 2002 I co-chaired a workshop on Digital Library Usability [Blandford and Buchanan 2002] at the Joint Conference on Digital Libraries. As part of running the workshop, it was clear that the problems of identifying appropriate criteria and methodologies for digital libraries generally are not yet clearly tractable, as evaluation methods were similarly focussed upon simple, incremental approaches, and substantial, quantitative studies were seen as premature, as the knowledge of which questions or measures were appropriate was seen as unclear.

As with Spatial Hypertext, only where a limited scope, direct comparison can be made – as with VQuery – are quantitative methods deployed. With the first implementations of certain types of tool, and especially when the number of potential adopters is perceived as small, the evaluations are usually qualitative and conducted in order to identify the influences on work practices and patterns.

Thus far, a certain consensus of practice seems to have emerged across both Spatial Hypertext and Digital Library evaluations. However, Spatial Hypertext and Digital Library researchers are not necessarily expert in evaluating the interactive properties of systems. For this, one must turn to Human Computer Interaction researchers, and study the means by which they evaluate such systems.

5.3 HCI Methodologies

Human-Computer Interaction is a wide-ranging field, and is impossible to address adequately as a whole in the scope of this thesis. In this section, the evaluation of novel interfaces to information repositories generally, and to digital libraries in particular, will be the primary focus. The studies reported in the previous sections have been conducted by experts in those domains. Evaluations of DL systems that have been performed by HCI experts will be reported in this section.

Unlike researchers who specialise in, e.g. information seeking, HCI practitioners are much more versed in appropriate evaluation techniques for the acceptability and usability of computer systems. Therefore, their evaluation methods bear closer examination.

One example of such an evaluation is Shneiderman *et al*'s paper [Shneiderman 1999] on the HierAxes browser presented at the ACM DL conference in 1999. HierAxes provides a means of interactively querying and browsing a set of documents with plentiful metadata through a rich, visual interface. To evaluate the system, Shneiderman *et al* performed two sets of studies. The first set was performed with 8 subjects from an academic Computer Science background, and resulted in a number of changes to the system. The second study was undertaken with 24 subjects from a more diverse background (library science and computer science totalling three quarters of the subjects). Subjects were given a simple two-minute introduction to the system before undertaking the main experiment. The results of task completion, user satisfaction, and error rates in mistaken actions were briefly reported in the paper. Results are reported only as simple percentages, e.g. of the number of users. The system was not compared to traditional search or browsing interfaces. The means of collecting information on the users, e.g. their degree of satisfaction, or how tasks were determined as being successful or unsuccessful, are not reported.

As a comparison, Chen and Dumais' paper on organising search result lists [Chen and Dumais 2000] entered into a direct comparison of a novel organisation of search results and directly compared it against the traditional ranked list. For their user study, eighteen subjects were recruited, and performed a series of tasks with both the traditional and novel search result presentation. Subjects' use of search technologies was obtained through a pre-study questionnaire. An online post-study questionnaire was taken, and the performance in both conditions (with traditional and novel presentation) was assessed quantitatively using t- and ANOVA-tests. Though subjects' subjective opinions on the interface were gathered as part of the questionnaire, they are reported quantitatively only. The details of the questionnaires – such as the questions posed to subjects – are not given in full.

Mary Czerwinski has carried out a number of HCI evaluations of information-oriented systems, e.g. the Data Mountain of George Robertson [Robertson *et al*, UIST, 1998]. She typically uses thirty or more subjects, in two or more conditions, to evaluate and compare one or more novel systems against an existing, standard system. As with Chen and Dumais' work, a post-study questionnaire was taken and is reported only quantitatively and as a summary. However, the individual questions asked are listed – 9 in total, for which the subjects gave a score on a five-point scale for each of the systems they used.

Ahlberg and Shneiderman [1992], in their earlier work on *dynamic queries* which prefigured the HeirAxes system also used a quantitative evaluation, but on that occasion they compared the performance of three different modes of querying a given database. Again, a quantitative approach and post-study questionnaire are used.

There are consistent circumstances which can be associated with the use of quantitative evaluations. Firstly, two or more systems that address a known problem exist. Secondly, the criteria upon which they will be compared are well defined and, arguably, understood. For instance, in the three systems just reported, each provides a new alternative to an existing system (form-based querying vs textual browsing for dynamic queries, textual versus graphical bookmarks for the Data Mountain, and ranked list versus classified groups in Chen and Dumais' study), and the criteria for judgement of the performance of the systems are well defined (retrieval of a given document or set of documents, time taken). Results are given in quantitative, summary form, without much description of individual patterns of use.

In the case of HierAxes, the search facilities available are consistently a rich and complex extension of the simple approaches of many digital libraries - and as designed well reflects the work patterns in some repositories with complex metadata which the interface is intended to address. The controls available to provide input to the search system include interactive graphical controls such as sliders. Comparing HierAxes to an existing search tool would require mapping several searches in a web interface to a single one in Hieraxes; mapping visual, direct manipulation controls to textual inputs; graphical 2-d displays to linear lists, etc. Thus, despite addressing a common task – searching – providing a sufficiently limited number of variables between conditions, as is required for sound statistical evaluation, is extremely fraught. Therefore, the criteria being hard to define in detail, quantitative data is reported in summary and indicative form, and qualitative user feedback is reported in greater detail.

Before concluding this section, I will examine one final area which has been studied by HCI researchers that is relevant to Garnet. There has been considerable study in recent years of Internet and Web usability. Given that the permanent noting of documents of interest is the central task, a direct comparison with the use of bookmarks in web browsers would be beneficial, particularly in the case of digital libraries. There are, however, a number of reasons why this approach should be approached with caution.

Firstly, the overall rate of use of bookmarks is low, and the coverage and quality of bookmarks is known to be poor (e.g. many are never used, many others point to now non-existent pages) [Abrams *et al* 1998]. Studies particular to digital libraries indicate that bookmark use is of a similarly low level as on the Web generally [Blandford *et al* 2001].

One source of this behaviour may be explained in part that bookmarking is but one of a range of activities to facilitate the later recall of documents [Jones, W.P., *et al* 2001, 2002] – emails to colleagues, post-it notes etc. all play a role. Furthermore, it is not clear whether the particular paradigm in which bookmarks are interacted with and presented in web browsers has an effect upon the degree of use of bookmarks. The interactive properties of a system like Garnet in which the equivalent objects are clearly visible is very different to the common web browser where bookmarks suffer from low visibility. The best means for making this sort of comparison would be a long-term study.

In the case of bookmarking, studies have been conducted over extended periods, e.g. [Catledge and Pitkow 1995]. However, within the context of a novel tool such as Garnet, the scope for providing comparable data to such studies is low. Garnet is not as complete as a commercial web browser, and digital library deployment in practice is generally unsupportive of access by novel interfaces. Therefore, users would find it hard to interact with the material that would be of most to them, resulting in low take-up. Therefore, the very ground upon which such a naturalistic, long-term study could operate is in fact deeply problematic.

Secondly, the emphasis of Garnet upon document saving is so opposite to the priorities of traditional digital libraries, that again only a meaningful study could be undertaken over a prolonged longitudinal study. Therefore, for pragmatic experimental reasons, it was not possible to undertake a direct feature-comparison approach with Garnet in order.

In the case of Garnet, or any spatial hypertext, observations have shown that the use of bookmarking or "document saving" in digital libraries is very rare, and occurs, as is the case on the web, over extended periods of time. A longitudinal study would, indeed be interesting, but there was not the scope for performing this in the case of Garnet, as a large number of interfaces would have to be prepared to connect it to a sufficiently large range of DL systems to make Garnet attractive.

Once again, in this section it can be seen that long-term evaluations are problematic for a number of reasons. In addition, quantitative methods are closely associated with an opportunity to compare the operation of two systems of similar maturity which differ in one, well-understood part only. This provides some validation for the approaches found in other fields, where expertise in evaluating interaction may be less developed. Therefore, the evaluative methods used upon Spatial Hypertext and Digital Library interfaces can be adopted with a degree of confidence.

The next section will now discuss the very different requirements for Information Retrieval evaluation, which needs now to be studied in order to understand how the quality of users' organisation patterns can be assessed.

5.4 Information Retrieval Methodologies

Information Retrieval research is carried out in a very well-defined framework which differs significantly from the approaches that have been seen above in other fields. In Information Retrieval, evaluation is done in a highly quantitative manner, using mathematical processes that do not require the evaluator to have a human interact with the system being studied. Various standard tests have evolved which readily allow for the comparison of different search engines, and these tests are usually based on fixed tasks and corpuses (e.g. DELOS or TREC), with the results of a search being judged against a pre-prepared list of 'relevant' documents. However, evaluations performed using this paradigm are used to judge the performance of an underlying technology, not its interface.

The TREC series of conferences has included an 'Interactive Track' which reports on experimentation with interactive search interfaces, but much of the evaluation remains tied to the TREC paradigm, though some such as [Toms 2001] do examine the extended information seeking process.

More commonly, Information Retrieval papers with a strong interface element and a focus upon interactive issues are reported in non-Information Retrieval conferences. For instance, Chen and Dumais' CHI paper [2000] on ranked versus classified results presentation uses a more HCI-type approach to assessing the system. This includes quantitative performance data together with questionnaire and other user-generated data, and the evaluation of the interactivity is essentially quantitative.

In the case of Garnet, the main use of IR evaluation would be in determining the effectiveness of the user's organisation of the document space into groups, the representation of document groups and the subsequent matching of other documents against those groups. The classic model of information retrieval is the rating of the effectiveness of given query terms against known information questions on closed, specified corpuses. However, this task model maps poorly onto the use of Garnet, where the document groups are formed by the user, and therefore cannot be semantically predicted, and similarly the consequence of that unknown organisation means that the particular 'information question' represented by a group cannot be predicted or controlled. Thus, the classic method of IR assessment certainly cannot be used in a direct and un-adapted manner when evaluating Garnet, and may not be at all relevant given the difference between standard query tasks and the matching of document groups.

Document clustering is another task that has been measured and tested by IR. In this case, the means of assessment is less well-defined than for many other areas of IR research. The key difficulties in systematic evaluation here are that it is unclear how many clusters should be created, and what topics each cluster should contain, for a given set of documents. The most common approach, followed by e.g. Cutting *et al* [1992], and later by Zamir *et al* [1997] was to evaluate the clusters as if they were used for ranking the results of a query, the largest cluster first. The resulting ordering by cluster was scored as for a defined question or query as with the traditional TREC/IR approach. However, in a spatial hypertext workspace, where more than one information question is represented, this entire approach itself presents significant problems – as all documents are of some relevance, and there is no closed question or corpus upon which to build an evaluation metric. Furthermore even those using this approach, such as Zamir *et al*, question its proper reliability as a useful indicator of in-use performance – it is merely presented as one potentially viable measure.

Leouski and Croft [1996] further adapted the TREC approach within the context of clustering. In their case, they took a standard query corpus, and generated a set of hand-built clusters for a number of specified document collections and topics. Algorithms were then tested for the proximity of their result to the 'original' clustering done by humans. Ranking, etc. as present in the Topic Search track of TREC was thereby eliminated, and grouping alone remained. However, though satisfied with this evaluation system in use, they note concerns over the labour cost of creating the hand-built clusters, and the method was only used on a small number of specialised collections. Furthermore, this research seems not to have been carried forward in the following seven years.

If human organisation of documents can be used to validate the organisation by computer, can the reverse be the case? Well, there are problems. Firstly, clustering algorithms are (pseudo) non-deterministic. Given the same set of documents, the same clustering tool will generally create different clusters on different runs, as a degree of 'randomness' is deliberately included in each clustering algorithm to permit certain efficiencies to be gained. Therefore there is little conception of a single canonical clustering of a set of documents. I have seen no evidence of the consistency of different runs by a clustering engine being evaluated, so there is little knowledge of how great the degree of variance is likely to be in practice. Humans are also known to have a high degree of variation from person to person, and many factors such as expertise are believed to play a role [Marchionini *et al* 1993]. However, the consistent pattern of the IR community has been to judge computer organisation against that of humans (be the task ranking or grouping), and to reverse this would be novel.

Thus, remembering that clustering involves the partition of document clusters into sets by a software system, it may seem inappropriate to use such measures for evaluating Garnet, where the document groups – analogous to the clusters of clustering algorithms – are created by humans. However, two remaining aspects of clustering suggest themselves as appropriate measures for evaluating Garnet and its use – in determining the consistency and quality of the users' organisation of their workspaces.

Firstly, clustering algorithms typically use a "coherence measure" [§3.8.3] to determine the topical consistency of a prospective cluster. In the case of Garnet, such measures could be applied to document groups in order to evaluate their likely utility as cluster substitutes in the use of clustering algorithms – i.e. whether users create document groups that have high or low topical consistency. These metrics are believed not to be sensitive to alternative, 'valid', organisations, but to be useful indicators of poor organisation, and should be consistent (though not identical) across several separate runs of a clustering tool, each of which may create a different structure to cluster the documents. Therefore, it should also be a viable measure across the differing organisations produced by human subjects.

Secondly, the selection of representative texts for each document group is a key issue in the performance of Garnet, and as in the case of the identification of key phrases and key words to represent documents, comparison of human-generated versus computer-generated words may be instructive, as a close match would be a positive indicator that the results of document group matching would compare with those performed by a human. However, those words chosen either by a human or by Garnet may perform better or worse, as perceived by a human, in use. Again, IR researchers such as [Jones 2001] have performed evaluations to assess such circumstances. Their methods have typically involved comparing representations created by humans against computer selections through measuring the consistency of one against the other. The human selections are treated as definitive, and the quality measure is the number of items found in both as the percentage of the total computer-generated list of items, plus the common items as a percentage of the total human-generated list. These two measures may be respectively treated as the equivalent of the precision and recall of 'classical' IR.

5.4.1 Summary

Information Retrieval offers an approach to evaluation that is highly structured, which operates on a very different basis to the interactive material we have seen in the previous three sections. However, the predominant model for evaluating query effectiveness of search engines is not well suited to the evaluation of Garnet. The evaluation of clustering systems is less well understood, and the available evaluation methods have not yet reached the same level of consensus as for querying. Also, the automatic generation of groups as done by clustering tools is not necessarily a good comparison against which to judge human organisation of documents.

However, clustering tools contain accepted techniques for the judgement of the consistency of potential clusters of documents, and these were used as the basis for judging the consistency of human created document groups.

5.5 Review

The evaluation techniques of four different areas of research relevant to Garnet have been reviewed: Digital Libraries, Spatial Hypertext, Human-Computer Interaction and Information Retrieval.

In each case, it has been found that quantitative methods have been tied to well-defined problems in which the context of what is being tested is well defined. In Information Retrieval, this was through the use of controlled corpora and search topics, and in the design of interfaces, the provision of two or more alternatives for a given task with specified parameters. Conversely, where the subject of research is less well understood, qualitative methods are more common, these being deployed to elicit better understanding of the underlying activities of users, or of the immediate issues with a particular technology.

In the case of Garnet, information seeking and information structuring are combined in a manner which has been observed in the physical world, but which has not been available in the digital world. As was the case during Marshall and Shipman's early work on Spatial Hypertext, it is not clear whether the transfer from physical to digital is direct, or whether work practices change in consequence. There is no ready alternative which one can turn to with which to compare Garnet at this stage – there being many changes between a traditional digital library and Garnet, and the alternative visual interfaces to digital libraries are not generally available, nor are their features directly comparable.

Therefore, in evaluating Garnet, the most appropriate means of evaluation for an initial study would be qualitative – eliciting more information about its use to better understand how more detailed distinctions in implementation or between it and other interfaces could be drawn. In addition, some simple quantitative data, such as the basic indicative measures reported by Sheiderman [1999] in his study of the HierAxes search interface may be beneficial.

In terms of the technique of observation, the clear theme throughout the various studies related to new software tools has been the use of laboratory, task-based experiments, whether for quantitative or qualitative work, formative evaluation or detailed comparison. The degree of interruption of the experiment by the observer, and the degree of involvement of the observer is less clear. Many researchers give little if any coverage of their experimental method in that regard.

Turning to the issues regarding determining whether the user organisation of documents can be evaluated, the coherence measure of clustering algorithms provides one means of obtaining indicative information as to whether users produce document groups of appropriate quality for performing text matching against them. However, relying upon a single measure may be unwise.

The acceptability of the matching of documents to document groups to users may be evaluated in the course of the general user evaluation, through eliciting the participants' reactions.

5.6 Experimental Design

With the principles of the basic approach to evaluation having been determined, the detailed experimental design can be done. To recap, there are three key questions regarding Garnet, which I will now slightly expand:

1) Is Garnet acceptable to users?
   a) Is Garnet acceptable as a whole to users and are they able to successfully adopt the new interface style of Garnet?
   b) Do any of the current interface components suffer particularly acute problems?
   c) How does the system compare against existing information seeking environments?
   d) Do users suffer any confusion between the behaviours of different elements of the system (e.g. between objects controlled by the system and those controlled by themselves)?
2) What are the patterns of behaviour displayed by users engaged with Garnet?
   a) Do their visual patterns of organisation match those of Marshall and Shipman?
   b) Are these visual patterns identifiable by Garnet's Spatial Parser?
   c) To what degree and how often do users move between information structuring and information seeking tasks?
3) Is the use of the user's information structuring artefacts by the system effective?
   a) Does the structuring of documents by users provide a sound basis for performing document matching (from an Information Retrieval viewpoint)?
   b) Are the actual matches performed by the system against the user's grouping acceptable to the user?

It has been seen that item 3a is best addressed using quantitative Information Retrieval methods of evaluation. The remaining items are more appropriately answered by other means. Question 2 is best identified from a Spatial Hypertext perspective using the qualitative approaches of Cathy Marshall. The first questions fall more into the sphere of human-computer interaction and, were 1c better defined, could be addressed using quantitative measures. The remaining parts of question 1 are more effectively addressed using qualitative measures, especially in obtaining detailed understanding of Garnet's features and the problems that they may cause. The final part, 3b, is another technology acceptance issue, and again qualitative measures will elicit more detailed information.

Given the early stage of Garnet's development, qualitative approaches are more likely to yield the detailed information that is insightful into its use. How an exact comparison could be made for Garnet as a whole against the use of a traditional digital library is not clear, so only those common elements are worthy of a comparison. As the interaction with the library is governed by the behaviour of the Digital Library protocol, it is not simply a Garnet feature. Therefore, the only insight that can be readily achieved is of any particular defect of Garnet's interface. This is not itself worthy of an extensive study, though it is clearly a required part of the evaluation of Garnet as a whole.

The limitation of Garnet in deployment, due to the limited number of DL systems that have quality documents that are readily accessed over a DL protocol, and the relative simplicity of a number of Garnet's features, an extended, naturalistic study is not viable. This, therefore, pragmatically limits the viable options to a laboratory, task-based study.

Similarly, the novelty of Garnet precludes detailed comparison with other systems, and many of the most immediate systems with which it could be compared (e.g. another visual interface to a Digital Library – say, DLITE) are not readily available. The pertinent questions as to how to differentiate the tangible benefits (say, speed, or number of retrieved documents) and intangible benefits (e.g. user confidence, user comprehension) of different systems is not at all clear. A more effective and useful approach would be to elicit a more detailed understanding of how users utilise Garnet – a focus upon Question 2 – both to understand how Garnet may be used in 'real life', and to identify those features that may be later compared between it and other systems.

So, whilst the quality of the user's organisation of documents (given that their organisation is properly identified by the Spatial Parser) may be evaluated using the techniques of information retrieval, the other features are best captured using a laboratory environment, capturing the user's work patterns and eliciting the problems and benefits that they perceive with the system in use, evaluated by qualitative means.

Given this, I will now consider the more detailed design of the evaluation.

Taking the qualitative evaluation first, the general approaches seen across both Digital Library and Spatial Hypertext research bear clear similarities to each other. As has been stated above, most of the studies discussed earlier were observational, laboratory and task-based studies including small numbers of subjects. Data capture was often by interview or questionnaire, complemented by the recording of the artefacts of the participants' work.

Given the novelty of the systems, subjects were often given a basic, introductory training to familiarise them with the basic features of Garnet, including performing simple searches, adding documents to the workspace and removing documents from both search results lists and the users workspace.

In the following sections, I will address some of the more detailed aspects regarding my own experiments within this common template. This pattern matches the general format of much of my other work [Buchanan *et al* 2001, 2002] in which users were introduced to a new system, given a task to complete, observed during its execution and finally interviewed and/or given a questionnaire on their experiences.

In the following sections, I will look in details at each of the elements of the qualitative aspects of the experiment: the means of eliciting information from participants, the capture of data during the work of the subjects, the training given in using Garnet, the task used in the experiment and the equipment and environment used during the experiments.

5.6.1 Elicitation Questionnaires and Interviews

As already seen, two key means of eliciting information about the experiences of subjects in a study are questionnaires and interviews. Each technique has shortcomings in the data obtained. Interviews suffer problems in obtaining consistent coverage of particular points of interest, and one more readily obtains descriptive information of individual issues, rather than a systematic comparison between features. However, there are shortcomings of questionnaires too. Firstly, their fixed format offers a 'closed' view of the key questions and consequently can limit the elicitation of information, and secondly, they are problematic to create, due to complex validation issues. I will now briefly address first questionnaires and then interviews.

Questionnaires require extensive expertise to design and validate. It is common, therefore, to rely upon pre-designed, validated questionnaires such as the IBM 19-point system acceptance questionnaire [Lewis 1995]. This questionnaire is often used to validate the overall usability of a novel system, and therefore may seem appropriate to the evaluation of Garnet. In the course of our work on small screen interaction, my colleagues and I [Buchanan *et al* 2001] used the IBM acceptance questionnaire. The study we performed included a large number of subjects (> 50) and we successfully identified the key failure points of the system being evaluated. The use of the questionnaire revealed two important shortcomings, however. Firstly, little insight was obtained into the particular components of the system that caused problems, or the precise difficulties encountered. Given the closed nature of the questions available to the subject, this allowed for easier analysis over a large sample; the cost was at the reduced opportunity for open feedback to elicit the more detailed information that we also desired. Secondly, the closed set of questions was not particularly adapted to the system being evaluated, and any adaptation would have resulted in the invalidation of the questionnaire. Therefore, particular aspects of the system in which we may have been interested were not covered, and distinctions that were particular to the system were not reflected in the questionnaire.

In the context of performing a qualitative study with a smaller set of users, the development of a fully validated questionnaire is of questionable benefit – requiring significant input effort for only a small amount of use. Combined with the limitations of questionnaires in eliciting detailed information, a questionnaire cannot be the exclusive basis of information capture. Nonetheless, the structured form of a questionnaire is useful for the purposes of analysis, as it allows for systematic comparison of certain elements of the subjects' experiences.

Interviews give more scope for the experiences and opinions of participants to be elicited than do questionnaires, and most of the other studies seen in this chapter have either preferred interviews over questionnaires, or used questionnaires to supplement the material obtained face-to-face. In addition to the prevalence of interviews over questionnaires in the research conducted by others, I have used interviews as an elicitation device [Buchanan *et al* 2001, 2002], and so I chose to use a post-study interview as the main means of eliciting the participants' experiences and preferences.

In the case of interviews, it is important to consistently address issues across subjects, to allow for a more structured comparison of participant's experiences during analysis. As noted above, questionnaires do provide a regular framework that allows for ready comparisons. Therefore, for the evaluation of Garnet, a post-study questionnaire was used in the process of the subject interviews in order to provide a consistent framework through which analysis could more easily be performed.

This approach of questionnaire and interview can be seen, for example, in the evaluation of the Daffodil Digital Library interface [Fuhr *et al* 2002]. Fuhr *et al* used questionnaires to capture general information seeking skill deficits in the user population, and interviews to obtain a better understanding of the interaction of users with the Daffodil system.

The equipment used to record the interview and questionnaire will be discussed later in this chapter.

5.6.2 Post-Study Questionnaire

The content of the post-study questionnaires was intended to focus upon:

1) The subject's satisfaction with the basic search facilities provided compared to Digital Library and Internet search engines.
2) The subject's satisfaction with the representation of individual components of Garnet (e.g. document labels, user labels)
3) The advantages they felt that they obtained from the ability to store documents on the workspace.
4) The effectiveness and acceptability of the suggestions facility.
5) Compare the storage and organisation of documents to familiar systems, such as bookmarking in a web browser and storage of files on a filing system.
6) The acceptability of Garnet as a system.

Most of these questions address the issues listed in question 1 seen at the head of §5.6 – i.e. of acceptability of Garnet generally, whereas question 4 here relates to 3.b given in §5.6 and regards the provision of the scattering facility and the placement of suggestions into the workspace. The questionnaire as used can be found in Appendix D.

The first question is a simple proof-of-concept test. Given the previous evaluations of visual DL interfaces, it was likely that another visual interface would be likely to be acceptable, but this should clearly not be presumed. Furthermore, as described in Chapter 4, Garnet also has a number of behaviours that are different from existing search result representations (e.g. the ability to delete documents), and whether these presented any obstruction to the user was therefore a clear issue.

The second question emerges from the presentation of documents in the interface, which could affect the user's performance of both the traditional information retrieval task and the information structuring task. The larger the representation of each document, the fewer documents that would be visible at one time. Conversely, the omission of key information would impede the user's interpretation of documents. Therefore, it was important to elicit the participants' perception of the shortcomings of the presentation that Garnet used. This problem was analogous with my work on the search representations on small screens [Buchanan *et al* 2001], where the impact of the displayed document descriptors upon search result selection was investigated. Furthermore, as the users also move, resize and select documents through their representative label, and a fluid interaction is sought, it was necessary to identify any obstruction that emerged from the particulars of the current implementation.

The third and fourth questions relate to the extension of the DL interface to cover document organisation, and one of the corresponding unique features of Garnet, whereas the question on structuring facilities was intended to identify whether formal as well as informal structures would be useful in Garnet – users being introduced to the latter only – and to discover something of the user's perception of their requirements.

Finally, the overall reaction to Garnet as a whole was an important validation of the overarching concept, and an opportunity to identify areas of need not presently served, or problems which were not elicited elsewhere in the interview.

5.6.3 User Artefacts and Observations

During and after the experiments, additional data can be captured to complement the material gleaned from interview or questionnaire. Again, this practice can be seen earlier in this chapter and in my previous work [Jones *et al* 1999, Buchanan *et al* 2001].

As the study at hand is not intended to perform a like-against-like comparison, timing information is much less useful than would otherwise be the case.. Therefore, an interruption by the observer to elicit more information about the user's behaviour would not impact upon timing data capture.

The think-aloud protocol is another common form of elicitation, and can result in a very rich data capture. However, it has the problem of being highly intrusive when subjects have to give too much effort to articulating what they are doing rather than engaging in their task. Personal experience has also shown that some subjects are very difficult to get to vocalise their actions, and thus the selection of subjects becomes more problematic, or one must accept a high degree of variation across users. Therefore, I chose to rely on the post-study interview, or to engage in a question-asking protocol if a user prompted an interaction in response to a problem that they encountered.

During each experiment, an audio/video recording was taken to capture the activity of each subject, including any dialogue and their activity on the computer screen. In addition, I was present throughout the experiments as an observer, and took notes of the user's activity for later reference in evaluation and in the post-experiment interview. In accordance with Marshall and Shipman's practice, the user's workspace was also saved at the conclusion of the study. This also permitted the later evaluation of their organisation of the workspace using Clustering Coherence measures etc. and the comparison of the organisation of documents between subjects and their workspaces.

5.6.4 User Training

Garnet provides a new and unfamiliar interface to a digital library for most users or experimental subjects. This clearly provides a difficulty when one wishes to observe the system in use. Much of the research described earlier in this chapter faced similar problems. For instance, the various visual interfaces to digital libraries [Cousins *et al* 1997, Fuhr *et al* 2002] – and their evaluative work frequently includes a period of familiarisation and training with the system before undertaking the part of the experiment in which data is captured.

Similarly, in my own work on small screen information seeking, a new browsing paradigm needed to be introduced to users [Jones *et al* 2000, Buchanan *et al* 2002]. In each case, my colleagues and I decided to give a brief introductory training session to minimise the effect of the learning that inevitably occurs in the early stages of the use of a new system.

For the evaluation of Garnet a brief introductory session of about ten minutes was given to each subject. This introduced the basic operation of Garnet and the features relevant to the aspects of the system that I wished to assess. This included the performance of searches, the organisation of documents and the use of the suggestions/scattering facility. The user was run through a fixed script of features, modelling how the system might be used for a simple search, and then permitted some free time to explore the system further and confirm its operation to their own satisfaction.

5.6.5 Subjects

The selection of subjects for an experiment is important. A poor match between subjects and the experimental task, or between the subjects and the proposed environment of use of a system would result in methodologically unreliable data. For the purposes of a digital library system, one would wish to use subjects who have an appropriate level of knowledge of the information within the library, and who regularly engage in information work.

The subjects used for the experiment were HCI and Computer Science Graduate students. Given that Garnet is intended for use by information workers, this required a skilled subject population that was consistent with the target user population for the system. Graduate students are required to write essays, reports and other works which cite from and build upon the knowledge in published academic documents, and therefore they are a representative population in that regard. They are also more likely to be capable in basic computer operation, thus reducing any effects from lack of experience. On the other hand, they are unlikely to have developed extremely effective information seeking skills as one would find with Information Science students or research staff.

Each subject was given a pre-experimental questionnaire to gain some background on the users' backgrounds. A previous study of information seeking that I conducted [Buchanan *et al* 2002] had shown that the length of user experience with Internet use had little or no effect on user skill or expectations. The same study showed that the the degree of use had a strong correlation with both user skill and their subjective acceptance of the new system. Therefore, both the factors of the length of use and degree of use of particular systems were incorporated into my experimental design for Garnet. Example technologies where these two factors were tested for were the Internet, physical libraries and Concept Map software . The familiarity of users with systems that used spatial cues to organise and structure information, such as Mind Maps™ or flowcharts was also identified, as this may have had an effect on their work with the spatial hypertext facilities of Garnet. Similarly, the tagging of documents using the favourites or bookmarks option of a web browser, or the equivalent facilities provided within digital libraries such as the ACM DL was also tracked for any subsequent effects. Again, this practice can be found in the work of Chen and Dumais, Czerwinski and other researchers cited earlier in this chapter.

5.6.6 Experimental Task and Environment

User studies can be either naturalistic, in the participant's normal environment, or laboratory-based, in a closed, artificial environment, and may follow either a task chosen by the subject, or a task given to them by the experimenters. In the studies of Digital Library and spatial hypertext systems, the predominant pattern is for subjects to engage in closed, task-based studies. The shortcomings of this approach include that participants may suffer from lower levels of motivation and attention when engaged in the task, may be distracted by the unfamiliar environment, and that their behaviour for these and other reasons may be very different to their normal pattern.

An extended, naturalistic observation would capture quality information which could be more readily compared to the studies of information structuring in paper-based information handling tasks which prefigured Spatial Hypertext and which still forms a common part of information structuring today. However, this would require a developed environment which could be connected to a large number of digital libraries of practical value, together with support for the access right, subscription authorisation etc. which are commonly found in digital libraries such as the ACM DL. Such a degree of development would be far beyond the available time for this thesis, and the subsequent longitudinal and observational studies would also be time consuming. As a result, for the evaluation of Garnet, such approaches are impracticable. That they are so for other researchers too may well be evidenced by the lack of similar work in the material we have seen earlier in this chapter.

Given that a naturalistic study was not viable, a laboratory-based experiment with a controlled task was the most practical option – again, the option most commonly chosen elsewhere.

There is also scope for the task to be chosen by the subject. This again permits for better fidelity to the 'natural' environment of a system in use. However, variation between users makes comparisons more problematic, and in the case of digital libraries, the quality of fit between the material in the library being studied and the participant's particular interest can have a significant impact upon the length and depth of interaction with the library. Conversely, one advantage of a controlled, given task is that all the participants engage in the same activity, and one can ensure that appropriate library material for the task is to hand. This fixed approach is seen across most of the research described in this chapter, within the norm of a common task to all users performed in a controlled, laboratory environment.

An additional consideration is that when one needs to compare the subjects' organisation of their workspaces, some of the Information Retrieval metrics for Cluster Coherence require a fixed underlying corpus of documents (e.g. the Scatter/Gather coherence measure [Cutting *et al* 1992]). Comparisons are clearly easier when more factors are common across the experimental subjects.

Given that fixed-task, laboratory-based experiments are not naturalistic, it becomes more important to ensure that the task selected is as close to the subject's real field of interest as practically possible. The Computer Science Technical Report collection of the New Zealand Digital Library represents a substantial corpus of academic material over several years. As it was known that the most readily available subjects were HCI and Computer Science graduate students, it was decided to use this collection as the basis for the material to be searched in the course of the experimental task. However, it was also important for the task not to be one in which the subjects were likely to have a high degree of expertise, as this is known to have an effect upon search strategies and outcomes [Marchionini 1995].

Thus, the particular field of digital libraries was identified as being the object of their exercise. The subjects were given a brief verbal and written briefing on the task requirements, and the goal that they should achieve. A list of topics that may prove relevant to the field of digital libraries was given to enable subjects to choose a wide variety of query terms and themes as they saw fit, rather than over-directing the task. Too limited a set of suggested search criteria may have resulted in subjects choosing too narrow a selection of documents, and reduced the scope for gaining evidence for the range of variation between user's information seeking strategies and personal interests.

Each subject was given the same task (See Appendix C for the task description given to each participant), and an open-ended amount of time to complete it. Though some researchers have used time limits in their experiments, this has generally proven unnecessary in my experience. Rather, I wished to capture the users' workspaces in as developed a state as possible, and therefore anything that might prematurely end the work of the user was undesirable. In order to give the subjects a sense of the amount of material they ought to be able to obtain and the maximum work expected, a guideline target number of documents was given: 12 to 16. Marshall and Shipman have used similar task goals in their evaluations of Spatial Hypertext [Shipman *et al* 1995, Marshall *et al* 1996].

The use of a common task better facilitates direct comparisons, both for the purposes of evaluating the organisation of the workspace, and the documents selected.

5.6.7 Equipment

The experiments were performed in two usability laboratories, with similar equipment. In each case, the subject used a PC running Windows 2000, a SVGA (800x600) screen display, on which ran the Garnet client software.

The digital library server was operated from a separate computer – a laptop computer running the Linux operating system. A separate computer was used to ensure some degree of fidelity to the speed of reaction that would be experienced with a remote DL server.

The screen size was chosen to ensure the minimum degradation of legibility of the screen when recorded onto video tape. Subject's voices and on-screen activity were both recorded onto video tape for later review, and notes were taken both during the task and in the course of the post-study interview.

Subjects also completed a pre-study and post-study questionnaire, the formats of which are given in Appendix A and Appendix D respectively.

5.6.8 Summary

In this section, I have outlined the considerations for the design of the actual experiment I conducted. The use of a laboratory, task-based study is comparable to the practice of many of the HCI, Digital Library and Spatial Hypertext researchers seen earlier in this chapter. In addition to the elicitation of user experiences through a post-study interview and questionnaire, the results of the subjects' selection and organisation of documents was recorded for later analysis, and the reorganisation, selection and searching done in the course of their task was recorded onto video tape. Thus, both the users' impression of the system, and the manner in which they actually used it would be captured.

5.7 Analysis

The outline of the means of capturing the material to be evaluated has been given in the preceding section. However, the analysis of that material is of at least equal importance. Referring back to the list of properties of Garnet that I wished to study at the beginning of §5.6, I will now briefly discuss how the various questions listed there could be addressed.

The set of questions regarding the patterns of behaviour of users (Questions 2a to 2c) are best addressed by comparing the observed patterns to the visual patterns of organisation reported on by Marshall and Shipman [Marshall *et al* 1993, Shipman *et al* 1995] and the workflow patterns observed by Kidd [1994] and O'Day and Jeffries [1993].

The remaining questions, excluding those addressed through the information retrieval evaluation, (§5.6 Questions. 1 & 3b) can be analysed through the textual analysis of the reported experiences of the subjects both during the experiment and in the post-study interview. As with the studies of Fuhr *et al* [2002] and Shneiderman *et al* [2000], the qualitative evaluation can be validated with simple indicative, quantitative measurements.

5.8 Information Retrieval

The previous section described the information capture undertaken for the purposes of the qualitative aspects of the study of Garnet that I undertook. However, some of the questions given as being pertinent at the beginning of §5.6 were identified as being properly investigated through the use of Information Retrieval measurements. The actual measurements which will be used have already been discussed in Chapter 3 §3.8.3, particularly the coherence measurements for user document groups used by Zamir *et al* [1997] and Cutting *et al* [1992].

In §5.6.3, I noted that the subjects' organisation of space was captured through saving their finished workspaces. This also allows for the later analysis of their workspace organisation using the different measurements of the internal textual consistency of the groups of documents that they created.

In addition to evaluating the quality of user's organisation of space, in terms of topical consistency, the study may address a corollary of the same issue. Users studied by Marshall and Shipman [1993, 1995] often created 'miscellaneous' groups of documents which showed little or no topical consistency. One would not wish to treat these groups as being meaningful, i.e. one would not wish to match documents against them. Therefore, the success of the IR coherence measures in identifying such heterogeneous groups also needs to be tested in the course of the information retrieval evaluation.

5.9 Conclusions

The experimental evaluation of Garnet has been described in this chapter, and consists of both a qualitative, user study and a numerical, information retrieval analysis of the organisational patterns that users created. The qualitative study both scrutinises Garnet as a system and the patterns of behaviour which users engage in within a combined information seeking and information structuring environment.

The study method is comparable to many other studies in the areas of Digital Libraries and Spatial Hypertext, being a laboratory, task-based experiment with a modest number of subjects (ten were recruited for the actual experiments).

The actual results of the study, and the insights gained from it, will be discussed in the next chapter.

Chapter 6: Evaluation

In the last chapter, the experimental apparatus and method was described. This chapter will discuss the results of the experiment that I conducted. Ten subjects were recruited for the experiment, which was conducted as described in the previous chapter.

This chapter proceeds in three parts: firstly, the subjects' evaluation of the interface will be examined; secondly, the patterns of user behaviour observed in the experiment, including the organisation of workspaces will be discussed, and finally an analysis of the document grouping performed by the user and the users' responses to the suggestions system will be presented. Afterwards, a brief summary of the main outcomes of the evaluation will be given. Detailed information from the study can be found in Appendices E to H.

6.1 Pre-Study Questionnaire

The pre-study questionnaire was taken to identify some possible influences on the subjects' behaviour during the study. All the subjects were from a common, university background, as described in the previous chapter. I also found that little use was made by any subject of the advanced features of digital libraries that they used, such as interest tracking and personal book lists – only one subject used either of these features. No subject used diagramming or concept-mapping tools extensively. However, six subjects had used diagramming tools, six had used concept-mapping software, and of these four had used both. All subjects used libraries and Internet information sources, though the use of each varied; similarly, the use of bookmarks in web browser software was universal, but the degree also varied.

The influence of these factors (library and web use, book-marking, semantic diagrams) was found to be undetectable within the main experimental data, which is discussed throughout the remainder of this chapter.

6.2 User Feedback

The post-study questionnaire and interview were intended to glean the users' impressions of using the Garnet system. The key questions that were to be addressed were:

- The acceptability of Garnet as a system.
- Users' satisfaction with performing searches within Garnet, compared to digital library and Internet search engines.
- Problems and successes experienced by the users in interacting with the interface elements (e.g. user labels, document labels).
- Perceived advantages of storing documents on the workspace

- Comparison of the storage and organisation of documents with other familiar organisational tools (bookmarks, user filing systems).
- Whether users experienced problems in discriminating between system- and user-owned objects.
- The users' perceptions of the effectiveness and acceptability of the suggestions facility.

These will be discussed in this section, and then summarised before the patterns of user behaviour observed in the experiment are discussed in the next section.

6.2.1 Acceptability of a Spatial Hypertext Interface

The first issue of interest was the acceptability of the spatial hypertext interface for basic DL tasks present in current systems – particularly searching for and later reading documents. All subjects rated Garnet either "Average" (3 subjects) or "Easy" (7 subjects) in regard to the ease of performing searches. The comparison against the search facilities of existing DL systems the users had used and internet search engines (Google was given as a suggested benchmark) is given in Table 6.1 below:

Garnet search compared to...	Slightly Harder	Similar	Slightly Easier	Easier	N/A
DL search	1	3	0	4	3
Internet search	4	6	0	0	0

Table 6.1: comparison of Garnet's search versus DL and Internet search engines.

Our recent experience of assessing DL searching versus Internet searching indicates that that response is typical for DLs generally, thus further reinforcing the impression of a broad similarity. Therefore, the spatial hypertext interface was not generally intrusive into the activity of searching.

Garnet also includes a search history facility, to facilitate the recall of earlier searches. In the post-experimental interview, this was mentioned positively by three users – e.g. Subject 4 commenting "it makes it easier to see what you've done, and bring it up again if you want to double-check something". All subjects used the history facility at least once.

As was discussed in Chapter 4, Garnet relies upon a standard Web browser to render the actual documents in the library. It was anticipated that reading in a separate application could impede the user. The need to switch between the workspace and the web browser window to read a document was commented on by four users – though they also observed that it was an inherent problem. e.g. Subject 10 reported "you can't see it all on the screen, and switching from the Garnet thingy to the browser is a pain – I can't see how you can avoid it, though, unless you used tabs or something", and the subject reported the same problem with web-browser interfaces to digital libraries: "I get that with the ACM Library too – having to change from window to window". Clearly, this could be addressed at an additional time-cost of developing a web document viewer inside Garnet.

When we moved onto studying the keeping of documents on the workspace, the subjects were universally positive, all rating the feature as "useful" (9 subjects) or "very useful" (1 subject). The comments received from the subjects were similar to those that have previously been reported by earlier spatial hypertext evaluations, such as those of Marshall and Shipman [Marshall *et al* 1993, 1994].

Two comparisons were made between this method of recording documents and other methods:

	Much Better	Better	Similar	Worse
Bookmarks	1	7	2	0
Filing system	0	6	4	0

Table 6.2: comparison of organising documents in Garnet versus web browser bookmarks and user filing systems.

Users consistently rated Garnet as superior to the organisational support of web browser bookmarks. Often users said that it was easier to review what you had seen – e.g. Subject 7 "the bookmarks are often hidden – you don't see them unless you go and look, and often I can't bother". The high visibility of the document labels was, therefore, important.

When compared to storing the documents in a filing system, the broad consensus was that there were complementary strengths and weaknesses – two users stating that the systems were "different" and being unable to give a direct comparison. The remaining subjects gave a rating but observed that the spatial hypertext was particularly beneficial in the short-term. It should be remembered that they only used the informal structuring tools of Garnet.

Seven subjects stated that being able to organise documents into separate sets in formal structures would be helpful in the long term, and was an advantage of filing systems. The subjects had not been introduced to the "collections" facility of Garnet, which does provide a hierarchical means of organising documents.

The ability to create formal, hierarchical sets seems to be advantageous in the long-term use of a spatial hypertext. However, an observation shared by four of the subjects who requested a hierarchical organisation facility was that such a mechanism would make documents less visible. Their reasoning was that as at any time most folders would be closed, obscured or for some other reason not making their content visible to the user, this would significantly reduce the benefit of being able to scan the workspace quickly and identify the documents selected to date.

There was a consensus that the flexible, casual Spatial Hypertext was much better in the midst of performing seeking tasks. Conversely, a formal hierarchy was seen as better in the long term – space being a reported difficulty in the Spatial Hypertext that would be eased with the hierarchical filing system. The informality of implicit structures was good in the short term, e.g. Subject 3 "you just drop the document onto the desktop – you don't have to choose where to file it or...er...whatever – choosing where to store it just takes more time – it is easier just to drop it quickly and, like, move on".

One alternative suggested by Subject 10 was being able to zoom out of the workspace and see just the labels – in fact, this alternative method can be found in the Pad++ spatial hypertext by Bederson and Hollan [1994]. However, this would be an alternative to the use of formal sets, as space would again be the underlying basis of distinguishing groups, rather than a hierarchy. Individual documents would, at a "wide" zoom, be less visible, but as no overlapping of separate windows could occur, groups of documents would not obscure each other.

The quotation from Subject 3 suggests something of a tangible interaction with the system, and this theme was something that came through repeatedly. Eight users made comments about the document labels as if they were "real" things – e.g. "you just drag the documents and fling them about" (Subject 9), as was the benefit of gaining an overview. Seven users made comments like "you can just glance and see what you've got" (Subject 1). Participants evidently appreciated the direct-manipulation, visual, interaction and this is consistent with the usual claims for such interfaces. [Shneiderman 2000]

6.2.2 Interface Objects

Garnet's interface is composed of a number of individual types of objects, each of which plays a different role. Firstly, there are document labels that are the shapes used to represent individual documents inside the spatial hypertext workspace. These usually appear inside the search result lists which are displayed as a result of the user's information seeking activity, and are then placed onto the main workspace by the user dragging them from the search result list onto the main workspace. Secondly, there are User Labels – which are coloured, rectangular shapes which do not represent individual documents, but can be freely added, moved and edited by the user as they see fit. It was intended that these could be used as aide-memoires for subjects to identify explicitly the role they foresaw for each document group. Thirdly, there are the search result lists that appear after the user triggers an individual search. Finally, the subjects could use the search history list that was consistently available throughout their interaction with the system.

Beyond the overall experience of performing task-level activities such as searches, I wished to check for any particular difficulties or opportunities in the role and operation of these separate interface components. This section will take these different items in turn and reflect upon the users' responses to each.

Document Labels

One property that was quickly identified by the subjects in their training was the opportunity to resize the document labels, and this is one method which Marshall and Shipman give for users of spatial hypertext systems to express significance. However, in these tests, as noted above, the subjects did not adjust the size of the document labels from their default – even when the entire title of the document was, consequently, not visible. Subjects who accidentally resized the document labels (this occurred once each with two separate subjects) then rescaled the labels back to their original size.

There were opinions expressed in the post-study questionnaire regarding the content of the labels, and two questions existed to elicit related user requirements. Users were generally satisfied with the labels, but six subjects expressed a desire to rename documents once they had selected them onto their own workspace. This arose for a number of reasons. Commonly, one complaint was over the quality of titles in the collection. For example, documents from the University of Maryland typically had titles such as "TR-97-486", which subjects found, unsurprisingly, uninformative.

Though that was an extreme case, many other more traditional titles were seen as opaque or, alternatively, though true to the whole document, unhelpful in reminding the users at a later point what the document's relevance was to their task. It was this latter point which users gave to justify their desire to re-title documents: the need to clarify the relevance of a document to their task. This behaviour would be much more similar to the use of bookmarks in a web browser, and indeed four of these subjects, e.g. Subjects 2 and 9, did report that they frequently renamed bookmarks as "the names are often meaningless".

The particular collection used – the Computer Science Technical Reports collection, does have a number of atypical problems in this regard. For instance, many technical reports are titled by an abstract series number, as in the example above. I was aware of this beforehand due to other research I was associated with which used the collection. However, the collection was one of the few available that had appropriate content for the sort of experimental subjects I was likely to recruit. Nonetheless, the problem is by no means unique to this collection, and bears further consideration.

A second requirement that emerged was for additional information, requested by 9 out of ten subjects. However, there was little agreement over what additional information was required. For example, Subject 9 asked for date information, Subject 5 for keywords, Subject 7 for author information, Subject 2 for the abstract, etc. There was also concern over the amount of additional space that would be needed to display the extra information, so subjects often suggested being able to have it optionally displayed, through a dialogue, tool-tip or similar temporary form, rather than presented on the document label permanently.

This requirement is necessarily problematic in a small display space, something that I had previously investigated in the context of small screen interaction research [Buchanan *et al* 2001b]. The difficulty, when working across different collections, of variable metadata labels (e.g. 'author' versus 'creator' versus 'written by') makes what is inherently a troublesome task even worse. Title was certainly one of the few properties in the internet domain to be consistently rated by subjects, as was the URL, but this latter property is particularly unlikely to be helpful within a single digital library, as any difference will appear late in the URL, and thus be unlikely to be visible, or may be entirely invisible if supplied to the DL server in particular ways. This particular issue may be worthy of further investigation in a manner similar to the methods used in [Buchanan *et al* 2001].

Finally, two subjects also wished to be able to have visual feedback as to which documents were close enough together to form a group.

Re-Occurring Documents

Across a number of related queries, it is unsurprising that individual documents recur often – either in two separate queries, or when a document that is on the workspace appears in a query.

This pattern certainly occurred within the experiments, and it was a subject of interest as to how well users identified recurrences of documents, especially as the system as tested did not highlight documents which had been seen before, nor those in a search list which corresponded to a document already stored upon the workspace.

Users did find this to be a problem, but it was observed that it was particularly acute in the case of documents with "meaningless titles", such as the Maryland technical reports noted in the 'Document Labels' section above. In such cases, the problem could actually be observed during the experiment. For instance, Subject 10 returned to one such document some five times, though they did not report duplication as being a problem in their subsequent interview except in the special case of bad titles (a complaint made also by Subjects 3 and 9). Thus, excepting the particular problem of documents with poor titles, only two subjects found recurring documents a cause of difficulty. Contrariwise, two subjects remarked upon duplicates being helpful as it helped them be sure that their current searches were still in the right area of interest, rather than addressing substantially different, and perhaps irrelevant, topics.

During the design of Garnet, the problems caused by re-occurrence of documents had been foreseen, but the means of highlighting re-occurring documents was not an easily tractable problem, and this had been left un-addressed due to limitations of time. As with the use of colour and other properties, using any given highlight would limit the scope of the user to use, say, their own colours without running into potential ambiguities should they use a colour also used by Garnet itself.

User Labels

In addition to using the document labels which each represented a single document, users had the option of creating their own labels to identify the role or purpose of sets of documents. This facility was included in their training session at the beginning of each experiment. Not all users used this facility, though six subjects did use them regularly, and one (Subject 2) who did not use their own labels did comment on the usefulness of such a function at the end of the session – reminded of the existence of user-defined labels, they then returned to the workspace to label their own groups.

Those subjects who created user labels expressed the opinion that it assisted them in remembering what a group of documents was about. Subjects did not usually adjust the name of their own labels after creating them – though Subject 9 did so three times as they re-organised the layout of their workspace.

The use of user labels was in part related to the comments on document labels above, where users commented upon the intractability of some titles in the collection with which they were working. However, other motivations were present too: Subject 6 reported using their own labels because the titles of documents did not always express the significance that they themselves had found in the document. Similarly, of the three remaining subjects who never used their own labels, two expressed the desire to rather alter the titles of the document labels for the purposes of clarification, than lose additional workspace which they could have used for storing documents.

Search Result Lists

Garnet provides a number of interactions with search result lists that are not commonly seen. For instance, documents can be removed from the search result list either by being deleted, or by being moved from the search result list to the user's workspace (users have the option to copy a document rather than moving it, which would retain the document in the search list). One issue of interest was how well users adjusted to the new behaviours.

In the case of either moving or copying documents from the result list to their workspace, most users chose to move the document, the default action, rather than copy it. However, subjects 1 and 3 consistently opted to copy the document, finding the moving action too intrusive; Subject 1 stated that they did not like losing the document from the search list. Whether users would generally have opted to move had it not been the default action, clearly one cannot judge.

In the case of deleting documents from the search result list, most of the subjects seldom used this action, though Subjects 8 and 9 used it extensively to eliminate documents which they had inspected as being irrelevant to their task, a technique also observable over shorter periods of time by Subjects 6 and 10. Both subjects 8 and 9 reported that the deletion occurred too rapidly – and often they felt that they needed to double-check whether a deletion had occurred. Each stated that seeing the remaining documents move up (in an animated manner) so that the action was slower and more visible would have helped.

Re-running a search from scratch (closing the existing window for it, and re-running it) would replace the deleted documents, a strategy deliberately used by Subjects 5 and 10 to ensure that they hadn't prematurely rejected a document at an earlier point in time.

Generally, therefore, the ability to 'edit' the search list was not intrusive, and was used by all but Subject 1 at some stage. The explicit delete action was used by seven subjects, and extensively by two of these. There may be a need to have an explicit action to reinstate deleted documents from a list to support a later re-review of removals.

Search History List

Garnet, like other visual digital library interfaces such as NaviQue [Furnas and Rauch 1998] and SketchTrieve [Hendry and Harper 1997], includes a simple history of the searches performed by the user within the interface. This object can be used to re-run closed searches, or to make open searches become visible (e.g. if obscured by later search result windows, a search is brought to the front of the window stack). Though this was not the focus of the evaluation, subjects did identify its presence during their interaction, and eight subjects used this facility at least once. Six of these subjects mentioned the presence of the search history as a positive feature of Garnet in the post-completion interviews. No problems were reported with the search history.

6.2.3 User versus System Ownership

In Garnet, documents could be in either the user's own workspace, or in a search result set produced by the system. Although generally documents behaved similarly in each, there were differences. For example, users could drag a document from a search result set onto their workspace, but the opposite did not apply. Few difficulties were experienced with the differences in behaviour – the only exception being that two subjects attempted to drag a document back onto a search result set after belatedly deciding not to keep it, some time after having placed it onto their workspace.

Due to the relatively simple set of features to which subjects were introduced, the scope for confusion over ownership was less than it might have been. For instance, if users had been able to create their own folders or collections of documents, the appearance of these would have been similar to the search result sets, though of a different background colour and generally different in structure, as the search result sets had a particular, regular positioning of documents within them. However, subjects did not demonstrate any attempts to move documents within a search result list, which is something that they naturally could do elsewhere.

6.2.4 General Workspace Issues

Returning to the workspace itself, a couple of common themes emerged in the post-experimental interviews. Six subjects observed that as the number of documents in the workspace increased, the total available workspace would become limiting in the long term. As the tests were run at a relatively low screen resolution (800 x 600) to avoid problems with recording the screen display onto video tape, this is in part due to the experimental conditions. Similarly, the size of the labels could have been carefully reduced by a small proportion (c. 10% of both width and height) However, such measures would have defrayed rather than eliminated the eventual problem of filling the immediately visible workspace.

Seven subjects also asked to be able to save the workspace that they created for later use – this is another feature of Garnet to which they had not been introduced, and the request suggests an interest in preserving search outcomes for longer-term use.

When asked about the value of storing documents on the workspace, all subjects reported that this was beneficial (two rated 'very useful', eight 'useful'). The benefits cited included being able to quickly scan for an overview of the documents to date (5 subjects), verifying that documents had already been chosen or viewed (9 subjects) and keeping documents for later reference (8 subjects).

6.2.5 Summary

Users were able to successfully interact with the components of the Garnet interface. The form of presentation of documents prove acceptable, though some improvement could be made both with the particular collection of documents used in the experiment and with the information displayed on the label. The visual, direct manipulation style of the spatial hypertext interface proved effective, as research by Shneiderman and other HCI experts would lead one to expect, and as the experiences of the Spatial Hypertext community would suggest.

The digital library elements of Garnet, such as the search result lists, were readily adopted by the users, and in their opinion were either superior or comparable to the equivalent systems on the Internet. Thus, the Spatial Hypertext paradigm does not seem to interfere with the interactions which users expect.

The novel behaviours of Garnet's search result lists, such as the ability to delete documents, did not cause problems, and were used by all users at least once. Whether the traditional or novel behaviours should be the default remains unclear.

The presence of labels for the user's own descriptive use was appreciated. However, some subjects wished to be able to alter the representations of documents in the workspace. This unveils ethical concerns regarding the probity of permitting documents to be 're-titled', and practical issues in ensuring that when such a change is made, the document appears consistently with one form of description, rather than in both the 'original' and 'user' forms. Furthermore, when a document appears more than once in the user's workspace, in different roles, there may be conflicting requirements.

One concern which I had was the degree to which the different behaviours of documents when controlled by the system (e.g. when appearing as items in a search results list) or when controlled by the user (i.e. when in the user's workspace) would cause confusion. In the simple environment tested in this experiment, this difficulty did not seem to arise. This does not guarantee that problems may not emerge in richer environments, but it suggests that the problem may not be as acute as suggested by Shipman [2001a]. As observed above, some users wished to be able to "re-title" documents. However, this may cause confusion over the ownership of documents.

One set of difficulties that can, however, be anticipated from the experiment is that if workspaces become large the visibility of individual documents will be reduced due to either being occluded by other objects or falling outside the visible area of the workspace. This is another difficulty Shipman observes, and may prove more immediately problematic.

Overall, the components of Garnet that were tested in the running of the experiment caused few problems. Those that did emerge are either indigenous to spatial hypertexts (e.g. visibility in large hypertexts) or to digital libraries (problematic document titles).

The participants anticipated benefits in being able to keep documents for later use, and the visual presentation of their work was reported as facilitating their monitoring of their task. This benefit is one of those predicted by Kidd [1994], O'Day and Jeffries [1993], and the early work on Spatial Hypertext by Marshall and Shipman [1993, 1994]. More detailed observations on the users' use of labels will follow later in this chapter.

6.3 Documents – Patterns of selection and organisation

Having discussed the subjects' experience of interacting with the system, I shall now turn to analysing the artefacts of their behaviour whilst using it.

Firstly, the degree of similarity between the documents which individual users selected will be discussed. If users choose vastly different sets of documents, then it would be more problematic to make comparisons between their workspaces, as will be done later in this chapter.

Secondly, the visual patterns of organisation that emerged will be enumerated and compared to those patterns already observed by Marshall and Shipman [1993]. If the patterns were chaotic, unknown or not identifiable by Garnet's Spatial Parser, then there would be no hope of acquiring meaningful data on the use of facilities that exploited the user's organisation of documents.

Finally, the degree of interleaving between information seeking and information structuring tasks over the span of each subject's work is of interest – for instance, did subjects in an electronic medium behave as one would expect from the previous literature on information organisation?

6.3.1 Selected Documents

The total number of documents selected by users varied from 6 to 17: the mean value was 13.6, with a standard deviation of 3.34. In fact, the single subject with only six was somewhat of an outlier – as the next smallest number selected was 11 documents, whereas two subjects each had 17 documents, and half the subjects had fifteen or more. Overall, subjects chose very similar numbers of documents.

Only a small number of documents were consistently found in the subjects' workspaces: only five were found in the workspaces of 5 or more of the subjects, and of these, only one was in all ten workspaces and one in nine workspaces. Of all the documents selected by the users onto a workspace, 40% appeared on the workspace of only one user.

6.3.2 Visual Patterns of Organisation

Typically subjects organised their documents into four groups, with one or two singleton documents. The size of document groups varied, commonly being between two to four documents (29 out of 34 groups). However, subjects 8 and 10 used grouping much less than other subjects – they both created a single list for the vast majority of the documents that they selected. Most subjects had one or two singleton documents, but Subject 7 had a larger number (5 out of 13 documents).

Therefore, though the subjects had collected a reasonably small set of documents, they did tend to group the majority of documents they selected (90% of all documents). This gives a reasonable, if modest, sample of document groups for study.

Marshall and Shipman [1993] note some patterns of visual layout which commonly emerge in spatial hypertexts. Subjects were not introduced to the use of colour or to the sizing of document labels as visual cues (though the latter was often come across in the course of their interaction). Thus, visual spatial hypertext patterns using such features were very unlikely. Similarly, I was unlikely to gain any insight into the role of colour or size in indicating significance or topic. This leaves those structures that rely upon shape and proximity to impart similarity: columns, rows and piles [Shipman *et al* 1995].

There was a strong tendency for participants to create "columns" of related documents, mirroring the layout used in presenting search results. A particular pattern that occurred in six participants' layouts was a column of documents headed at the top by a label, which is technically, by the usage of Marshall and Shipman, a composite of a singleton of one type (the label), and a column of another (the documents). In VIKI and VKB, the consequence of this would be that the whole arrangement of label and documents would be treated as a second-order pattern rather than a first-order pattern, requiring additional effort for the user to select as a whole. Given the inextricable link between label and documents in this case, a specialised pattern parser would permit the easier selection of the entire group. For Garnet itself, its simple spatial parser, relying on distance alone, does not suffer this particular problem, but were a more advance parser created, this would be an important design consideration.

Subjects 1, 5 & 9 used the less regular "pile" layout extensively. "Rows" were observed with only one subject (4), and even then were used only once, whilst one subject (6) used a small grid arrangement for one group.

The visual organisation used by subjects thus closely corresponds to the patterns observed by Marshall and Shipman, and these patterns should be discernible to a spatial parser.

It is unclear whether the column layout commonly used by the participants is particular to these subjects or whether the column layout was influenced in some way by Garnet's interface. It is possible that the use of columns was primed by the search list, or perhaps by the visible workspace often being in a "portrait" shape, encouraging the more vertical arrangement of documents.

6.4 Accuracy of Spatial Parser

The visual groupings created by the users were in general readily distinguishable to the spatial parser. However, two subjects did present difficulties to the parser as implemented.

Subject 10 placed their document labels very close together, and generally in a list as discussed above. This subject also had two columnar groups headed by a label at the top. These were placed very close to the column of 'ordinary documents'. (See Appendix H.) The column arrangement is one which Shipman's parser is able to identify, so an improved spatial parser should be able to identify these groups more accurately, using the "column" pattern of [Marshall and Shipman 1993]. However, with as tight a packing of space as seen here, the positioning of suggestions may become more problematic, given the difficulties of the visual overlap which would probably result.

A second difficulty was observed in Subject 2's workspace. Two small labelled groups were created with a document placed ambiguously between them, as the subject was unable to determine which of the groups to place the document in. Given the informal arrangement of both groups, it is unclear whether any spatial parser would be readily able to identify such incidences, though it is a useful test case. The two groups and common document were semantically sufficiently similar for the resulting combined group to be deemed a quality group by both information retrieval metrics (see below).

Thus, some improvement to the Spatial Parser is required: though the internal form of groups (e.g. rows, columns) may not play a key role in extracting any semantic significance, it is important for ensuring that groups are properly distinguished from each other. Given the issues of space identified by the participants, a better parser which was less sensitive to densely packed workspaces may be an important improvement.

6.5 Information Seeking and Structuring

The visual patterns which subjects used have been discussed in the preceding sections. However, it is also interesting to note the work patterns that were adopted across the time of their participation in the study.

As has been noted §6.3.2, two subjects showed little inclination to organise documents thematically. Their post-experimental interviews revealed a divergence between the plans of these two subjects and the other participants. Subject 8 reported that they would have organised documents thematically at a later date, as part of their reading in depth. Subject 10, on the other hand, claimed that they would not have done so – instead, the theme of the workspace as a whole, for one task, was itself a sufficient organisation. These two subjects therefore demonstrated a coarse-grained, loose connection between seeking and structuring activities with the two parts more clearly separated than previous studies suggest is the norm.

The remaining eight subjects all demonstrated a closer interconnection between their seeking and structuring activities. In each case, documents were usually organised shortly after their selection, though this could be reworked later in the task when the user decided to alter the organisation that they had previously used (seven subjects performed one or more reorganisations of their workspace at some point in the experiment). These findings are consonant with the observations of O'Day and Jeffries [1993], Kidd [1994] and Bates [1989]. These researchers found that information seeking and information structuring are closely interwoven.

Given the eight subjects who organised their selected documents during their seeking, and the further subject who expected to organise the documents as a separate activity after their seeking was completed, there would appear to be good evidence for structuring being an activity required by digital library users. A further study could reveal what tie exists between their visual organisation of space, and their later intentions as to how to use the documents that they selected.

Three subjects (1, 6, 9) verbalised a decision not to seek any more documents upon a subject because, having referred to their workspace, they observed that they already had sufficient documents with that theme. Such patterns of behaviour are, according to Bates [1989] to be expected, and are consonant with the benefit of overview that the subjects claimed the workspace gave them in §6.2.1 above.

Subjects also clearly identified new topical themes in the course of attempting to organise a new document into their workspace. At a trivial level, this would result in the start of a new document group, or the labelling of an existing group to clarify its purpose. However, the identification of a new theme could result in new searches being performed. For example, when Subject 4 discovered two documents that discussed networking issues in digital libraries in the course of a search, these were selected to form a new document group (which they labelled 'Networking') and they initiated a new search with the query terms "digital library networking". Two further documents were later identified as being relevant to the new topic. Similarly, when Subject 6 was adding a new document to a group about the usability issues, she spotted a similarity with another document already in the group. This resulted in the two documents being moved to form a new, specialised group and a new search to recover further documents through a new search on "usability information retrieval" – a significant adjustment from the previous "digital library usability" search.

As subjects were not being required to express all their thoughts, further subjects may have made similar decisions. Subjects appear to have been demonstrating interleaved seeking and structuring as observed in physical environments, and were altering their seeking activities in consequence.

6.6 Suggestions in Practice

When the "scatter" system was used, the acceptability of the results was sensitive to the user's organising strategy. Users with distinct thematic organisation were generally satisfied with the results, whereas subjects with poor topical segmentation of their documents were disappointed.

For instance, Subjects 8 and 10 both used essentially temporal ordering – treating the workspace as a list in which any new chosen document was added to the bottom. In both cases, the subjects reported that they had only just collected documents – e.g. Subject 8 stated: "I hadn't really organised them yet – I'd do that later on…". These subjects found the results of "scattering" were unconvincing. A low level of topical consistency and thematic separation would result from their simple organisation strategy, and Garnet would attempt to match documents against a few frequently occurring words of little common meaning. It is unsurprising that the results of such an action would be of little value.

Conversely, subjects 6 and 9 managed to utilise the scattering system with great effect – the latter calling it "the magic document bringer" after a particularly successful triggering of it, from which he obtained four new documents. Later, the same subject failed to obtain any documents from another scatter, but then on inspecting the search result list stated that "these are all completely different – this is new" and during the post-experiment interview said "I was disappointed then – but when I looked at them, I realised that actually it was in effect telling me it was new stuff in there, which was true, so that was alright".

Five of the remaining six subjects found the suggestions made relevant, and the matches closely correlated to the documents they themselves would have suggested after a more detailed search or reviewing their own workspace. These subjects were evenly divided between those that used scattering in the course of their interaction and those that only experienced it during the separate final task used to test the effectiveness of the scatter.

Subject 7 used a very specific strategy of organising document groups by Author, which Garnet was able to track with good precision. However, the resulting suggestions were therefore more related by author than by topic (though topic similarity remained, it was clearly secondary). This particular pattern mimics one of the chaining strategies observed by David Ellis [1989] – exhausting the publications of successive authors as a means of ensuring a broad and systematic search on a topic. While subject 9 also used an author-centred strategy in selecting retrieved documents, they organised the results discovered topically. Subjects 1 and 9 also scanned references to derive a framework of authors who were related to the topic upon which they were searching. Although only one user used the organise-by-author approach; it prove less effective as a means of discovering relevant documents through scatter than was the case with equally systematic organise-by-topic use. This may be worth exploring further to evaluate the effectiveness of author-based retrieval generally.

6.7 Information Retrieval Evaluation

After the subjects had completed the experiments, the arrangement of their workspaces were then evaluated to determine the level of consistency in the organisation and selection of documents which had been chosen. As has been seen, the workspaces were generally well-ordered, and in a form which Garnet's spatial parser was successfully able to identify. This section will discuss the textual properties of the groups that emerged from the processing of the subjects' organisations by the Spatial Parser. The subjects' organisation of documents were compared against each other and against the nominal performance of two well-accepted clustering algorithms.

Before addressing the more particular details of document groups, it is worth noting the consistency or otherwise of the words which would have been used by Garnet to describe the whole workspace of each subject – those occurring most frequently among the chosen documents. Individual subjects varied in the style of document that they focussed upon, but there should be consistent themes relevant to the topic that the subjects had been asked to study – i.e. digital libraries. Unsurprisingly, 'digital' and 'libraries' occurred among the six most common words in every workspace. 'User' also appeared in every workspace, and 'retrieval', 'data', 'interface', 'information' and 'system' occurred in eight or more workspaces. This demonstrates that although participants often chose different documents, there were common themes that were not simply the product of the search task given to them. Furthermore, this reduces the likelihood that the scores obtained in the quality functions discussed throughout this section are more a product of the document selections made by a subject than their organisation of the documents that they chose.

6.7.1 Subjects' Organisation of Space

In Information Retrieval, Clustering Algorithms are used to automatically divide a set of documents into a number of smaller sets, each of which will ideally have a common topic across all of its constituent documents. Candidate document groups are tested for quality using a quality function, and those that fail the test are rejected, with good candidates being accepted for the next phase of the clustering. The exact method and strategy for dividing or merging varies from algorithm to algorithm, but there is a much smaller variety of measures for judging the quality of clusters.

The predominant form of measuring group quality is a "coherence measure". Higher coherence scores indicate better candidate groups. Therefore, using the clustering coherence measures on the subjects' organisation of documents may allow some insight into the quality of the organisation done by participants, in terms of the topical (i.e. textual) consistency of groups. If high ratings are achieved, this would suggest that the grouping of documents by a user is likely to prove a satisfactory basis for organising further documents at a later date, whereas low scores would be a negative indicator.

A number of small alterations to the scoring system were used:

1. The original Global Quality Function (GQF) of Zamir *et al* [1997],
2. The GQF altered to use the word-weighted scoring of Scatter/Gather
3. The GQF altered as described in §4.6.2 that is used to generate the representative words for each document group.

These separate scores should rank users similarly if the scores are to be trusted. If scores and rankings varied widely, then this may have indicated that such measures were unreliable. The original GQF function is susceptible to small differences in organisation, as if one document in a group does not include a word, then that word does not score in the group at all. The score of the overall organisation is also consequently lower. Conversely, the Scatter/Gather system always scores something for each word that occurs in one or more documents in a group. However, words vary their score in inverse proportion to their frequency. Thus, the weight of a rare word found in only one document is large, which distorts the score, and even common words have some effect. Together, these effects can mean that the scoring distinction between groups is somewhat lower than with GQF. The effect of these features is that the GQF is prone to yielding very different scores for subtly different organisations, and behaves in a highly 'polarised' manner whereas the Scatter/Gather system demonstrates very 'dampened' scoring, with substantial changes required to alter the scoring significantly.

How, then did the ten subjects' organisation of documents score according to the clustering algorithms? The score of each subject was compared against the other subjects and also against the score achieved by two clustering algorithms: Grouper [Zamir *et al* 1997] and Scatter/Gather [Cutting *et al.* 1992]. In addition to the overall score for the entire workspace, the individual score of each document group was calculated. In this section, I will only discuss the scores when compared to other subjects – the comparison with the clustering algorithms themselves will follow in the next section.

The scores for the individual groups created by a subject and the score for the subject's overall organisation of space proved to be similar in rank.

Four subjects (3, 4, 6, 9) achieved high scores when compared against other subjects and the nominal clustering scores, whereas three (1, 7, 10) performed poorly. A further subject (2) consistently scored in the middle of the rankings. The remaining two subjects' scoring and ranking varied between scoring methods. These two scored middling values and ranks, but the use of the relatively stringent original GQF or of the 'liberal' Scatter/Gather scores played a key role in their variation in scoring. This offers some evidence that the 'compromise' scoring system that I introduced in creating Garnet may provide a more stable indicator when evaluating human organisation of groups. In any case, the closer examination of which scoring system should be used would be a worthwhile further exercise.

6.7.2 Nominal Scores versus Actual Scores

In addition to comparing the quality scores of each subject against the others, each subject was compared with the maximum quality score that could be achieved for the documents that they chose when using the Zamir *et al*'s "Grouper" algorithm. The Grouper algorithm was run across the documents stored in workspace of each subject, generating a nominal score for each subject. The relative score of the Grouper algorithm and the subject's own organisation was then calculated. The results for this test were consistent with the ranking achieved when comparing the score of each user's workspace, as organised by the participant, against the other subjects.

This scoring also permitted the identification of the influence of document selection upon the scores achieved for document organisation. One concern that might arise is that the scores for document organisation were in fact the product of variation between the underlying document selections – i.e. that the subjects' choice of documents was more significant than their skill in organising their documents. As can be seen from Appendix G, there is little correlation between the nominal score for the workspace and the subject's score for their organisation of the documents. Subjects who scored well when compared against other subjects also scored well when compared to the nominal score that could be achieved.

Finally, another comparison can be made – between the scores achieved for each subject's workspace and a "random" organisation of the same documents. This permits comparison against a naïve organisation strategy. Each subject's documents were organised into two forms – grouped into sets of three or four documents[1] in order of accession to the workspace (i.e. {1,2,3}{4,5,6}{7,8,9} etc.) and secondly in order with successive documents appearing in successive groups (e.g. {1,4,7}, {2,5,8}, {3,6,9}). The two scores achieved were averaged, and the subject's scores compared against the scores for these random organisations. If the organisational strategies of participants were chaotic, then subjects would achieve similar scores to the background, 'random' score. Again, these scores can be read in Appendix G.

For most subjects, the score that their own organisation achieved was higher than the "random" scores. I observed above that there was a correlation in rank and score when a participant's organisation was compared with the organisations created by other subjects and that of a clustering system. Likewise, there appears to be a relationship between a participant's rank under either or both of these other two systems and their rank in this third system. Given this consistency, one can be more confident that the relative position of subjects is accurate and that each test is trustworthy.

6.7.3 Identifying 'Miscellaneous' Document Groups

One problem that could be foreseen whilst designing Garnet was the problem of users creating one or more 'miscellaneous' piles in their workspace. When a "scatter" was performed, this may lead to the user receiving poor quality suggestions, matching documents against a highly generic and diffuse group of no topical significance.

The first aspect of this issue that I considered was to what degree users did create piles of unsorted documents, and in which forms were these found. Some of the features of the participants' organisation of space have already been identified, and Subjects 8 and 10's use of a single unsorted list is a clear example of one form of 'miscellaneous' pile. As will be seen, these users gave a low rating to the suggestions placed near these long heterogeneous lists after a "scatter". In these cases, the users' admitted organisational strategy and the low quality scores given to these groups under IR analysis can give us a high confidence in the document groups being unsorted. Subject 8, however, did create a small second group of documents which boosted the score for their overall workspace.

Given the simple structure of the workspace in both these cases, a single list, the low workspace score directly relates to a poor individual group. However, in clustering there is a tension between a good overall division of documents, resulting in a high score for the workspace as a whole, and high quality within document groups, which may result in one

[1] The size of group is consistent with the behaviour of subjects in their own organisation, and of the best scoring organisations created by the Grouper algorithm.

or two well-defined, high-scoring clusters offset with a larger number of more topically varied, low scoring clusters.

This is reflected in Subject 7, who used a strategy of having a large number of singleton documents. The overall workspace score is low (singleton documents score '0' in most clustering quality measures), but the individual scores for the two document groups that the user did create scored close to mean group scores (within one standard deviation for both Cutting and Karger's and Oren and Zamir's measures). This user's unusual organisation strategy is discussed elsewhere in this chapter.

Other subjects created workspaces that consistently scored highly under IR analysis. Subjects 3, 6 and 9 all scored well in the different IR systems for their overall workspace quality. The individual group scores for these users were consistently above the mean (12 of 16 groups). Of the four groups that scored below average, three were outside one standard deviation, and one was marginally within one standard deviation. The degree of separation between these groups and the others would seem to be clear. In each case, the identified group corresponded with a set of documents that the subject had reported as being associated with uncertainty of relevance or of theme.

Mid-scoring subjects repeated the pattern observed with high-scoring participants. Of the fifteen groups created by these four subjects, five scored less than one standard deviation below the average group score, and nine above the average – again suggesting a distinction between two separate types of groups.

This tends to indicate a bipolar distribution – with 11 low-scoring groups (e.g. < 0.6 with Cutting and Karger's self-similarity score[2]) and 23 high-scoring groups (> 0.675 self-similarity score). One subject (subject 7) has no low scoring group. Two subjects each have two (Subjects 1 and 5), and the remaining all possess only one in their workspace.

Subjects did report 'miscellaneous' piles, but not all of the 11 low-scoring piles match this description. For example, Subject 1 had no 'miscellaneous' pile, and their two low-scoring groups were labelled "Technical Issues" and "General Library Access". In each case, the selection of documents was somewhat diffuse, but they cannot be described as being simply "random". Subject 3's low scoring group was labelled 'Interfaces', yet contained one document on the verification of systems, perhaps an erroneous selection.

The overall impression is that the quality measures for groups are good at identifying heterogeneity, but that the user's perception of the group may be somewhat different to it being a 'miscellaneous' pile. This subject would appear to remain worthy of further study.

---

[2] This score is between 0 and 1 on a logarithmic scale

6.7.4 Impact of Organisation on Suggestion Effectiveness

Three subjects particularly engaged with the scattering/suggestions facility of Garnet, and proved effective with it: Subjects 3, 6 and 9. It is notable that when their workspaces were ranked by Zamir and Eztioni's Global Quality Function, these three subjects were the highest rated. Subjects 6 and 9 rated the facility as 'Very Helpful', and 3 as 'Helpful'.

On the other hand, Subject 7, whose organisation was rated poorly by the GQF, received suggestions which they believed to be relevant to the individual document groups, but not related to their task. Their organisation was unusual, by author, and the documents retrieved matched that pattern well.

Subject 10 received notably low quality suggestions, with no documents that they rated as relevant, and their workspace organisation was rated a very poor 10th by both GQF measures. Subject 8 did fare better, despite having a superficially similar organisation, i.e. the primary group was a single list of nearly all documents, receiving two suggestions which they rated as of moderate relevance to the task. However, the rating of their workspace organisation was rated higher (the raw score being twice that of Subject 10).

Subject 5, the third low-scoring subject in the GQF test, received two suggestions of moderate relevance, but rejected them as being too specialised (this subject selected only six documents in total – citing the same reason for selecting so few documents generally).

All of these less successful subjects gave indifferent responses to the suggestions system, as described above in the 'Suggestions in Practice' section of this chapter.

The remaining subjects all rated the suggestions system 'Helpful', and were able to validate one or more suggestions made as 'Relevant'. Thus, there appears to be a relationship between a subject's quality of organisation, as measured by the clustering quality functions and the perceived quality of suggestions that they received from Garnet.

6.7.5 Summary

This section reported the results of the information retrieval evaluation of the subject's workspace organisations. There was a common pattern to the ordering of subjects' workspaces when scored against each other, against the nominal score achievable by two clustering algorithms upon the same documents, and against a random organisation of their workspace. This ordering also relates to the perceived quality of suggestions made by Garnet – participants with high-scoring organisations rating the suggestion quality more favourably than those with low-scoring organisations.

Together, this suggests that suggestion quality may be estimated from the quality score of a user's workspace.

6.8 Discussion

As a basic DL interface, Garnet proved acceptable to all the subjects who undertook the experiment. No effects were identified from the pre-study questionnaire upon the subject's rating of Garnet, either from library or internet use. Therefore, a spatial hypertext interface to a digital library would seem to be a viable alternative to the traditional, web-based interface.

In the case of the new, organisational facilities that Garnet provided, all users used the facility to some degree. Two subjects used the workspace to build up a list of documents rather than grouping the documents thematically, and one subject placed many documents individually. However, the remaining seven subjects organised the vast majority of their chosen documents into groups of varying sizes.

All subjects reported that keeping documents on the workspace had benefits – most commonly, that it assisted them rediscover documents that they had already seen, and that it helped them gauge the current progress of their seeking. Similarly, subjects consistently reported favouring the visual and interactive style of the Garnet interface.

Improvements were also suggested: in addition to the casual organisational tools tested in the experiment, users reported that formal organisational tools such as folders in traditional filing systems would have complementary benefits; subjects also reported a desire to be able to adjust the titles of documents to reflect their own understanding or clarify ambiguous or confusing titles.

The success of users in using the suggestion facility which built upon their organisation of their workspace varied but was generally positive. The most positive subjects about the feature were also those subjects that scored well on an information retrieval analysis of the quality of their document groupings, either in terms of the overall organisational quality or in terms of individual group quality. Therefore, such measures may be an effective basis upon which to assess whether to match documents as potential suggestions against a group or within a workspace – those groups that score low being ignored. However, that metadata may in fact play a role is demonstrated by the fact that a well-defined structuring task – the organisation of documents by author – scored low. If metadata were not used, or were not available, some laxity of group coherence scores would be required were sound suggestions not to be prematurely rejected. This aspect of design is worthy of further investigation.

Returning to the research issues of Spatial Hypertext, much of what has just been discussed within this section is relevant to the investigation of computation over documents. It would appear that the textual processing of underlying documents may indeed be capable of providing useful information for some information seeking tasks. This clearly has not been processed in depth in the experiment, but there is sufficient indicative evidence to validate that further, detailed, study is warranted. Given the noted scarcity of examples of computation over hypertext, Garnet therefore stands itself as a contribution to how such computation could be done, and the potential benefits. Secondly, in Garnet both the system and the user play a role. In the experiments, subjects demonstrated no confusion over which documents were placed by themselves or by the system. However, given the limited range of expression which subjects used, the scope for confusion was probably reduced. Therefore, further investigation into the relationship between expressive freedom and confusion of ownership is required. Nonetheless, it is clear that multiple actors working in the hypertext need not necessarily lead to confusion. Finally, Garnet provides an example of how an information environment may be placed within the context of a spatial hypertext. Here, subjects were able to interweave their interaction with digital library and spatial hypertext facilities without confusion, and adopt a workflow that previously would have occurred in two discrete parts and within two discrete systems.

Within the context of Digital Library research, users were able to adopt a Spatial Hypertext interface onto a digital library, and effectively use the library facilities. In addition, they demonstrated an interwoven work pattern of both information seeking and structuring which reaffirms the claims of researchers such as Kidd [1994], and O'Day and Jeffries [1993] that the organising of information and the discovery of information are two tasks which interplay at a fine-grained level. The experience with Garnet is a positive indicator that this can be achieved in an electronic as well as a physical environment.

These insights will be discussed more fully in the following, concluding chapter.

Chapter 7: Conclusion

At the beginning of this thesis, I introduced digital libraries as a means of supporting the information seeking of users. In Chapter 2, information structuring was introduced as an important complementary task to information seeking, and spatial hypertexts, as software systems which supported information structuring activities. It was noted that spatial hypertext systems had, however, seldom been connected to a system from which documents could be retrieved. Digital libraries, on the other hand, contain few if any tools for information structuring. Given the interwoven nature of information seeking and information structuring reported by researchers such as Kidd [1994], O'Day and Jeffries [1993] in the physical environment, combining the two elements of seeking and structuring in a single electronic environment was worthy of investigation.

Two research questions emerged. One, could Spatial Hypertext be used as an interface to digital libraries – would the combination of information seeking and structuring be as useful in a digital system as a physical one, and could an information source such as a digital library be successfully accessed through a spatial hypertext interface? Secondly, in a spatial hypertext, could an information retrieval system exploit the user's organisation of documents?

A spatial hypertext interface to digital libraries – Garnet – was implemented as described in Chapter 4. Garnet was subsequently evaluated in a small, indicative user study, which was reported in the previous chapter. In this chapter, the outcomes of the study and the findings of this thesis will be summarised, the contributions made discussed and future work outlined.

Firstly, the use of Garnet as an interface to the information seeking tools of a digital library will be discussed. This will be followed by a summary of the information structuring and seeking behaviours of participants that were observed during the experiment. The contribution of Garnet to the field of Spatial Hypertext will then be reviewed. The chapter will then move onto the features of Garnet that combine information seeking and structuring, which focuses more upon the issue of whether the user's organisation of documents in the spatial hypertext workspace could be used to improve information seeking.

7.1 Spatial Hypertext as a Digital Library Interface

The first element of the experimental study was the study of Garnet as a visual interface to digital library facilities. Given the existence of other visual interfaces to digital libraries, e.g. [Cousins 1997][Hendry and Harper 1997], there was good reason to have confidence that a spatial hypertext interface, as a particular form of visual interface, would be effective.

The study found evidence to support the suitability of Spatial Hypertext as a digital library interface. When the basic information seeking facilities were compared against web search interfaces and digital libraries, Garnet was rated as similar to or easier than a web-based digital library, and as similar or slightly harder when compared to a web search system such as Google™ §6.2.1. Subjects were able to perform searches without any noticeable hindrance, and viewed documents to gauge their suitability at will. Other, less commonplace, digital library features such as the search history list, were also adopted by the subjects and used without error. Garnet's visual interface and workspace were positively received – subjects reported advantages in having an overview upon the documents that they had selected, saved documents to the workspace without experiencing problems, and organised those documents into thematic groups.

Existing visual interfaces to digital libraries have been mentioned in §3.3. Benefits of overview and document retention have been claimed for other visual digital library interfaces [Cousins *et al* 1998], [Rauch and Furnas 1998], [Hendry and Harper 1997]. Some evidence is found in these earlier studies to support those claims, and the outcome of the study of Garnet gives those claims further credibility.

However, in addition to the information seeking activity, I was also interested in the information structuring activity undertaken. Garnet represents a novel visual interface style based upon Spatial Hypertext – emphasising information structuring. As was outlined in §3.3, the existing visual interfaces have a number of shortcomings when considered from a Spatial Hypertext perspective, and have much more limited support for the activity of information structuring. The experimental use of Garnet indicates that users of digital libraries do value information structuring, as one would expect from the insights of information seeking research. Thus, the consideration of information structuring tools should be part of the design of future visual interfaces to digital libraries.

As will be seen in §7.3, the evaluation of Garnet provides evidence that there are benefits to be gleaned from the information that is implicit in the visual organisation of structures in visual workspaces. These benefits may also be found in the organisation of items other than documents – for instance, in the organisation of search result lists, notes, etc. Already, digital library researchers such as Hendry and Harper [1997] have speculated upon this possibility and future visual interfaces may learn from the experience of Garnet. Conversely, other visual interfaces have provided facilities that could be added to future versions of Garnet – for instance, DLITE provides drag-and-drop means for combining query terms with different search engines rather than re-typing the query terms into each, amongst other advanced means of interacting with multiple digital libraries.

7.2 Information Structuring

The organisation of documents selected by users during information seeking is seldom supported in traditional library interfaces. This is the sort of activity that Marshall and Shipman [1993] argue is supported by Spatial Hypertext systems.

My interest in information structuring is also informed by the studies of Kidd [1994] and O'Day and Jeffries [1993], who observed particular patterns of behaviour in physical information seeking environments. Whether those patterns occurred in an electronic environment was also an issue of interest.

To evaluate its support of information structuring, Garnet was compared to the use of both bookmarking systems in a web browser and filing systems in a personal computer. Subjects had been introduced only to the informal structuring tools of Garnet. For short-term work, informal structuring was seen as superior to the formal structures prevalent in filing systems. This reliance on informal structures was also observed in physical environments for short-term and localised organisation [Malone 1983].

Participants also reported a requirement for the explicit structures favoured for large-scale, long-term storage. In the context of extended information work in information-rich environments, the number of documents referred to by the user is likely to be large, and the topical range broad. The requirement for explicit structures is therefore likely to be strong, and thus relying upon implicit structures alone, as is the case with systems such as Pad++ [Bederson *et al* 1994] or NaviQue [Furnas and Rauch 1998], is open to question.

Thus, the inclusion of formal structures in Garnet seems validated, but the informal structures are clearly valued by users in the immediate context of acquiring documents.

Against browser bookmarks, Garnet's facilities were favourably rated §6.2.1. The prominent, consistent role given to documents visually in the workspace, and the tangible, direct manipulation interface of a spatial hypertext seems preferable to web bookmarks for information structuring tasks. Comments on both visibility and tangiblity were made by the experimental subjects.

Subjects used the structuring facilities throughout their interaction with Garnet. The subjects interleaved seeking work between periods of structuring activity. This behaviour mirrored information structuring seen in the study of information seeking in physical environments [Kidd 1994][Bates 1989]. In the field of spatial hypertext, Marshall and Shipman [Marshall *et al* 1994] observed the pattern of interleaved searching and structuring in physical environments.

Furthermore, the users reported the benefits of the combined system in terms of improving their review of the work that they were doing, and in helping them co-ordinate their ongoing seeking activity. Again, this echoes the findings of researchers such as Ellis, Kidd and Bates in their studies of information seeking. Hendry & Harper's [1997] evaluation of their SketchTrieve system observed similar advantages. Given that both the behaviours observed and the benefits reported by the subjects correspond with existing research, this can give added confidence to the findings of this study.

To conclude, users interleaved information seeking and information structuring activity in Garnet as previous researchers observed in physical environments. In addition, the short-term use of informal structures, supported by explicit structures for long-term work, seen in physical environments was also observed in the subjects' use of Garnet. The correspondence between previous studies and the one presented in this thesis further supports the hypothesis that information structuring is a beneficial support for information seeking in electronic as well as physical environments.

7.3 Spatial Hypertext

In §4.10.1 I discussed the relevance of the implementation of Garnet to Spatial Hypertext research. Also in that section, some anticipated problems that might affect a system such as Garnet were briefly introduced. The embedding of an information environment into a spatial hypertext, the synchronous collaboration of multiple agents in a single workspace and the performance of computation over a spatial hypertext, all of which occur within Garnet, were identified as areas where outstanding issues in Spatial Hypertext research arose.

Experience with Garnet seems to suggest that some of the problems to be expected when spatial hypertexts are connected to a wider information environment, such as representing the information environment within the hypertext and conflicts over ownership of objects, are less pointed than was previously feared. However, further work needs to be done to explore whether the computation over hypertext can yield benefits – either through more detailed analysis of the suggestions system, or through other directions.

In Chapter 6 [§6.2.1] it was observed that there was little evidence that a spatial hypertext interface hindered interaction with a digital library. Experience with other visual interfaces to digital libraries suggested that a visual interface would not obstruct interaction with the digital library, so this finding can be held with some confidence. However, the converse problem – the degree to which adding other facilities into a spatial hypertext workspace hindered the use of the spatial hypertext – is also pertinent when integrating spatial hypertext with the information seeking environment.

Subjects were able to exploit the organisational facilities of spatial hypertexts, and reported favourably upon the benefits of these facilities. The findings in Chapter 6 suggest that the integration of spatial hypertext and digital library did not interrupt the users' ability to use the spatial hypertext facilities. Subjects selected and organised documents in a manner consistent with the behaviours reported by Marshall and Shipman's studies. However, it must be borne in mind that participants did not have prior experience of spatial hypertexts, and so it was not possible to gauge the obstructive effect of the digital library elements upon the spatial hypertext as in the opposite case.

Spatial Hypertexts do not use hypertext links for organisational purposes, instead they use the position of objects in their workspace. The visual structures created in the hypertext can be detected by a Spatial Parser, which thus provides a foundation for computation over spatial hypertext. The existing, simple, use of computation over spatial hypertext that identifies the visual structures in the hypertext (i.e. spatial parsing) was noted in §4.10.1. Previous systems have facilitated the selection of implicit document groups that were identified by the spatial parser. This does not require any use of the content of the documents – an opportunity which has not previously been taken.

Computation over traditional, hyper-linked hypertext systems has proven at times to be an effective boost to computational and user-centred tasks, particularly where document content is used in addition to the hypertext links themselves. One example is the exploitation of links between internet pages in the PageRank algorithm [Page and Brin 1998], used as the basis for the Google™ search engine.

The advantages of using the document content within computation over a spatial hypertext are therefore worth exploring. Garnet's processing of the text of document groups thus provides a contribution to Spatial Hypertext research merely by its implementation. However, that a process can be performed does not tell us if that process is useful. Therefore, Garnet's "scatter" facility and its "find similar" search facility are of interest in terms of their effectiveness – does computation over a spatial hypertext provide any potential benefit, in terms of improved document retrieval?

The technical discussion of the detection of the user organisation of documents, and the textual characterisation of the document groups will be appear in length in the next section. However, the findings of §6.7 certainly indicate that there is promise in the exploitation of computation over Spatial Hypertext, even if the methods are clearly at an early stage of development. Subjects found that the suggestions made by the "scatter" facility were pertinent to their interest.

The visual organisation has also been used in a small way to determine the position of suggestions placed during a "scatter" – but this could be used with more refinement to also affect the position of suggested documents being confirmed as belonging to a document group. The thematic groups extracted from the visual organisation may also be used to provide profiles for notifying the user of new or changed documents that are relevant to their interests. Such notifications could be displayed on the workspace in a similar method to that already used for 'scattering'. A document being written could be matched against these groups to identify appropriate citations, in a manner similar to the Niles system of [Jones, S., *et al* 1999a]. Finally, the topical groups could be matched with classifications in the subject hierarchy of a previously unseen digital library to focus the user's initial browsing to the areas most relevant to their existing fields of interest.

The experience of combining a spatial hypertext system with a digital library suggests that the general information environment can be used without disrupting the structuring activity of the spatial hypertext,

7.4 Detecting User Organisation of Documents

Garnet endeavoured to use the information structuring activity of users to improve their later effectiveness in information seeking, through its "Scatter" facility §4.2.3. For the resulting suggestions to be potentially acceptable, the user's visual organisation of documents must be discernible to the Spatial Parser and the thematic organisation had to be consistent so that the resulting textual representation of the group is, by the standards of Information Retrieval, sound.

In §6.3.2 I observed in my experiment that each subject created a number of visually distinct document groups. These groups typically consisted of three to five documents. Given the apparent spatial differentiation between most groups, there was some hope that these groups might be discernible to Garnet's spatial parser. The Spatial Parser was indeed able to identify most of the document groups created by users, §6.5. The shortcomings which occurred will be discussed in the 'Future Work' section of this chapter.

The next question was whether the visually identified groups would be thematically (i.e. textually) consistent or not. The text of each informal document group and each entire workspace was analysed using two different quality functions from clustering to evaluate the textual coherence and consistency of the subjects' organisation of documents [Zamir *et al* 1997][Pederson *et al* 1995]. The quality score of most groups was comparable to that achievable by a typical clustering algorithm §6.7.2, and where groups were provably heterogeneous (as in the case of the simple list structures used by two subjects) the scores were substantially lower.

Clustering algorithms such as those whose quality measures were used here have proven viable means of organising documents to support the information seeking of a user. The groupings performed by our subjects scored considerably higher than a random organisation, and were comparable to the scores achieved by well-accepted clustering systems. Given that the cluster systems are considered a good means of grouping documents, then the subjects' grouping of documents may well be a viable basis for the same task.

Each of the elements described in this section – the visual identification of groups and the analysis of their textual homogeneity – could be performed without eliciting the experiences of the subjects. However, this would not enable the effectiveness of the use of the user's organisation of documents to be discerned. Therefore, the actual use of the user's organisation of space to place later documents – the "scatter" facility – was studied. The "scatter" feature placed documents into their workspace, next to groups to which each document bore a textual similarity, §4.2.3. Seven of the ten participants rated the suggestions that they received positively. Two of the three remaining subjects had scored notably low values in the information retrieval evaluation of their workspace §6.7.4. Therefore, poor quality organisation of documents seems to have an impact upon the acceptability of the scatter facility. However, the study performed is not sufficiently large to be certain of this.

Thus, the second research question – could an information retrieval system exploit the user's organisation of documents – is partially satisfied. Users create structures that are of potential value in the context of Information Retrieval, and there is indicative evidence that this may be beneficial in its support of information seeking.

In the case of Garnet, I chose to attempt to represent the topic of each identified group of documents by building upon the techniques used in document clustering. The particular representation I chose prove to be successful in use, but alternative approaches to representing document groups exist and the question as to the advantages of each approach is worth further exploration – this will be discussed in the 'Future work' section of this chapter.

One interesting phenomenon observed in §6.7 was that when a group of documents or a workspace has a low clustering coherence score, this is indicative of low levels of topical organisation. 'Miscellaneous' piles are particularly prone to low coherence scores, but, involved in a relatively low proportion of the documents of a single workspace their effect on the overall workspace score is small. This suggests that 'miscellaneous' groups could be successfully identified using their quality score. A clustering tool could assist the user in their organisation of large miscellaneous piles of documents, as could the text matching technology behind the "scatter" facility. Such support of the information structuring task may be useful as support of the information seeking task.

To conclude, Garnet was able to identify implicit visual structures, as have previous spatial hypertexts. The user-created visual groupings of documents that the spatial parser can detect are also comparable to the groupings computed by an automatic grouping (i.e. clustering) algorithm. It may be possible to identify groups of miscellaneous documents and/or poorly organised parts of the workspace using the clustering quality measures. Where documents within a group were similar to each other, the participants approved of the documents that Garnet found to be topically similar.

7.5 Wider Contribution

The contribution of Garnet in both Spatial Hypertext and Digital Library research has been discussed in §7.1 to §7.4. This section discusses the relevance of this thesis to other fields.

In the evaluation of the potential benefits of exploiting the organisational information implicitly stored in spatial hypertexts, I have drawn on the field of clustering within Information Retrieval. Clustering quality measures have traditionally been used to choose from alternative automatically generated organisations of a set of documents [§3.8.3]. In Garnet, I used the same measures to evaluate the consistency and quality of an organisation performed by a user. This use of clustering quality measures to assess the consistency of a user's organisation of documents is another novel aspect of the evaluation of Garnet.

Systems that use human-generated classifications to order search results, such as the Categorical Search of Chi and Dumais [2000], have been demonstrated to improve information seeking performance when compared to the presentation of search results using the traditional ranked list. Similarly, researchers such as Hearst *et al* [1996] have compared computer-generated classifications as a presentation method against relevance ranking. Further study can be given to comparing each of these three established presentations (ranked list, matching to a predefined human-created classification, dynamic computer-created classification) to the presentation of search results in structures created by a user themselves. Such a comparison would have been premature before some confidence could be had in the quality of organisation performed by users and the ability of a computer system to detect the implicit structures that are commonly created by users in the course of information structuring. Now that some confidence can be placed in the ability to do this, a more direct and systematic comparison can be commenced.

In Chapter 2, the enhanced filing systems of Dourish *et al* [2000] and the Piles metaphor introduced to filing systems by Mander *et al* [1992] were noted as being relevant to information structuring activities. In Chapter 7 I briefly noted the comparison of Garnet to the structured hierarchy of classical filing systems, where users of Garnet found informal, lightweight organisation advantageous for short-term organisation. The experiment presented here gives a further datapoint to support the design intention expressed by Mander. More importantly, Garnet indicates that descriptive text could be obtained on groups of documents implicitly. In Dourish's enhanced filing systems, document descriptions are created manually. Garnet suggests that it may be possible to reduce of the amount of this entry work by exploiting document content.

7.6 Future Work

The creation and initial study of Garnet suggests that information structuring can be supportive of information seeking in an electronic environment. However, much work remains to be done to give a more detailed picture of how this occurs, and how to improve the means by which the two support each other.

In the case of the role of information structuring artefacts in information retrieval, improvements to the simple means by which Garnet performs its textual matching may be possible. Though Zamir *et al*'s [1997] algorithm is of proven effectiveness in clustering, matching through the use of machine learning techniques or alternative information retrieval techniques such as latent semantic indexing may yield better results in the case of representing the informal document groups created in a spatial hypertext.

The Spatial Parser can be extended to identify some of the idioms identified in the study – e.g. columns of document labels headed by a user label and the closely, sequential spaced columns of Subjects 8 and 10. This latter pattern is also relevant to the improvement of the suggestions system, given the problems that arose in suggestion quality §6.7.4. Another aspect that could be improved is the adaptation of the Spatial Parser to assist in placing confirmed suggestions in an appropriate place, rather than in their position when they are confirmed. At present, confirmed suggestions remain in the position in which they are initially placed after the "scatter". This may well prove incongruous with the spatial layout of the group that they match, and automatically positioning them in a manner more consistent with the group's visual structure may yield benefits in both time and consistency.

A more extensive study is required to identify the effect of any structuring activity on information seeking effectiveness, either as a result of the user's engagement in the activity, or as a consequence of the structures being used by an information retrieval system. In the present study, no work was done to identify any unconscious beneficial effects acquired through the act of organising documents. Similarly, the advantage in time or accuracy of using features such as "Scatter" were not quantified.

Methodological issues also arise. For the purposes of the current evaluation of Garnet, some measurements were gleaned from the clustering field within the information retrieval community. However, which measures for effectiveness should be used is an open question in regard to automatically generated classifications such as clustering, and has been even less addressed in regard to human-created classifications. How comparisons can be made between classification organisation of search results and the more commonplace ranked list is also an open issue.

The implementation of Garnet involved a number of design decisions, and some of these can be confirmed through the user feedback from the study of Garnet. For example, the compromise between scope for the expression of the system and the user's ownership of documents, and the representation of the digital library elements were both found to be acceptable by the experiment participants. In each case, previous research indicated that problems might arise.

However, further design decisions have to be made, and certain decisions are poorly understood. For example, the trade-off between giving the system scope for more expression and the user's scope for expression requires further study. At present documents that recur in the workspace are not identified. Doing so may be useful, but as discussed in §3.2.4, this would necessarily reduce the unambiguous range of expressions for the user to use in their own organisation. The activity of browsing the digital library within the spatial hypertext was not studied within the current evaluation of Garnet, and the particular study of that important information seeking technique is a necessity if our understanding of spatial hypertext interfaces is to be complete.

Finally, the results of the current study represent only a starting point in the development and validation of the concepts upon which Garnet is built. Now that there is some indicative evidence that the organisation of documents created by users may be useful for supporting information retrieval, a more substantial and detailed examination is justified. Improvements in information retrieval may appear in a number of different forms. Users may retrieve more documents, or retrieve a similar number more rapidly, than is the case without the support of their own structural work. Any study would have to address both possibilities. Beyond the retrieval benefits of the workspace, the role of the structuring tools themselves in supporting information seeking is worthy of further study.

Even without the use of the "scatter" facility, users' behaviours in terms of quantifiable measures, such as the number of documents that they retrieve may be altered when compared against the use of a system without a workspace on which to store their documents. Benefits other than retrieval effectiveness may be of particular interest. For instance, in evaluating Scatter/Gather Hearst *et al* [1996] investigated whether the user's conception of the material available in the collection was improved by the search system automatically structuring and organising of the documents that matched their searches. Similar techniques could be used to identify any benefit attained by the organisation of documents by users themselves in a Garnet workspace.

7.7 Summary

To conclude, I have introduced a combined spatial hypertext and digital library system, Garnet, which has demonstrated that information seeking and structuring are as interconnected in the digital environment as in the physical, and that users value the combination. The organisation of documents performed by users is generally scrutable to a simple Spatial Parser, the textual representation of which is viable input to information retrieval systems, and the resulting suggestions are frequently accepted by users. The quality of users' organisation of space can be in part evaluated through traditional clustering measurements. In addition, Garnet is of interest to Spatial Hypertext researchers as an example of integration with the information environment and of computation over hypertext.

References

Abrams, D., Baecker, R., Chignell, M., "Information Archiving with Bookmarks: Personal Web Space Construction and Organization", Proceedings of the ACM Conference on Computer-Human Interaction (CHI), ACM Press, pp. 41-48, 1998.

Allen, B., "Cognitive Differences in End User Searching of a CD-ROM Index", Proceedings of SIGIR, ACM Press, pp 298- 309, 1992.

Anick, P.G., Brennan, J.D., Flynn, R.A., Hanssen, D.R., Alvey, B., Robbins, J.M. "A Direct Manipulation Interface for Boolean Information Retrieval via Natural Language Query" Proceedings of the thirteenth international conference on Research and development in information retrieval (SIGIR), ACM Press, pp. 135 – 150, 1990.

ANSI/NISO, "Information Retrieval (Z39.50 version 3): Application Service Definition and Protocol Specification (ANSI/NISO Z39.50-2003)", NISO Press, Bethesda, MD. Available on-line at lcweb.loc.gov/z3950/agency/document.htm, 2003.

Baldonado, M.Q.W, and Winograd, T. "SenseMaker: An Information-Exploration Interface Supporting the Contextual Evolution of a User's Interests". In Proceedings of the Conference on Human Factors in Computing Systems. ACM Press, pp. 11-18, 1997.

Bates, M.J. "The design of browsing and berry-picking techniques for the on-line search interface". Online Review, 13(5), pp.407--431, 1989.

Bates, M.J. "The Design of Databases and Other Information Resources for Humanities Scholars: The Getty Online Searching Project Report No. 4." Online and CD-ROM Review 18, pp. 331-340, December 1994.

Bates, M.J. "Learning about the Informtion Seeking of Interdisciplinary Scholars and Students." Library Trends 45, pp. 155-164, Fall 1996.

Bederson, B. and Hollan, J. "Pad++: A zooming graphical interface for exploring alternate interface physics." In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '94, Marina del Rey, CA, Nov.) ACM Press, pp. 17-26, 1994.

Bederson, B.B, Hollan, J. D., Stewart, J., Rogers, D., Druin, A., Vick, D., Ring, L., Grose, E., Forsythe, C., "A Zooming Web Browser," Human Factors and Web Development, Eds.: Forsythe, C., Ratner, J., and Grose, E., Lawrence Earlbaum, pp. 255-266, 1998.

Belkin, N., Oddy, R.N, Brooks, H.M. "ASK for Information Retrieval: Parts 1&2", Journal of Documentation, 38(2,3), ACM Press, pp. 61-71; 145-164, 1982.

Belkin, N. and Croft, B. "Information filtering and information retrieval". Communications of the ACM, 35(12) pp. 29-37, 1992

Belkin, N., Marchetti, P.G., and Cool., C., "BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval". Information Processing and Management 29 (3), Elsevier. pp.325-344, 1993.

Belkin, N., "The Human Element: Helping People Find what they don't know" Communications of the ACM, 43(8) , ACM Press, pp. 58-61, August 2000.

Blandford, A., Stelmaszewska, H., Bryan-Kinns, N., "Use of Multiple Digital Libraries: a case study", Proceedings of Joint Conference on Digital Libraries (JCDL), ACM Press, pp. 179-188, 2001.

Blandford, A. & Buchanan, G. (2002) Eds. Proc. Usability for digital libraries workshop at JCDL 2002

Blandford, A., Buchanan, G. & Jones, M., Eds. "Usability of Digital Libraries" Special Issue, Journal of Digital Libraries, Springer-Verlag, 2003,in press.

Botafogo, R. and Shneiderman, B. "Identifying aggregates in hypertext structures", Proceedings of the ACM Conference on Hypertext, ACM Press, pp. 63-74, 1991.

Brin, S., Page, L., "The anatomy of a large-scale hypertextual (Web) search engine", Proceedings of the 7th World Wide Web Conference, (WWW7) also as Computer Networks and ISDN Systems, Vol. 30:1-7, pp. 107-117, 1998

Buchanan, G., Blandford, A., Jones, M. & Thimbleby, H. "Spatial hypertext as a reader tool in digital libraries," 1st International Workshop on Visual Interfaces to Digital Libraries, ACM/IEEE Joint Conference on Digital Libraries, June 2001.

Buchanan, G., Jones, M., Thimbleby, H., Farrant, S., Pazzani, M. "Improving mobile internet usability," Proceedings of the Web 2001 Conference, pp673–680. ACM Press, 2001.

Buchanan, G., Blandford, A., Jones, M., & Thimbleby, H., "Spatial Hypertext as a reader tool in Digital Libraries," Lecture Notes in Computer Science, 2539, Springer Verlag, ISBN 3-540-00247-2, pp. 13-24 (Extended from 1st Workshop on Visual Interfaces for Digital Libraries), 2002.

Buchanan, G., Jones, M. & Marsden, G., "Exploring Small Screen Digital Library Access with the Greenstone Digital Library". Proc. European Conference on Digital Libraries, Springer-Verlag, pp. 583-596, 2002,

Buchanan, G., Blandford, A., Thimbleby, H., Jones, M., "Integrating Information Seeking and Structuring in a Spatial Hypertext", Proc. ACM Hypertext 2004, in press.

Buchanan, G., Blandford, A., Thimbleby, H., Jones, M., "Supporting Information Structuring in a Digital Library", Proc. European Conference on Digital Libraries, 2004, in press.

Campagnoni, F. R. and Ehrlich, K. Information retrieval using a hypertext-based help system. In ACM Transactions on Information Systems (TOIS), Volume 7(3) pp.: 271 – 291, July 1989.

Card, S. K., Mackinlay, J. D. & Shneiderman, B. "Readings in Information Visualization – using vision to think." Morgan Kaufmann Publishers, Inc. 1999.

Carroll, J.M., Rosson, M.B., "The paradox of the active user", in "Interfacing thought: Cognitive Aspects of Human-Computer Interaction", J.M. Carroll (Ed.), MIT Press, 1987.

Catledge, L.D., Pitkow, J.E., "Characterizing Browsing Strategies in the World Wide Web", Computer Networks and ISDN Systems, 27(6), pp. 1065-1073, 1995.

Chen, C. and Czerwinski, M. , "From Latent Semantics to Spatial Hypertext -- An Integrated Approach", Proceedings of the ACM Conference on Hypertext, ACM Press, pp. 77--86, 1998.

Chen, H, Dumais, S. "Bringing order to the web: Automatically categorizing searchh results", Proceedings of the Conference on Computer-Human Interaction (CHI), Human Factors in Computing Systems, ACM Press, pp. 145-152, 2000.

Chi, E., P. Pirolli, and J. Pitkow. "The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a Web site". Proceedings of the Conference on Computer-Human Interaction (CHI), Human Factors in Computing Systems. ACM Press, pp. 161-168, 2000.

Chimera, R., Shneiderman, B. "An Exploratory Evaluation of Three Interfaces for Browsing Large Hierarchical Tables of Contents". ACM Transactions on Information Systems (TOIS) 12(4): pp. 383-4, October 1994.

Combs, T., Bederson, B., "Does Zooming Improve Image Browsing?", Proceeding of the ACM Conference on Digital Libraries, pp.130-137, 1999.

Cousins, S. B., Paepcke, A., Winograd, T., Bier, E., and Pier, K. 1997. The digital library integrated task environment (DLITE). In Proceedings of ACM Conference on Digital Libraries. ACM Press, New York, NY, pp. 142-151, 1997.

Cutting D., Karger D., Pedersen J., and Tukey, J. W. "Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections", Proceedings of the 15th Annual International ACM/SIGIR Conference, Copenhagen, 1992. pp. 318-329.

Doan, K., Plaisant, C., and Shneiderman, B. "Query previews for networked information systems". In Proceedings of the 3d Forum on Research and Technology Advances in Digital Libraries (ADL) '96 (Washington, D.C., May 13–15), IEEE Computer Science Press, Los Alamitos, Calif., 1996, pp. 120–129.

Dourish, P., Edwards, W. K., Lamarca, A., Salisbury, M., "Presto: An Experimental Architecture for Fluid Interactive Document Spaces" ACM Transactions on Computer-Human Interaction, 6(2), ACM Press, pp. 133-161, June 1999.

Dourish, P., Edwards, W. K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., Terry, D. B., Thornton, J., "Extending Document Management Systems with User Specific Active Properties" ACM Transactions on Information Systems, 18 (2), ACM Press, pp. 140-170, April 2000.

Druin, A., Bederson, B.B., Hourcade, J.P., Sherman, L., Revelle, G., Platner, M., Weng, S., "Designing a Digital Library for Young Children: An Intergenerational Partnership", Proceedings of the Joint Conference on Digital Libraries (JCDL), ACM Press, pp. 398-405, 2001.

Ellis, D., "Progress and Problems in Information Retrieval", Library Association, 1996.

Ellis, D., "Modelling the information seeking patterns of engineers and research scientists in an industrial environment" Journal of Documentation 53(4):pp.384-403, 1997.

Ericsson, K., Simon, H., "Protocol Analysis", MIT Press, Cambridge, MA, 1984.

Frank, E., Paynter, G.W., Witten, I.H., Gutwin, C., Nevill-Manning, C.G., "Domain-specific keyphrase extraction", Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, Morgan Kaufman Publishers, San Francisco, CA, pp. 668-673, 1999.

Fuhr, N., Klas, C.P., Shaefer, A., Mutschke, "Daffodil: An Integrated Desktop for Supporting High-Level Search Activities in Federated Digital Libraries", Proceedings of European Conference on Digital Libraries 2002, Springer, pp. 597-612, 2002.

Furnas. G. W., "Generalized fisheye views". In Proceedings of ACM CHI'86, (Boston, MA), pp. 16-23, April, 1986.

Furnas, G. W., and Rauch, S., "Considerations for information environments and the NaviQue workspace". Proceedings of the Third ACM Conference on Digital Libraries (DL '98), ACM Press, New York, pp. 79-88, 1998.

Giles, C.L., Bollacker, K.D., Lawrence, S., "CiteSeer: an automatic citation indexing system", Proceedings of the Third ACM Conference on Digital Libraries (DL '98), ACM Press, pp. 89-98, 1998.

Goh, D., Leggett, J., Patron-Augmented Digital Libraries, In Proceedings of the Fifth ACM Conference on Digital Libraries, (DL 2000, San Antonio, Texas) ACM Press, New York, NY, pp. 153-163.

Golovchinsky, G., "What the Query Told the Link: The Integration of Hypertext and Information Retrieval", Proceedings of the Eighth ACM Conference on Hypertext, ACM Press, pp. 67-74, 1997.

Halasz, F. R.. "Reflections on Notecards: Seven issues for the next generation of hypermedia systems". Comm. of the ACM, 31(7), ACM Press, pp. 836-852,1988.

Hancock-Beaulieu, M., and Walker, S. "An evaluation of automatic query expansion in an online library catalogue". J. of Documentation 48(4), pp. 406—421, 1992.

Hearst, M.A. "TileBars: visualization of term distribution information in full text information access". Proceedings of ACM CHI Conference '95, pp. 56-66; May, 1995

Hearst , M.A. ,Pedersen, J. O.. Reexamining the cluster hypothesis: Scatter/Gather on retrieval results. In Proceedings of the 19th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 76-84, 1996.

Hearst, M. A. and Karadi C. "Cat-a-cone: An interactive interface for specifying searches and viewing retrieval results using a large category hierarchy". In Proceedings of the 20th Annual International ACM/SIGIR Conference, pp. 246-255, 1997.

Hemmje, M., Kunkel, C., and Willett, A., "LyberWorld - a visualization user interface supporting fulltext retrieval." In Proc. of the 17th Annual International ACM SIGIR Conference, ACM Press, pp. 249-259, 1994.

Hendry, D. G., and Harper, D. J., "An informal information-seeking environment". Journal of the American Society for Information Science, 48(11), pp. 1036-1048, 1997.

Hightower, R.R., Ring, L.T., Helfman, J.I., Bederson, B.B., and Hollan, J.D. "Graphical Multiscale Web Histories: A Study of PadPrints", Proceedings of the Ninth ACM Conference on Hypertext, pp. 58-65, 1998.

Hseih, H., and Shipman, F., "VITE: A Visual Interface Supporting the Direct Manipulation of Structured Data Using Two-Way Mappings", Proceedings of ACM Conference on Intelligent User Interfaces (IUI) 2000, pp. 141-148, 2000.

Jones, M., Buchanan, G., Mohd-Naisr, N. "An Evaluation of WebTwig—a site outliner for handheld Web access,", Proc. International Symposium on Handheld and Ubiquitous Computing, pp343–348, 1999.

Jones, M., Buchanan, G. & Thimbleby, H. "Sorting Out Searching on Small Screen Devices". Proc. Mobile HCI, Springer, pp. 81-94, 2002.

Jones, S., McInnes, S., Staveley, M.S., "A Graphical User Interface for Boolean Query Specification", International Journal on Digital Libraries, Special Issue on User Interfaces for Digital Libraries, Vol. 2. 2-3, Springer, pp. 207-223, 1999.

Jones, S., Staveley, M.S., "Phrasier: a system for interactive document retrieval using keyphrases", Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp 160 – 167, 1999a.

Jones, S., Paynter, G.W., "Human Evaluation of Kea, an automatic keyphrasing system", First ACM/IEEE-CS Joint Conference on Digital Libraries, ACM Press, pp. 148-156, 2001.

Jones, W.P., Bruce, H., Dumais, S., "Keeping things found on the Web", Proceedings of the Conference on Information and Knowledge Management (CIKM), ACM Press, pp. 119-126, 2001.

Jones, W.P., Dumais, S., Croft, H., "Once Found, What Next? A study of keeping behaviours in the personal use of web information", Annual Meeting of the American Society of Information Science, 2002.

Kidd, A., "The Marks are on the Knowledge Worker", Proceedings of the ACM CHI Conference, Boston, MA, pp. 186-191, 1994.

Koenemann, J., and Belkin, N. "A Case for Interaction: A study of Interactive Information Retrieval Behavior and Effectiveness:" , Proceedings of CHI '96, Human Factors in Computing Systems, ACM Press, pp. 205–212, 1996.

Kuhlthau, C. C., "Seeking Meaning: a process approach to library and information services" Ablex Publishing, Norwood, New Jersey, 1992.

Lagoze, C. and Fielding, D. "Defining Collections in Distributed Digital Libraries", D-Lib Magazine, Vol. 4:11, Available on-line at www.dlib.org/dlib/november98/lagoze/11lagoze.html, Nov. 1998.

Lagoze, C., Van De Sompel, H., Nelson, M., Warner, S., "The Open Archives Initiative Protocol for Metadata Harvesting", Version 2, Available online at: www.openarchives.org/OAI/openarchivesprotocol.html, 2003

Lakin, F., "Visual Grammars For Visual Languages", Proceedings AAAI-87, IEEE, pp. 683-688, 1987

Lamping, J., Rao, R. : "The Hyperbolic Browser: A Focus + Context Technique for Visualizing Large Hierarchies". Journal of Visual Languages and Computing 7(1): pp. 33-55, 1996.

Larson, R. R., McDonough, J., O'Leary, P., Kuntz, L., "Cheshire II: Designing a Next-Generation Online Catalog", Journal of the Society of Information Science, 47 (7), pp 555-567, July 1996.

Lawrence, S., Bollacker, K.D., Giles, C.L., "Indexing and Retrieval of Scientific Literature", Proceedings of the Eighth International Conference on Information and Knowledge Management, ACM Press, pp. 139-146, 1999.

Leouski, A.V., Croft, W.B., "An Evaluation of Techniques for Clustering Search Results", Technical Report IR-76, Department of Computer Science, University of Massachusetts, Amherst, 1996.

Lewis, J. R. "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use". International Journal of Human-Computer Interaction, 7(1), Elsevier, pp. 57-78, 1995.

Li, J., Plaisant, C., Shneiderman, B., "Data Object and Label Placement for Information Abundant Visualizations", Proceedings Workshop on New Paradigms in Information Visualization and Manipulation (NPIV), ACM Press, pp. 41-48, 1998.

McNab, R., Witten, I., Boddie, S., "A Distributed Digital Library Architecture Incorporating Different Index Styles", Proceedings of the IEEE International Forum of Research and Technology Advances in Digital Libraries, IEEE Computer Society Press, pp. 36-45, 1998.

Malone, T. W., "How do People Organise their Desks? Implications for the Design of Office Information Systems", ACM Transactions on Information Systems, v. 1 #1, January 1983, pp. 99-112, 1983.

Manber, U., Patel, A., Robison, J., "Experience with personalization of Yahoo!", Communications of the ACM, 43(8), ACM Press, New York, pp. 35-39, 2000.

Mander, R., Salomon, G. and Wong, Y. Y., A "pile" metaphor for supporting casual organization of information, Conference proceedings on Human factors in computing systems, ACM Press pp. 627 – 634, 1992.

Marchionini, G., Dwiggins, S., Katz, A., Lin, X., "Information seeking in full-text end-user-oriented search systems: the roles of domain and search expertise", Library and Information Science Research, 15 (1), Elsevier, pp. 35-69, 1993.

Marchionini, G. "Information Seeking in Electronic Environments". Cambridge Series in Human Computer Interaction, Cambridge University Press, Cambridge, UK, 1995.

Marshall, C., Halasz, F.G., Rogers, R.., and Jansen, W.C. "Aquanet: a hypertext tool to hold your knowledge in place." Proceedings of Hypertext 91, San Antonio, Texas, (Dec.16-18, 1991), pp. 261-275.

Marshall, C. and Shipman, F. "Searching for the Missing Link: Discovering Implicit Structure in Spatial Hypertext". Proceedings of Hypertext 93, ACM Press, pp. 217-230, 1993.

Marshall, C., Shipman, F. and Coombs, J., "VIKI: Spatial Hypertext supporting emergent structure". Proceedings of the ACM European Conference on Hypermedia Technology (ECHT), ACM Press, pp. 13-23, 1994.

Marshall, C. and Shipman, F., "Spatial Hypertext and the practice of information triage". Proceedings of the Eighth ACM Conference on Hypertext, ACM Press, pp. 124-133, 1997.

Marshall, C., Price, M., Golovchinsky, G., Schilit, B., "Introducing a Digital Library Reading Appliance into a Reading Group", Proceedings of ACM Conference on Digital Libraries, ACM Press, pp. 77-84, 1999.

Marshall, C., "The Haunting Question of Intelligibility", First International Workshop on Spatial Hypertext, ACM Hypertext 2001, Aarhus, Denmark, 2001. Online at: http://www.csdl.tamu.edu/~shipman/SpatialHypertext/SH1/marshall.pdf

Marshall, C., Ruotolo, C., "Reading-in-the-small: a study of reading on small form-factor devices", Proceedings of the Joint Conference on Digital Libraries (JCDL), ACM Press, pp.56-64, 2002.

Muramatsu, J., Pratt, J., "Transparent Queries: Investigating Users' Mental Models of Search Engines", Proceedings of SIGIR, ACM Press, pp. 217-224, 2001.

Nürnberg, P.J., Leggett, J. J., Schneider, E. R., Schnase, J. L. "Hypermedia operating systems: a new paradigm for computing", Proceedings of the Seventh ACM Conference on Hypertext, ACM Press, pp. 194 – 202, 1996.

Nürnberg, P.J., Wiil, U.K., Leggett, J.K., "Structuring Facilities in Digital Libraries", Proceedings of the European Conference on Digital Libraries, LNCS 1513, Springer-Verlag, pp. 295-313, 1998.

O'Day, V, Jeffries, R.,"Orienteering in an Information Landscape: How Information Seekers Get From Here to There", Proceedings of INTERCHI, ACM Press, pp. 438-445, 1993.

Oard, D.W., and Marchionini, G., "A conceptual framework for text filtering". Technical Report CS-TR3643, University of Maryland, College Park, MD, May 1996.

Paepcke, A., Baldonado, M., Chang, C. K. , Cousins, S. and Garcia-Molina, H., "Using distributed objects to build the Stanford digital library Infobus", IEEE Computer, Vol. 32:2, pp. 80-87, Feb. 1999.

Paepcke, A., Brandriff, R., Janee, G. , Larson, R., Ludaescher, B., Melnik, S. and Raghavan, S., "Search Middleware and the Simple Digital Library Interoperability Protocol", D-Lib Magazine, Vol. 6:3, Available on-line at www.dlib.org/dlib/march00/paepcke/03paepcke.html, March 2000.

Pirolli, P., Card, S. K. "Information Foraging in Information Access Environments". Proceedings of ACM CHI, ACM Press, pp. 51-58, 1995.

Pirolli, P., "Computational Models of Information Scent-Following in a Very Large Browsable Text Collection". Proceedings ACM CHI, ACM Press, pp. 3-10, 1997.

Pirolli, P., Card, S.K., Wege, M.M.V.D.. "The effect of information scent on searching information visualizations of large tree structures". In Proceedings of Advanced Visual Interfaces, (AVI) Palermo, Italy: ACM Press. pp. 161-172, 2000

Pratt, W., Hearst, M. A., and Fagan, L. M. "A knowledge-based approach to organizing retrieval documents". In Proceedings of the 16 th National Conference on AI (AAAI '99) held at Orlando, Florida, IEEE, pp. 80-85, 1999.

Rao, R., Pedersen, J., Hearst, M., Mackinlay, J., Card, S., Masinter, L., Halvorsen, P.-K., and Robertson, G. "Rich interaction in the digital library". Commun. ACM 38, 4 (Apr. 1995), pp. 29–39, 1995.

Rauber F, Merkl, D. "The SOMLib Digital Library System". Proceedings of the 3rd Europ. Conf. on Research and Advanced Technology for Digital Libraries (ECDL'99), Paris, France, September 22. - 24. 1999, Lecture Notes in Computer Science (LNCS 1696), Springer, 1999.

Reinert, O., Bucka-Lassen, D., Pedersen, C. A., Nürnberg, P.J., "CAOS: A Collaborative and Open Spatial Structure Service Component with Incremental Spatial Parsing", ACM Hypertext Conference, ACM Press, pp. 49-50, 1999.

Salton, G. "Automatic Text Processing", Addison Wesley, 1989

Salton, G., and Buckley, C. Improving retrieval performance by relevance feedback. JASIS 41(4), pp. 288-297, 1990.

Shipman, F., Marshall, C., and Moran T. "Finding and Using Implicit Structure in Human-Organized Spatial Layouts of Information", Proceedings of Human Factors in Computing Systems (ACM CHI), ACM Press, pp. 346-353, 1995.

Shipman, F., Furuta, R.. and Marshall, C. "Generating Web-Based Presentations in Spatial Hypertext", Proceedings of the ACM Intelligent User Interfaces Conference, Orlando, Florida, pp. 71-78, 1997.

Shipman, F., "Seven Directions for Spatial Hypertext Research", First International Workshop on Spatial Hypertext, ACM Hypertext Conference 2001, Aarhus, Denmark, 2001a. Online at: http://www.csdl.tamu.edu/~shipman/SpatialHypertext/SH1/shipman.pdf

Shipman, F., Hsieh, H., Airhart, R., Maloor, P., and Moore, J.M., "The Visual Knowledge Builder: A Second Generation Spatial Hypertext", Proceedings of the ACM Conference on Hypertext, ACM Press, pp. 113-122, 2001.

Shipman, F., Moore, J.M., Maloor, P., Hsieh, Akkepeddi, R., "Semantics Happen: Knowledge Building in Spatial Hypertext", Proceedings of the ACM Conference on Hypertext, ACM Press, pp. 25-34, 2002.

Shipman, F., Price, M., Marshall, C., Golovchinsky, G., "Identifying useful passages in documents based on Annotation Patterns", Proceedings of the European Conference on Digital Libraries, Springer Verlag, pp. 101-112, 2003.

Shneiderman, B., Byrd, D, Croft W B, "Sorting out Searching: A User Interface Framework for Text Searches" Communications of the ACM 41(4), April 1998, pp.95-98, 1998.

Shneiderman, B., "Creating Creativity: User Interfaces for Supporting Innovation" ACM Transactions on Computer-Human Interaction; Vol. 7 No. 1, March 2000; pp 113-138, 2000.

Shneiderman, B., Feldman, D., Rose, A., and Grau, X.F. "Visualizing Digital Library Search Results with Categorical and Hierarchical Axes". in Proceedings of ACM Conference on Digital Libraries, ACM Press, pp. 57-65, 2000.

Suleman, H., Fox, E.A., "Designing Protocols in Support of Digital Library Componentization", Research and Advanced Technology for Digital Libraries, Proceedings 6th European Conference (ECDL 2002), LNCS 2458, Springer, pp. 568-582, 2002.

Theng, Y. L,. Duncker, E., Mohd-Nasir N., Buchanan G. & Thimbleby H. "Design Guidelines and User-Centred Digital Libraries," Third European Conference on Digital Libraries, pp167–183, 1999.

Toms, E., Kopak, R. W.; Bartlett, J., Freund, L., "Selecting Versus Describing: A Preliminary Analysis of the Efficacy of Categories in Exploring the Web", Tenth Text Retrieval Conference (TREC), National Institute of Science and Technology (NIST), 20021, pp. 653-662.

Witten, I.H., Moffat, A., Bell, T.C., "Managing Gigabytes", 2nd Edition, Morgan Kaufman, ISBN 1-55860-570-3, 1999.

Witten, I.H., Frank, E., "Data Mining", Morgan Kaufmann, ISBN 1-55860-552-5, 1999.

Witten, I., McNab, R., Boddie, S., Bainbridge, D. "Greenstone: A Comprehensive Open-Source Digital Library Software System". Proceedings of the Fifth ACM Conference on Digital Libraries, June 2000, ACM Press, pp.113-121.

Zamir, O., Etzioni, O., Mandani, O. and Karp, R. M. `Fast and Intuitive Clustering of Web Documents'. Third International Conference on Knowledge Discovery and Data Mining, August 14-17, Newport Beach, California,. AAAI Press, Menlo Park, California. pp. 287-290, 1997.

Appendices

Appendix A

Pre-study Questionnaire

1) Which of the following have you used/do you use?

Information Source	Frequency					Numb of Yea
	Daily	Weekly	Monthly	Seldom	Never	
Public Library						
Academic Library						
Internet (Web)						
Digital Libraries						
Online Journals						
Web Information sites (e.g. BBC, online newspapers)						
Other electronic information service (describe below)						

2) Have you ever used Concept Maps or Spider Diagrams?

If so, how often?

3) Have you ever used a Computer-based diagramming tool?

If so, how often?

4) How often do you use the "bookmark" or "favourites" facility of your web browser to go to web pages?

Never/Occasionally/Monthly/Weekly/Daily

5) How often do you add documents to the "bookmark" facility of your web browser?

Never/Occasionally/Monthly/Weekly/Daily

6) How often do you organise your web bookmarks into folders:

Immediately/Weekly/Monthly/Occasionally/Never

7) If you have used digital libraries, have you used the "favourites" options within a library?

8) If you have used digital libraries, have you used a facility to track topics of interest?

Appendix B – Tutorial Script

Garnet is a workspace for performing searches on digital libraries which allows you to organise and store documents of interest. We will first do some searches on an example collection to familiarise you with how the program works.

Do search on "Snail" in "Demo" collection - returns a few results.

Explain the document labels: their close buttons, how to move and resize.

Double-click on an example document to view it (try again if okay)

You can move a document from the search space onto the background window to keep it.

(repeat)

Documents can be deleted from the search list.

(check ok)

Closing a search removes it from the screen.

Its icon may still exist (in which case, one can double-click on that) - example

Closing the icon closes both

Do search on "Buffalo" in "Demo" collection - returns a few results.

Select more items in a separate group

DESCRIBE GROUPS AND WHAT THEY ARE.

Close both search windows.

(check ok)

Explain history list.

Demo double-clicking on the snail item re-opens snail search...

Close the window again

Do a third & final search on "banana"

Explain that Garnet can find documents related to those in the groups on the page

Ask if they can find any documents in the list

Perform a "scatter"

Explain the suggestions & why they are there

Explain how to confirm a suggestion

Explain how suggestions can be cleared

(check ok)

Appendix C: User Task Description

Introduction - Digital Libraries

Digital libraries are online collections of electronic documents, from which papers and books of interest can be retrieved and read. Like physical libraries, books are organised by topic, and this organisation can be used to find relevant material. However, in addition, they possess search capabilities like those found in online search engines such as Google, or within the catalogues of physical libraries like those at UCL.

The Task

The Computer Science Technical Reports (CSTR) collection contains a selection of papers produced by academic researchers up to 1997.

Digital libraries, online repositories of electronic books and documents, became an identifiable field of research around the end of this period, starting in 1995.

Your task is to collect approximately 12 to 16 documents from the CSTR collection which would form a good basis for a literature review of research papers relevant to digital libraries. Place the documents that you choose on the main workspace. You may organise and group the selected documents in any way that you find appropriate. Some relevant topics could include:

Online journals	Searching document collections	Document storage
Browsing hierarchical classifications	Reading electronic documents	Web technology and standards
Information retrieval	Using multiple libraries	Electronic publishing and copyright
Networking & communications	Systems architecture	Web-accessible catalogues and databases
Bibliographic databases	Information Visualisation	Multimedia – sound and graphics in the library

In addition to documents about Digital libraries, related research in the areas of hypertext, databases and the internet or world-wide web may also be helpful in discovering relevant papers.

A Final Task

Human-Computer interaction is an important area of research, which has relevance to digital libraries and the other fields on which digital libraries build. Discover some documents on HCI which you believe are particularly relevant to any of these areas.

Appendix D

Post Study Questionnaire

How easy did you find performing a search with Garnet?

Very Difficult/Difficult/Average/Easy/Very Easy

How did searching with Garnet compare to searching with an internet search engine:

Much harder/a little harder/Similar/Easier/Much easier

How did searching with Garnet compare to searching with a digital library:

Much harder/a little harder/Similar/Easier/Much easier

What were the best and worst elements of searching with Garnet (list up to 3)

How useful did you find being able to keep documents on the workspace

Very useful/Useful/Somewhat useful/Not useful/Obstructive

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser?

Much worse/Slightly worse/Similar/Slightly Better/Much better

How does keeping documents on the workspace compare to storing and organising documents in a filing system?

Much worse/Slightly worse/Similar/Slightly Better/Much better

What was good about keeping documents on the workspace? (list up to 3)

What was bad about keeping documents on the workspace? (list up to 3)

What was good about the document labels in Garnet?

What was bad about the document labels?

Appendix E: Pre-Study Questionnaire Results

	1	2	3	4	5	6	7	8	9	
Public Lib	Monthly	Weekly	Weekly	Seldom	Seldom	Seldom	Monthly	Seldom	Never	Seldom
	10	40	12	10	15	10	23			
Academic Lib	Monthly	Monthly	Weekly	Monthly	Monthly	Weekly	Academic	Academic	Seldom	Seldom
	4	7	3	7	12	3	4		8	
Internet	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
	10		5	6	10	5	11		10	
DLs	Weekly	Monthly	Monthly	Never	Monthly	Seldom	Seldom	Weekly	Seldom	Weekly
	2		3	5	8	<1	2		5	
Online Journals	Weekly	Monthly	Monthly	Seldom	Monthly	Seldom	Seldom	Monthly	Weekly	Weekly
	2		3	4	8	<1	1		5	
Web info sites	Daily	Occassion	Daily	Weekly	Weekly	Weekly	Daily	Weekly	Daily	Weekly
	10		5	4	5	1	11		10	
Concept Maps	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
		Once	Occassion	Once		Occassionally				Monthly
Diagramming	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
	Monthly			Weekly	Occassion	Occassion	Occassion	Occassionally		Monthl
Bookmarks	Occassion	Occassion	Daily	Occassion	Daily	Daily	Weekly	Weekly	Daily	Occass
Add Bookmarks	Never	Never	Daily	Occassion	Daily	Weekly	Monthly	Weekly	Weekly	Weekly
Organise Bookmarks	Occassion	Never	Occassion	Never	Weekly	Occassion	Immediate	Occassion	Monthly	Occass
DL Favourites	No	No	No	No	No	No	No	No	No	Yes
DL Profiles	No	No	No	No	No	No	Yes	No	No	No

Appendix F: Post-Study Questionnaire Results

Subject 1

How easy did you find performing a search with Garnet? **Easy**

How did searching with Garnet compare to searching with an internet search engine: Similar

How did searching with Garnet compare to searching with a digital library: **Easier**

What were the best and worst elements of searching with Garnet (list up to 3): **Ranking (positive)**

How useful did you find being able to keep documents on the workspace: Useful

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser? **Similar**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Similar**

What was good about keeping documents on the workspace? (list up to 3) **Finding documents again** was easy

What was bad about keeping documents on the workspace? (list up to 3)

What was good about the document labels in Garnet? **Easy to read**

What was bad about the document labels? **Perhaps more information on label?**

Was scatter useful? **Useful**

Subject 2:

How easy did you find performing a search with Garnet? **Easy**

How did searching with Garnet compare to searching with an internet search engine: Similar

How did searching with Garnet compare to searching with a digital library: **Easier**

What were the best and worst elements of searching with Garnet (list up to 3): **Ranking (positive)**

No toolbar search

How useful did you find being able to keep documents on the workspace:Very Useful

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser? **Better**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Similar**

What was good about keeping documents on the workspace? (list up to 3) **Finding documents again was easy**

What was bad about keeping documents on the workspace? (list up to 3) **Not able to rename**

What was good about the document labels in Garnet? **Easy to read, easy to remove**

What was bad about the document labels?

Was scatter useful? **Useful**

Subject 3:

How easy did you find performing a search with Garnet? **Average**

How did searching with Garnet compare to searching with an internet search engine:

A little harder

How did searching with Garnet compare to searching with a digital library: **A little harder**

What were the best and worst elements of searching with Garnet (list up to 3): **Too many windows**

How useful did you find being able to keep documents on the workspace: **Useful**

How does keeping documents on the workspace compare to using bookmarks/ favourites on a web browser? **Better (in the short term)**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Similar**

What was good about keeping documents on the workspace? (list up to 3) **Helped refine searches/track alternative terms, overview of chosen documents**

What was bad about keeping documents on the workspace? (list up to 3) **Could run out of space after a long time, forgot group topics**

What was good about the document labels in Garnet? **Easy to read**

What was bad about the document labels? **Need more information**

Was scatter useful? **Useful**

Subject 4:

How easy did you find performing a search with Garnet? **Average**

How did searching with Garnet compare to searching with an internet search engine: A little harder

How did searching with Garnet compare to searching with a digital library: **N/A**

What were the best and worst elements of searching with Garnet (list up to 3): **Easy to find top of list, history to recall searches, too many windows, want toolbar search**

How useful did you find being able to keep documents on the workspace: **Useful**

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser? **Much Better**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Different – better for short term**

What was good about keeping documents on the workspace? (list up to 3) **Helped organise work and thoughts, gave an overview of discoveries**

What was bad about keeping documents on the workspace? (list up to 3) **, Long-term organisation needs structures, in the long-term space may run out.**

What was good about the document labels in Garnet? **Easy to read**

What was bad about the document labels? **Need more information – tooltips perhaps?**

Was scatter useful? **Useful – though initially confusing.**

Subject 5:

How easy did you find performing a search with Garnet? **Easy**

How did searching with Garnet compare to searching with an internet search engine: **Similar**

How did searching with Garnet compare to searching with a digital library: **N/A**

What were the best and worst elements of searching with Garnet (list up to 3): **Having visual bookmarks**

How useful did you find being able to keep documents on the workspace **Useful**

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser? **Better – particularly in the short term**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Better – particularly in the short term**

What was good about keeping documents on the workspace? (list up to 3) **Finding documents again**

What was bad about keeping documents on the workspace? (list up to 3) **Long-term space problems?**

What was good about the document labels in Garnet? **Nice appearance**

What was bad about the document labels? **Need more information – keywords? Visible group boundaries? Topic labels (i.e. user labels) got in the way**

Was scatter useful? **No.**

Subject 6:

How easy did you find performing a search with Garnet? **Easy**

How did searching with Garnet compare to searching with an internet search engine: **A little harder**

How did searching with Garnet compare to searching with a digital library: **N/A**

What were the best and worst elements of searching with Garnet (list up to 3): **Scatter facility**

How useful did you find being able to keep documents on the workspace **Useful**

How does keeping documents on the workspace compare to using bookmarks / favourites on a web browser? **Similar overall – different benefits – good in the short term**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Better – particularly in the short term**

What was good about keeping documents on the workspace? (list up to 3) **Finding documents again was easy**

What was bad about keeping documents on the workspace? (list up to 3) **Would like hierarchical organisation, space may run out eventually.**

What was good about the document labels in Garnet? **Easy to read**

What was bad about the document labels? **Need more information. Could they be renamed?**

Was scatter useful? **Very Useful.**

Subject 7:

How easy did you find performing a search with Garnet? **Average**

How did searching with Garnet compare to searching with an internet search engine: **Similar**

How did searching with Garnet compare to searching with a digital library: **Similar**

What were the best and worst elements of searching with Garnet (list up to 3): **Deleting bad documents – but delete was too quick to see!**

How useful did you find being able to keep documents on the workspace **Useful**

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser? **Better**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Better – particularly in the short term**

What was good about keeping documents on the workspace? (list up to 3) **Checking on documents found already**

What was bad about keeping documents on the workspace? (list up to 3) **None.**

What was good about the document labels in Garnet? **Clear, simple.**

What was bad about the document labels? **Need more information. Rename/retitle them perhaps?**

Was scatter useful? **Useful.**

Subject 8:

How easy did you find performing a search with Garnet? **Easy**

How did searching with Garnet compare to searching with an internet search engine:

A little harder

How did searching with Garnet compare to searching with a digital library: **Easier**

What were the best and worst elements of searching with Garnet (list up to 3): **History of searches – a shortcut would be helpful**

How useful did you find being able to keep documents on the workspace **Useful**

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser? **Better**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Better – particularly in the short term**

What was good about keeping documents on the workspace? (list up to 3) **Reviewing what I'd tried already – seeing if I'd seen a document before.**

What was bad about keeping documents on the workspace? (list up to 3) **Concern that space would run out in the long term.**

What was good about the document labels in Garnet? **Easy to read**

What was bad about the document labels? **Need more information. Could the title be changed?**

Was scatter useful? **Partly useful.**

Subject 9:

How easy did you find performing a search with Garnet? **Easy**

How did searching with Garnet compare to searching with an internet search engine: **Similar**

How did searching with Garnet compare to searching with a digital library: **Similar**

What were the best and worst elements of searching with Garnet (list up to 3): **Meaningless titles of documents, too many windows**

How useful did you find being able to keep documents on the workspace **Useful**

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser? **Better**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Similar**

What was good about keeping documents on the workspace? (list up to 3) **Overview of progress and topics, finding documents already seen**

What was bad about keeping documents on the workspace? (list up to 3) **,None.**

What was good about the document labels in Garnet? **Simple.**

What was bad about the document labels? **More information – rename perhaps?**

Was scatter useful? **Very Useful.**

Subject 10:

How easy did you find performing a search with Garnet? **Easy**

How did searching with Garnet compare to searching with an internet search engine:

Similar

How did searching with Garnet compare to searching with a digital library: **Easier**

What were the best and worst elements of searching with Garnet (list up to 3): **Very visual (good) – too many windows.**

How useful did you find being able to keep documents on the workspace

Useful

How does keeping documents on the workspace compare to using bookmarks/favourites on a web browser? **Better**

How does keeping documents on the workspace compare to storing and organising documents in a filing system? **Better – particularly in the short term**

What was good about keeping documents on the workspace? (list up to 3) **Track work done**

What was bad about keeping documents on the workspace? (list up to 3) , **Space shortage eventually?**

What was good about the document labels in Garnet? **Clarity**

What was bad about the document labels? **Need more information – rename bad document titles?**

Was scatter useful? **Partially useful.**

Appendix G: Subject Clustering Scores

Word	Subject										Subjects	W…
	1	2	3	4	5	6	7	8	9	10	(Total)	Ra…
Library	1	1	1	3	1	5	1	1	1	2	10	
Digital	2	2	2	5	2	6	2	2	5	3	10	
Information	3	5	3			1	3	8	3	1	8	
User	5	3	4	1	8	4	4	12	4	4	10	
System	4		7		3	2	6	5	7	7	8	
Image		6					8	3	8		4	
Interface	6	7	5	2	11		7		12	5	8	
Retrieval		4	9	20	7	3	5	4	2	16	9	[illegible]
Data	8	20	8	6		17	10	6	6	6	9	[illegible]
Searching		10	6				16		9	20	5	

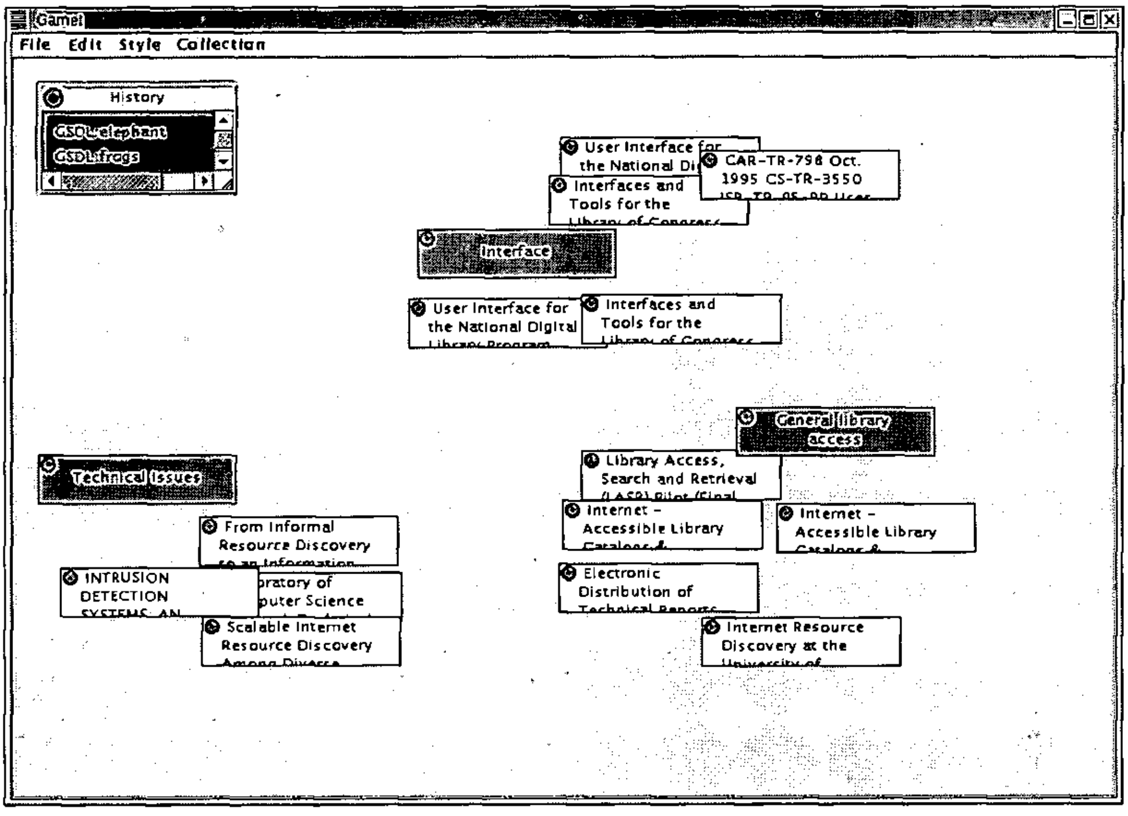
Frequency rankings for terms found in subject's workspaces – only ranks within the first twenty most frequent terms in each workspace are given.

Score/Rank	***Subject***									
	1	2	3	4	5	6	7	8	9	10
GQF (Zamir and Etzioni)										
Nominal GQF	63	56	57	36	17	46	43	40.5	55	39
Nominal Rank	1	3	2	9	10	5	6	7	4	9
GQF Score	47	66	90	65	6	89	42	38	108	0
Finel Score	22.70316	34.78505	40.24922	28.14583	4.242641	38.61362	23.29741	26.87006	42.77437	0
GQF Rank	8	4	2	5	9	3	7	6	1	10
Improved GQF	107	122	194	187	76	184	73	62	240	10
Improved Final	51.68591	64.29965	86.75944	80.97338	53.74012	79.83041	40.49311	43.84062	95.05416	10
Improved Rank	7	5	2	3	6	4	9	8	1	10
Factor (GQF/Nominal)	0.360368	0.621162	0.706127	0.781828	0.249567	0.839427	0.5418	0.663458	0.777716	0
Factor Rank	8	6	4	2	9	1	7	5	3	10
Coherence (Pederson et al)										
Summed Coherence	1.49365	1.157033	1.620183	1.550583	0.5496	1.7798	0.877867	1.317467	1.798917	0.941417
GQF'd Coherence	1.29406	1.119025	1.466905	1.361089	0.777252	1.485193	0.766274	0.903612	1.614178	0.5135
Wghtd Coh Rank	5	7	3	4	10	2	9	6	1	8
GQF'd Coh Rank	5	6	3	4	8	2	9	7	1	10
Ranks										
Best Rank	5	4	2	2	8	1	7	5	1	8
Worst Rank	8	7	4	5	10	4	9	8	3	10
Average Renk	6.6	5.6	2.8	3.6	8.4	2.4	8.2	6.4	1.4	9.6
Deviation	1.516575	1.140175	0.83666	1.140175	1.516575	1.140175	1.095445	1.140175	0.894427	0.894427
Wilcoxon Renk	7	5	3	4	9	2	8	6	1	10

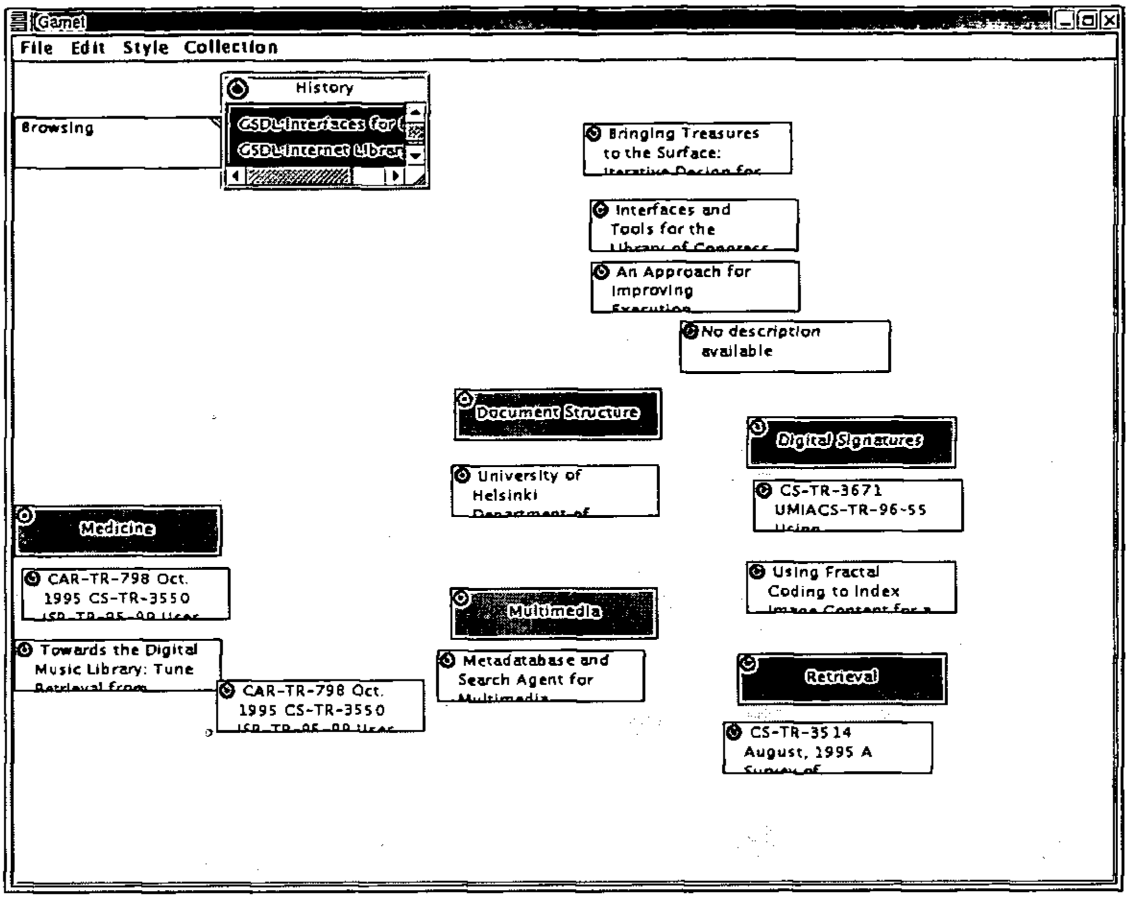
Sample scores for ranking subject's workspaces under a clustering algorithm

Appendix H: Subject Workspaces

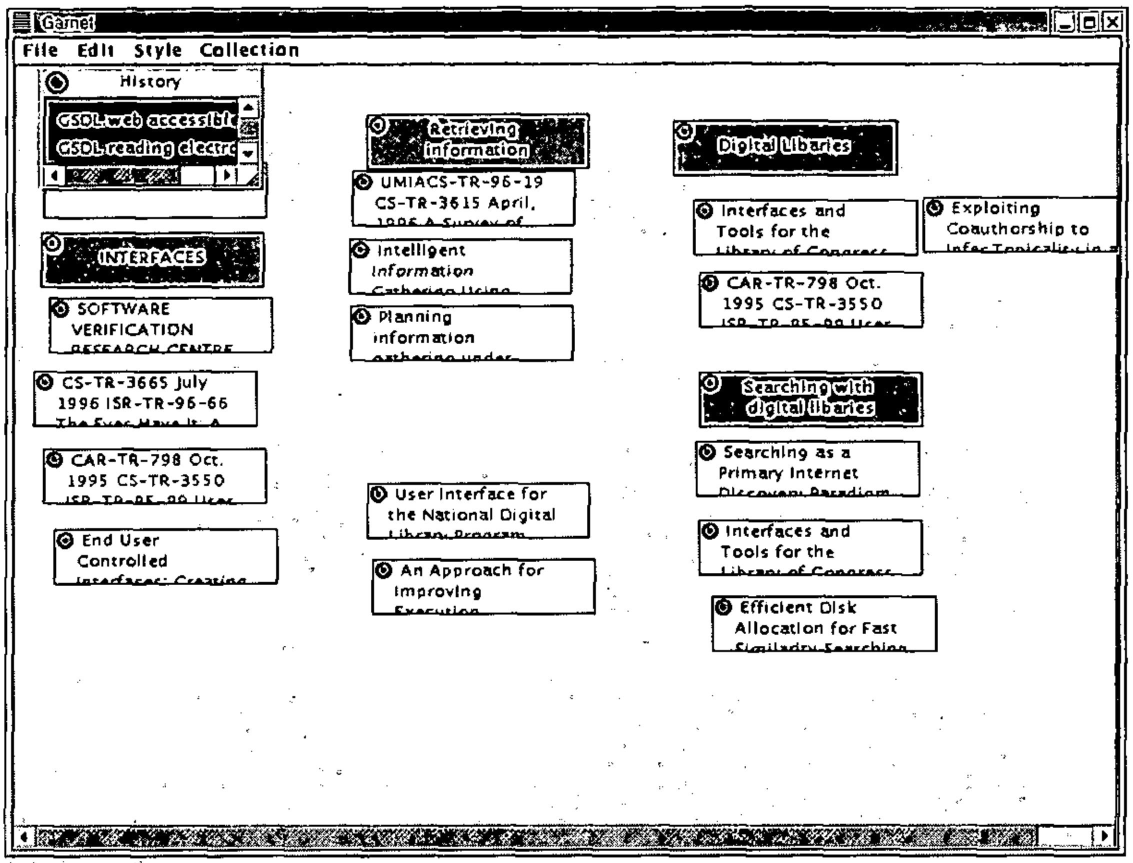
Subject 1



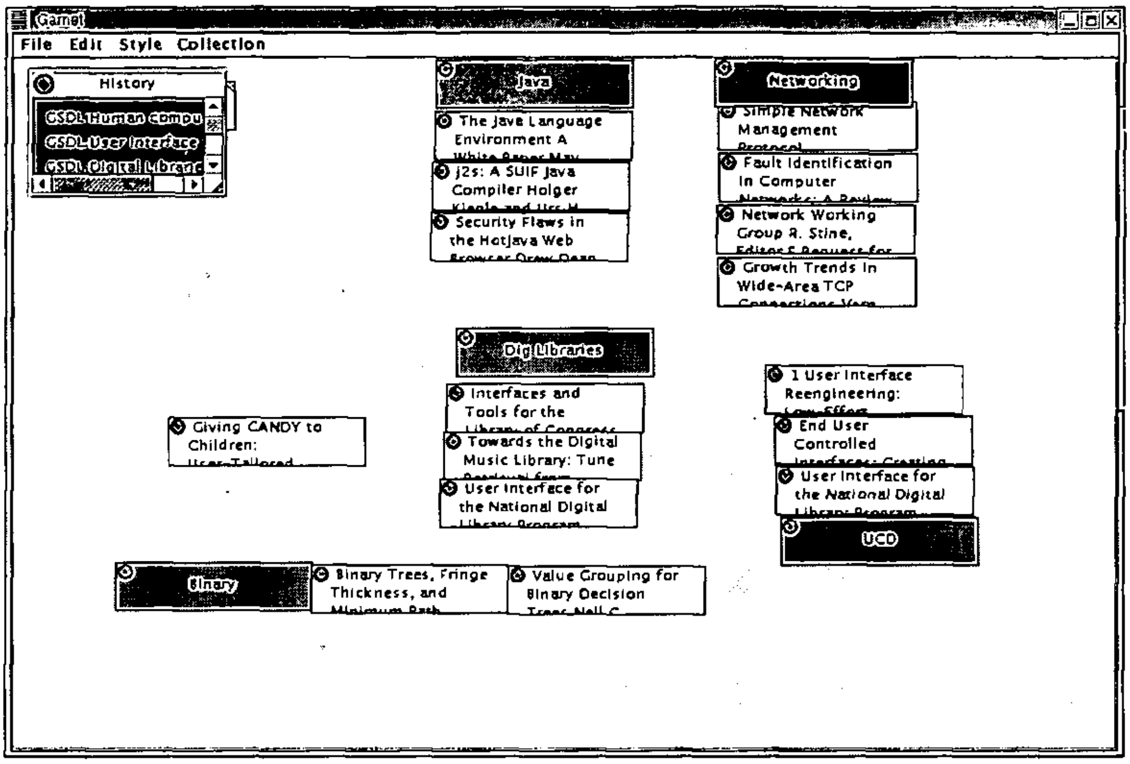
Subject 2


Garnet
File Edit Style Collection
History
GSDL:Interfaces for
GSDL:Internet Librar
Browsing
Bringing Treasures to the Surface:
Interfaces and Tools for the
An Approach for Improving
No description available
Document Structure
Digital Signatures
University of Helsinki
CS-TR-3671 UMIACS-TR-96-55
Medicine
CAR-TR-798 Oct. 1995 CS-TR-3550
Using Fractal Coding to Index
Multimedia
Towards the Digital Music Library: Tune
CAR-TR-798 Oct. 1995 CS-TR-3550
Metadatabase and Search Agent for
Retrieval
CS-TR-3514 August, 1995 A

Subject 3

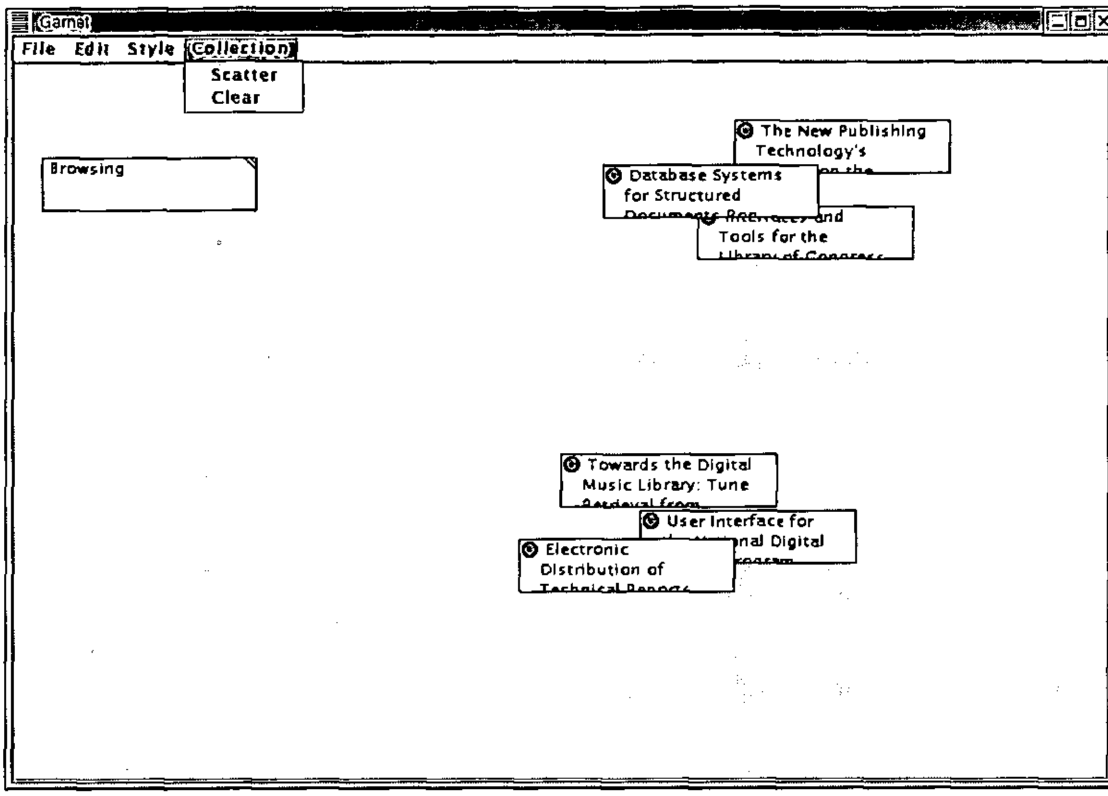

Garnet
File Edit Style Collection
History
Retrieving information
Digital Libaries
UMIACS-TR-96-19 CS-TR-3615 April,
Interfaces and Tools for the
Exploiting Coauthorship to
INTERFACES
Intelligent Information
CAR-TR-798 Oct. 1995 CS-TR-3550
SOFTWARE VERIFICATION
Planning information
CS-TR-3665 July 1996 ISR-TR-96-66
Searching with digital libaries
Searching as a Primary Internet
CAR-TR-798 Oct. 1995 CS-TR-3550
User Interface for the National Digital
Interfaces and Tools for the
End User Controlled
An Approach for Improving
Efficient Disk Allocation for Fast

Subject 4

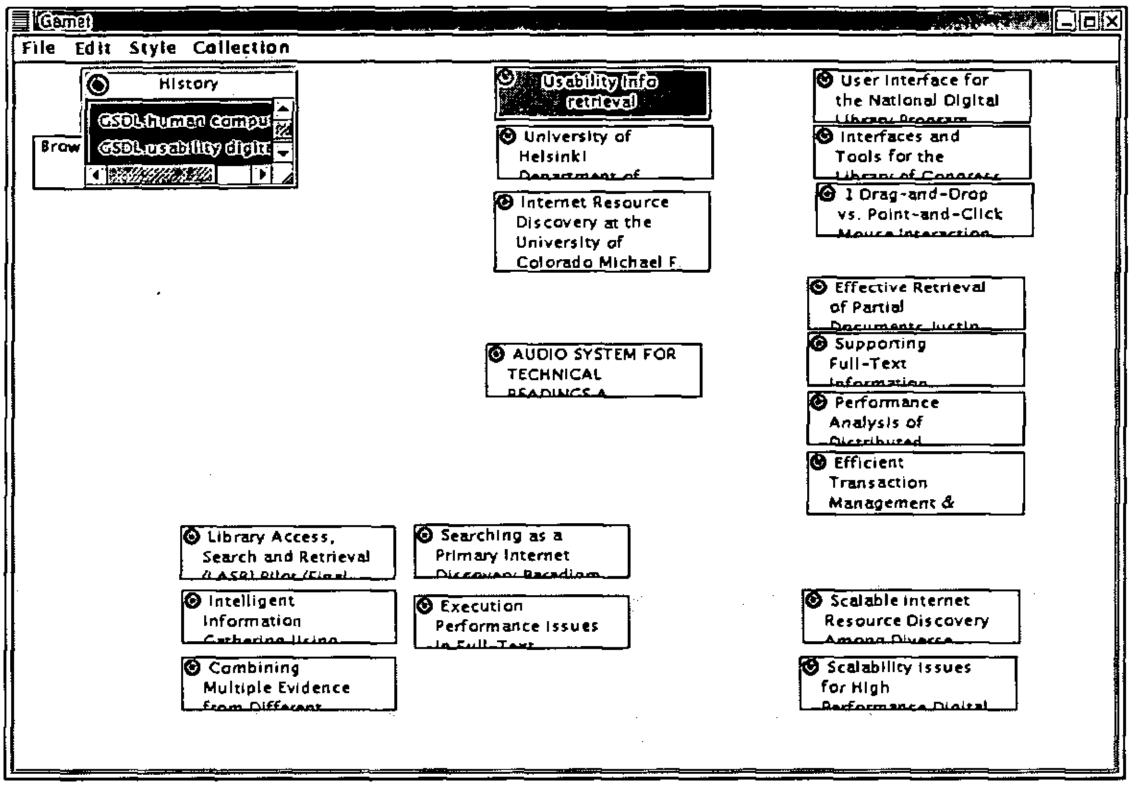


Garnet
File Edit Style Collection
History
Java
The Jave Language Environment A
J2s: A SUIF Java Compiler Holger
Security Flaws in the HotJava Web
Networking
Management
Fault Identification in Computer
Network Working Group R. Stine,
Growth Trends in Wide-Area TCP
Dig Libraries
Interfaces and Tools for the
Towards the Digital Music Library: Tune
User Interface for the National Digital
Giving CANDY to Children:
1 User Interface Reengineering:
End User Controlled
User Interface for the National Digital
UCD
Binary
Binary Trees, Fringe Thickness, and
Value Grouping for Binary Decision

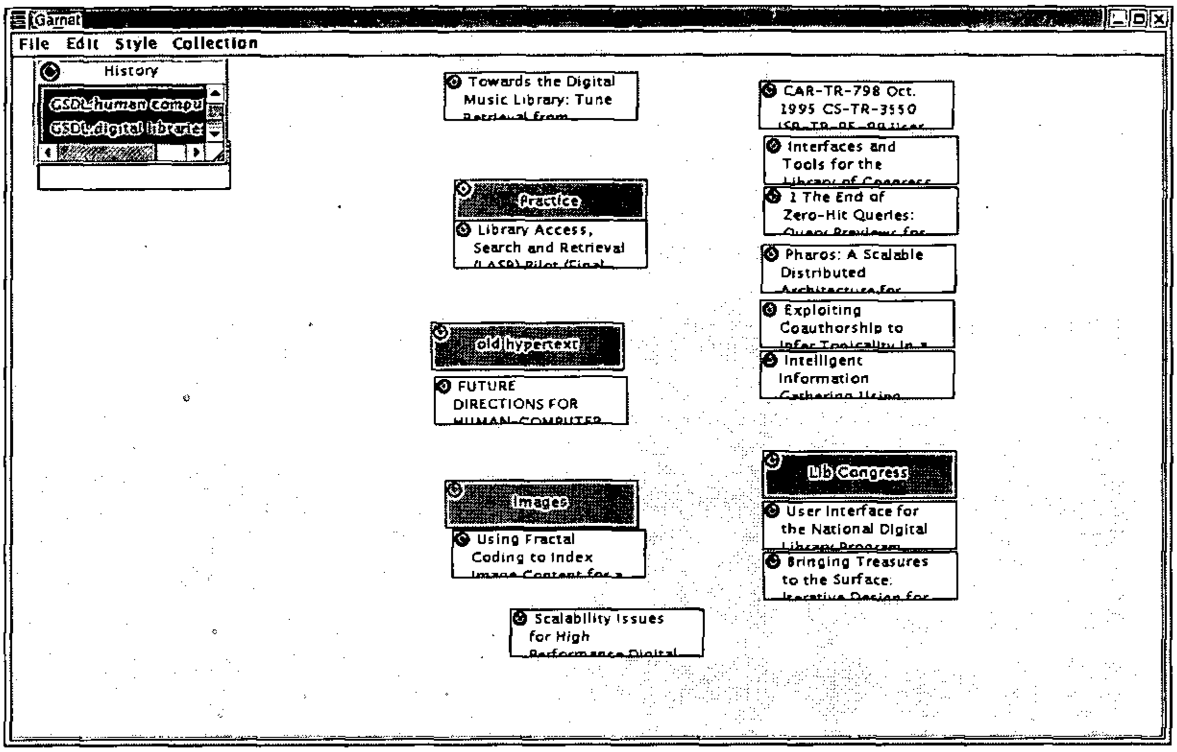
Subject 5


Garnet
File Edit Style Collection
Scatter
Clear
Browsing
The New Publishing Technology's
Database Systems for Structured
Tools for the
Towards the Digital Music Library: Tune
User Interface for
Electronic Distribution of

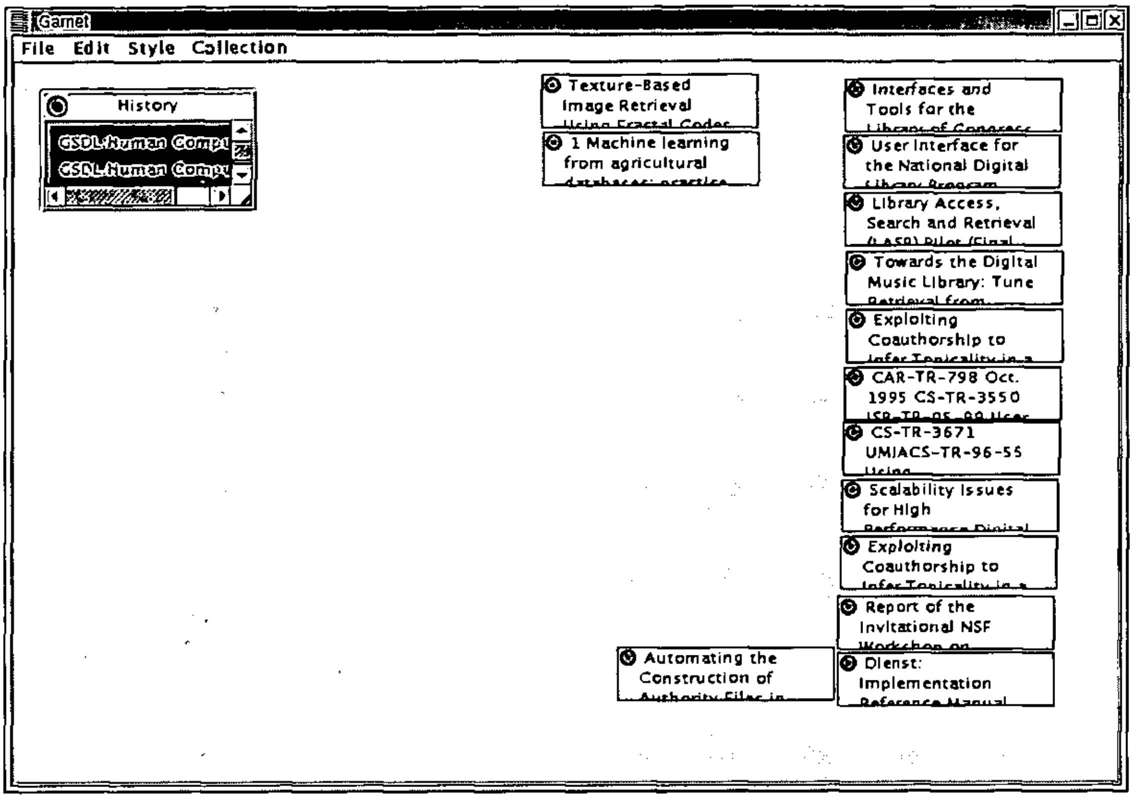
Subject 6


Gamet
File Edit Style Collection
History
GSDL human compu
GSDL usability digit
Brow
Usability info retrieval
University of Helsinki
Internet Resource Discovery at the University of Colorado Michael F.
AUDIO SYSTEM FOR TECHNICAL
User Interface for the National Digital
Interfaces and Tools for the
I Drag-and-Drop vs. Point-and-Click
Effective Retrieval of Partial
Supporting Full-Text
Performance Analysis of
Efficient Transaction Management &
Library Access, Search and Retrieval
Searching as a Primary Internet
Intelligent Information
Execution Performance Issues
Combining Multiple Evidence
Scalable Internet Resource Discovery
Scalability Issues for High

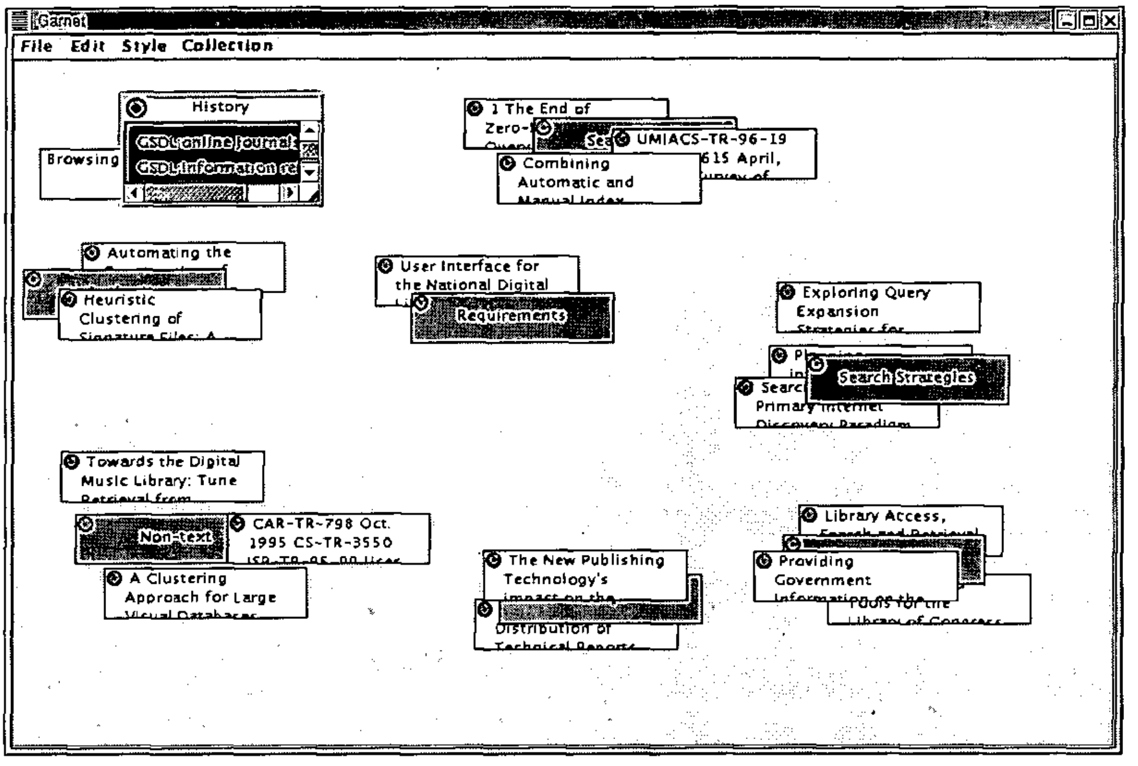
Subject 7


Garnet
File Edit Style Collection
History
Towards the Digital Music Library: Tune
CAR-TR-798 Oct. 1995 CS-TR-3550
Interfaces and Tools for the
1 The End of Zero-Hit Queries:
Practice
Library Access, Search and Retrieval
Pharos: A Scalable Distributed
Exploiting Coauthorship to
Intelligent Information
old hypertext
FUTURE DIRECTIONS FOR
Lib Congress
User Interface for the National Digital
Images
Using Fractal Coding to Index
Bringing Treasures to the Surface:
Scalability Issues for High

Subject 8

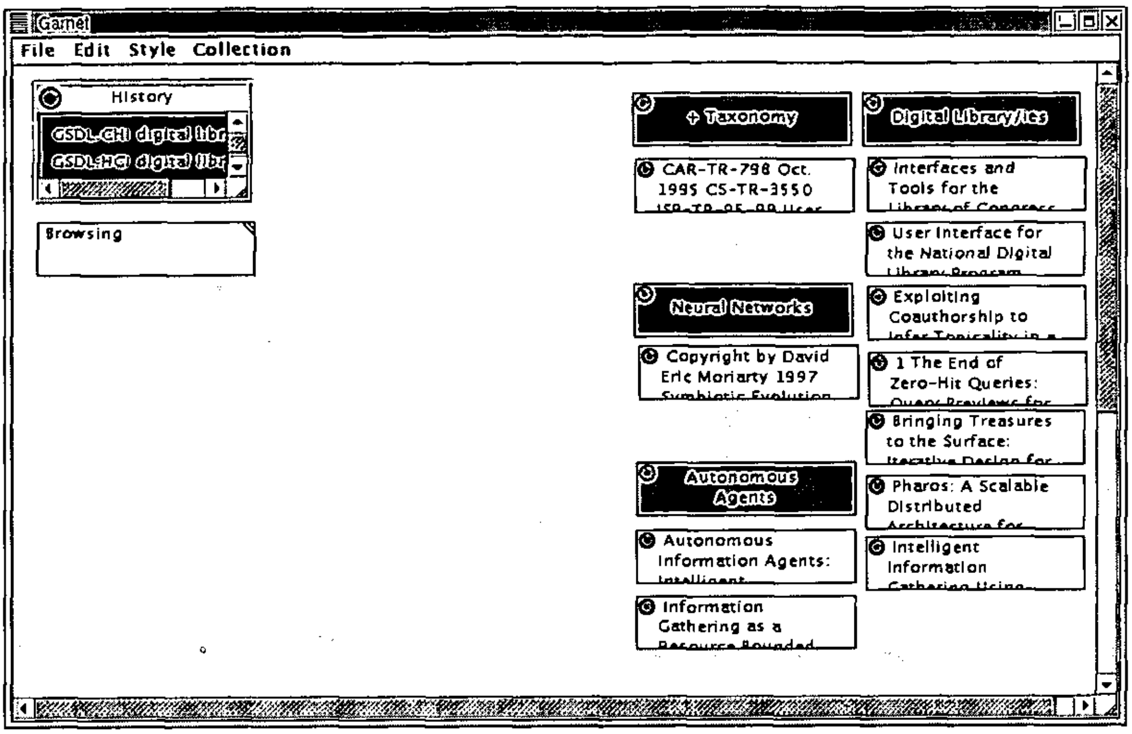

Garnet
File Edit Style Collection
History
Texture-Based Image Retrieval
1 Machine learning from agricultural
Interfaces and Tools for the
User Interface for the National Digital
Library Access, Search and Retrieval
Towards the Digital Music Library: Tune
Exploiting Coauthorship to
CAR-TR-798 Oct. 1995 CS-TR-3550
CS-TR-3671 UMIACS-TR-96-55
Scalability Issues for High
Exploiting Coauthorship to
Report of the Invitational NSF
Automating the Construction of
Dienst: Implementation

Subject 9



Garnet
File Edit Style Collection
History
Browsing
Automating the
Heuristic Clustering of
User Interface for the National Digital
Requirements
1 The End of Zero-
UMIACS-TR-96-19
615 April,
Combining Automatic and
Exploring Query Expansion
Search Strategies
Primary Internet
Towards the Digital Music Library: Tune
Non-text
CAR-TR-798 Oct. 1995 CS-TR-3550
A Clustering Approach for Large
The New Publishing Technology's
Distribution of
Library Access,
Providing Government

Subject 10


Garnet
File Edit Style Collection
History
GSDL:CHI digital libr
GSDL:HCI digital libr
Browsing
Taxonomy
Digital Library/ies
CAR-TR-798 Oct. 1995 CS-TR-3550
Interfaces and Tools for the
User Interface for the National Digital
Neural Networks
Exploiting Coauthorship to
Copyright by David Eric Moriarty 1997
1 The End of Zero-Hit Queries:
Bringing Treasures to the Surface:
Autonomous Agents
Pharos: A Scalable Distributed
Autonomous Information Agents:
Intelligent Information
Information Gathering as a