COLLABORATIVE BROWSING AND VISUALISATION OF THE SEARCH PROCESS

Michael B. Twidale and David M. Nichols

Computing Department, Lancaster University, LA1 4YR, UK.

Introduction

The term browsing is frequently applied to information searching activities although it is has been defined in many different ways[4]. In this paper we highlight the social and collaborative aspects of browsing and discuss how they may be assisted by explicit computerised support. Specifically, we describe the Ariadne system – an interface which visualises the search process as a manipulable digital object. Visualisation of the process offers a range of activities that can support more effective searching activities by individuals and through collaboration with others.

The need for collaboration

It is likely that the exponential growth of available online information will continue and that a growing proportion of existing information in traditional libraries will become available electronically. From the general history of computer systems, we should expect that the information access mechanisms will continue to evolve. Therefore for these twin reasons of information overload and perpetual systems evolution that we propose the consideration of the techniques of collaboration and visualisation particularly as they can be applied to an explicit interaction with a representation of the search process. There are two interrelated areas which collaboration can support: learning and working.

By learning we mean:

- learning both generic information searching and browsing skills
- techniques and activities specific to a particular system
- issues relating to the domain being searched

By working we mean:

- jointly searching with other people
- sharing hints and tips
- asking for advice on particular problems

Existing collaboration

The use of library resources is often stereotyped as a solitary activity, with hardly any mention in the substantial library science and information retrieval literature of the social aspects of information systems. Other than the traditional reference interview which has been extensively studied, there seems to have been very little consideration of issues of collaboration either with paper based or computerised systems.

However, previous library computerisation makes all the other people invisible. If anything, computer scientists have been even more determined to ignore the 'problem' of collaboration. We have designed multi-user databases (such as library catalogues) to appear to their users as much as possible as single-user systems.

We are arguing that introducing support for collaboration into information retrieval systems would help users to learn and use the systems more effectively. However it should be noted that people are already using existing systems collaboratively *despite* the systems failing to support this. We have performed informal observations[25] in Lancaster University Library which indicate significant informal computer-based collaboration between users. They share a terminal or lean over to look at an adjacent terminal and discuss and point at screens. Many students learn how to use the system not from attending introductory talks, reading leaflets or using the help facilities on the system, but from their friends. This approach, at least at Lancaster, appears to be very effective with a

high degree of contentment with the system, even from users who have little or no experience with computer systems (such as word processors). For our investigations, we had great difficulty in finding people afraid or unwilling to use the OPACs.

Spontaneous collaborative learning is effective in introducing novices to the basics of using an OPAC. However, we found that many users only learned a minimal subset of the available facilities (such as almost exclusively performing author-title searches). This is a sensible learning strategy for novices. Unfortunately many seem to settle into a feeling of competence with little desire to improve by either formal or informal learning. They may acquire sub-optimal working habits such as failure to refine a search, contenting themselves with reading hundreds of hits or never investigating how to use more sophisticated searching options. Such habits can be addressed through collaboration with other learners, working with librarians or computerised help. We believe that systems that actively support collaboration should also support this incremental learning, by providing facilities to illustrate and explain browsing techniques.

Social and Collaborative Browsing

Information retrieval systems have been largely designed to give the impression of being single user systems – the existence and activities of other users have been hidden from each other. Library catalogues often contain little recognition of users' activities beyond noting that some anonymous other person has borrowed the book you were after.

In a computerised environment the searches of users can be easily recorded and re-used. There are several mechanisms for using this information to improve future searching efficiency[27]:

- collaborative filtering[2,6,18,21]
- social filtering/recommending[10,16,24]
- enhancing indexing mechanisms and browsing structures[11,14,19,30]

The common feature of all these approaches is that one user's activity can have an influence on future users – search information can be re-used. Instead of regarding the library catalogue as a static structure and searches as private the library system grows with usage and becomes a repository of social information[12,13]. The ownership and distribution of such information raises many interesting issues of privacy[9].

A broad distinction can be drawn between those systems which automatically gather information about past usage[8,19,24] and those which require explicit additions from users[11,17]. The presence of a (perceivable) cost from the user's point of view is an important factor in the success or failure of cooperative systems[7]. When there is a cost to adding information to a database (e.g. links[11], annotations[6,23] or ratings[2,17,21], then users will tend to expect a clear benefit. Although explicit ratings by users may be expected to be more useful in assessing the value of information inferred ratings may be more effective as the perceived cost may be zero[27].

The inferred ratings approach is particularly suited to applications with large numbers of users – the volume of interaction allows small incremental additions to produce significant results: this concept has been termed *cyberspace leveraging*[29]. A related concept is that of the *history enriched digital object*[9] – structures that contain information about their own history.

Digital libraries, with a large user base and the facilities to record their history, are able to provide social browsing structures where the activities of previous users have enhanced their catalogues[12]. Although the approaches referred to above aim at long term improvement of information structures, the preservation of searches can also be used to support direct human collaboration about information searching activities.

In computer supported cooperative work (CSCW) a common distinction is made between the temporal and spatial nature of activities. Activities are either co-located or remote and either synchronous or asynchronous[22]. Figure 1 shows library activities on these dimensions: the approaches referred to above are largely contained in the term *social information filtering* in the remote-asynchronous quadrant.

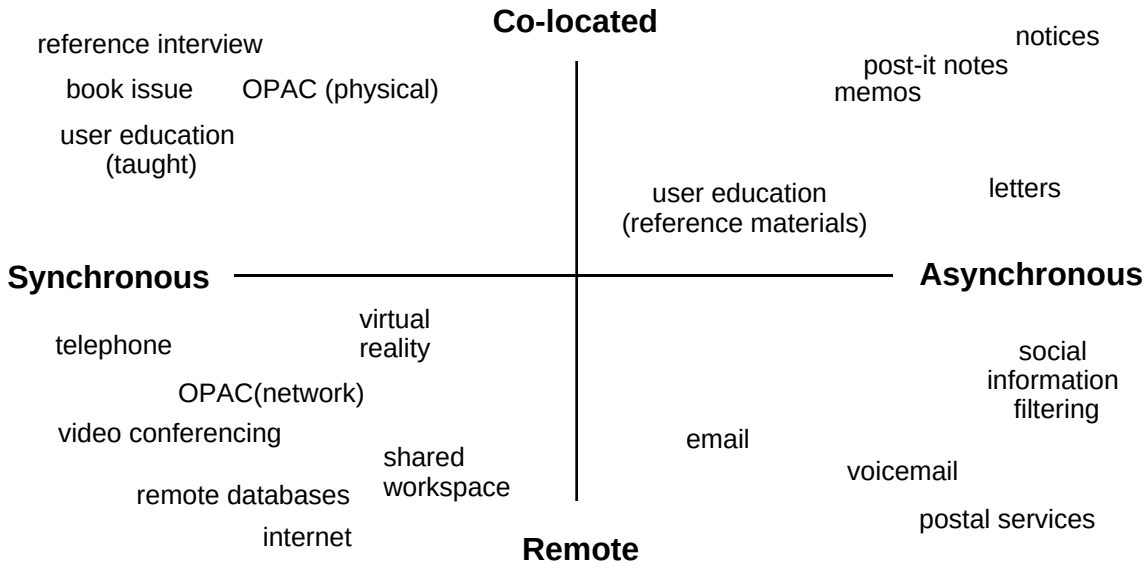

Figure 1 Library activities on spatial and temporal dimensions

In conventional libraries collaborations are usually co-located and synchronous. In digital libraries, as the following sections show, we believe that most interactions will be remote and asynchronous. Staff at a remote and asynchronous reference desk will not be able assist users without explicit computerised support. To investigate these differences we observed users both in Lancaster University Library and connecting remotely. It is our view that understanding collaboration 'as it is' will be helpful in understanding collaboration 'as it may be' in future digital libraries.

Preliminary study

A preliminary study[26] was made observing remote collaboration using existing simple communication tools in order to reveal the problems that the system to be developed should address. For this study we supported synchronous remote collaboration between an end user and an information expert. The user performed a sequence of searches and asked the expert for advice. Thus we had chosen to replicate the reference interview but using technology to permit remote collaboration. Participants communicated by typing into windows and using a simple window replication program to allow them to see what the other was doing when undertaking bibliographic searches. A key problem was the misinterpretation of collaborators' written messages to each other. For example, the expert would use a technical information systems term that was misinterpreted by the user. It was clear that what was needed a means of supporting discussions between people who might be using very different vocabularies.

In retrospect our choice of synchronous-remote collaborations was inappropriate. The difficulties of coordinating a time for synchronous working and the heavy workloads of university library staff imply that remote asynchronous collaboration is a more suitable approach. They also imply that we should build systems that help end-users to seek help from each other in addition to (or before) seeking help from an expert.

Product and Process

Most examples of visualisation in information retrieval concentrate on the document records - both before and after a query has been applied[20]. We can categorise them as addressing the problem of visualising the search product. Some of the virtual reality approaches allow a form of synchronous collaborative browsing by including embodiments of users in the document space[1,3]. Whilst these areas are important there is another candidate for visualisation - the search *process.* A representation of the process of searching can be a valuable aid to recognising and recovering from errors – both for yourself and for helping others.

Once a search has been represented as a digital object it can be manipulated: edited, replayed, annotated, re-executed and shared. This representation can become the focus of asynchronous collaborative activities between remote users. A representation of the search process could simply be

a textual listing of actions taken by a searcher. However, graphical representations have shown to be useful in abstracting features of problem-solving processes[5]. Lin *et al*[15] use such a visualisation technique, as researchers, to examine the behaviour of searchers. We believe that such techniques should be made explicit to the users so they can benefit directly – rather than indirectly via incremental system refinements.

Collaborative working implies a need to share information: both the end *product* (the 'hits') and the *process* (the search strategy/tactics). Similarly there is a need to share this information with the librarians, for whom inspection of the search process can reveal not only gaps in the user's browsing techniques but also an indication of their degree of searching sophistication. In addition an externalised representation of the search process reduces cognitive load and facilitates reflection – a vital component of learning.

Ariadne: Visualising the search process

We have developed an interface, the Ariadne system, that supports collaboration by promoting the awareness of the activities of others, better visualisation of the information data structures being browsed and more effective means of communicating the browsing process. The system captures the users' input and the database's output and forms them into a search history consisting of a series of *command–output* pairs. This means data capture is done transparently: so that users can work as if they were interacting directly with their chosen database. It is only in the subsequent playback phase that the new form of working becomes available.

The simplicity of the approach ensures that it can be used for *any* text based interface for *any* library for which remote access via Telnet is possible. This separation between capture and display is necessary because of the lack of separation between the user interface of a database and the database functionality itself. Therefore, the separation we provide allows the graphical Ariadne interface to work both with data captured in this way and (potentially) through other methods (e.g. by the Z39.50 protocol). Although the current version of the system only produces the visualisation after a search has been completed, we intend to extend it to allow the history to be dynamically created and presented in real time.

Ariadne bridges the informal distinction referred to earlier (automatic gathering *v* explicit additions) in that the information is captured at zero perceived cost to the user but is then made available to users who can add value (e.g. an annotation) and communicate it others.

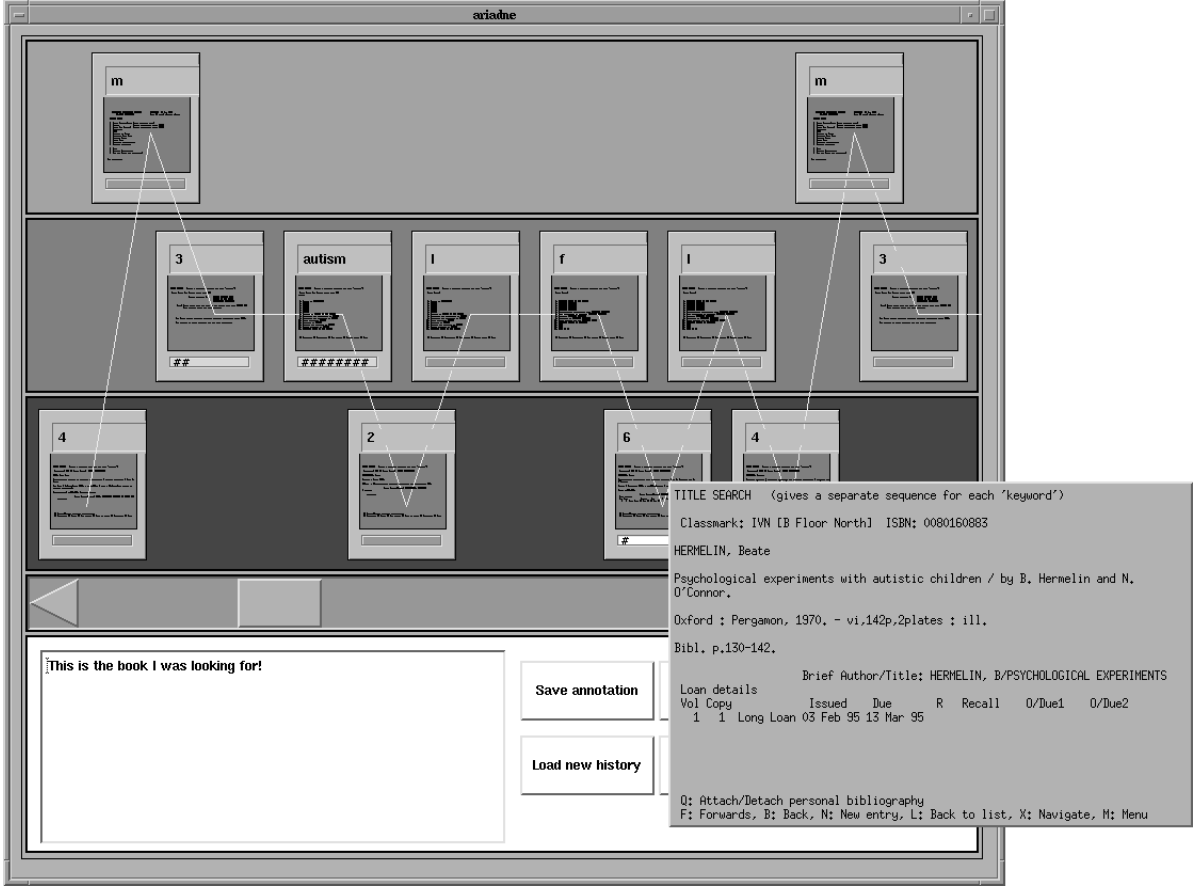

Figure 2 The Ariadne interface

Figure 2 illustrates the system. The visualisation of the search process consists of a sequence of the command-output pairs looking like small playing cards and containing thumbnails of the screendump of the outputs. These can be expanded to full size by clicking on them. The sequence of cards through time progresses horizontally while the vertical position of the card on one of three rows gives information of the semantics of the action it represents: choices from top level menus on the top row, search commands in the middle row and looking at particular book details on the bottom row. The aim is to give an impression of 'diving' into a database by composing queries and going down to actual data entries. A session consists of numerous 'dives' into detail, interspersed by 'higher level' activities of composing and combining searches, selecting display options etc. This visualisation makes it easy to pick out certain characteristic patterns of behaviour such as reading through large numbers of hits rather than redefining a query. This externalisation of the search process into a digital object allows it to be annotated, discussed with colleagues around the screen and distributed to remote collaborators for asynchronous commenting.

Studies with Ariadne

Testing of the system[28] has involved the use of volunteers who bring a problem that they already have to solve, rather than our imposing a standardised problem upon them. We believe that this more authentic form of testing avoids the dangers of implicit assumptions about envisaged use being embodied both in the system design and in the evaluation task. Students from a wide range of academic backgrounds (including Psychology, Computing, Women's Studies, Chemistry, Religious studies and Environmental Science) have used the system. The typical case is that they are about to write an extended essay, dissertation or group project and need to do a literature search. The testing informed the iterative development of the system. The results were promising, not least that complete novices were able to understand the concept of the visualisation and found it useful in informing their understanding of the search process, and in discussing this understanding with colleagues. The visualisation reduces the potential for misinterpretation and makes it easier to check for remaining misinterpretations and to remediate them through pointing and discussion. In particular novices can discuss searching strategies with experts even when they lack the appropriate information retrieval

vocabulary by pointing to examples from the history.

Issues of privacy and ownership

Recording the interaction history of the user of an information system raises a number of issues relating to privacy and ownership[9]. If the interaction history is only recorded by the active choice of the user and if they only use the history themselves, or share it with a small group of known individuals who do not forward it outside the group, then there is little problem. However the potential of recording the histories of users as outlined above may lead to other patterns of usage.

It may be decided to routinely record all search activities of all users. Given the rapidly falling costs of computer memory and storage this is becoming a feasible option. It would offer many advantages including permitting more detailed analysis of usage patterns in order to identify problems users were having with the existing system, which features were popular and which approaches were most productive. In addition, other users could search these histories to support their own searching.

Information professionals may submit a search history (process) along with the results of their searching (product) as part of the justification of their fee. This may be particularly appropriate when part of the information they are providing is the lack of information available on the subject in question and they wish to indicate that their negative results are not from want of trying. By contrast, the same professional may be quite adamant that she definitely does not wish to give away the tricks of her trade. Likewise, information professionals in a cooperative environment may be willing to share histories within their group but be unwilling to see them spread out any further. We can expect to see levels of access permissions across different groups, perhaps analogous to the levels of permissions to conventional computer files.

Librarians have a long tradition of concern for the privacy of their users. Collaboration usually involves some reduction of this privacy. When the collaboration is with known individuals then the users can freely choose the degree to which they make personal activity information available. However an electronic system offers the opportunity (should people want it) of collaboration with strangers. This may be by the system identifying similar research interests and offering to introduce the participants to each other, leading to a conventional form of collaboration, or it may involve permanent anonymity where the actions of others are used to inform one's own searching behaviour. Where traces are aggregated they become more anonymous even though their utility to future users remains.

The problem arises of who owns a search history and whether the person who created it should have rights over how it is used in the future. Even if a history is anonymised by deleting any link to its author, the pattern of interests revealed in a complete search may well be sufficient to uniquely identify the person. For example, a search history that involved the concepts of visualisation, collaboration, browsing and digital libraries may well be sufficient to identify it as performed by the authors of this paper. Breaking up a search session into smaller parts renders identification less easy but may greatly reduce the usefulness of the resulting histories. Note that anonymity of the kind outlined in this paragraph would preclude the potential advantages of the introduction service outlined in the previous paragraph, illustrating the tradeoff between privacy and the potential benefits from collaboration.

It would seem that to be acceptable to users, a system supporting collaboration should make clear to users the benefits that accrue from their loss of privacy along with precise details about the way in which information about their search activities are, and may be, used. Users should also be provided with substantial control over the information about themselves that is stored, including options to veto the storing of certain searches or to anonymise their submissions.

Conclusions

Browsing computerised information resources has a social and collaborative dimension which will be increasingly remote and asynchronous. The storage and re-use of the search process provides a mechanism to support a variety of activities which users may wish to undertake. The visualisation of a search process can be a useful means to abstract information and aid collaboration between information workers. In addition, collaborative systems must fulfil not merely the single user attributes of being useful and usable but also the social attribute of being acceptable to the user community, particularly by clearly addressing concerns of privacy and ownership.

References

1. Benford, S., Snowdon, D., Greenhalgh, C., Ingram, R., Knox, I. and Brown, C. VR-VIBE: a virtual environment for co-operative information retrieval, *Computer Graphics Forum*, **14**(3), 1995, pp. C-349 - C-360.
2. Brewer, R.S. and Johnson, P.M. *Toward Collaborative Knowledge Management within Large, Dynamically Structured Systems*, Technical Report CSDL-TR-94-02, Department of Information and Computer Sciences, University of Hawaii, 1994.
3. Chalmers, M. Design perspectives in visualising complex information, in Spaccapietra, S. and Jain, R. (Eds.), *Visual Database Systems 3: Visual Information Management*, New York: Chapman & Hall, 1995, 103-11.
4. Chang, S.J. and Rice, R.E. Browsing – a multidimensional framework, *Annual Review of Information Science and Technology*, **28**, 1993, pp. 231-76.
5. Foss, C.L. *Learning from errors in Algebraland*, Technical Report IRL-87-0003, Institute for Research on Learning, Palo Alto, CA, 1987.
6. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. Using collaborative filtering to weave an information tapestry, *Communications of the ACM*, **35**(12), 1994, pp. 61-70.
7. Grudin, J. Groupware and social dynamics: eight challenges for developers, *Communications of the ACM*, **37**(1), 1994, pp. 92-105.
8. Hill, W.C. and Hollan, J.D. Edit wear and read wear, *Proceedings of the Conference on Human Factors in Computing Systems. CHI'92*, Monterey, CA, ACM, 1992, 3-9.
9. Hill, W.C. and Hollan, J.D. History-enriched digital objects: prototypes and policy issues, *The Information Society*, **10**(2), 1994, pp. 139-45.
10. Hill, W.C., Stead, L., Rosenstein, M. and Furnas, G. (1995), Recommending and evaluating choices in a virtual community of use, *Proceedings of the Conference on Human Factors in Computing Systems (CHI'95)*, Denver, CO, ACM, 194-201.
11. Kantor, P.B. The adaptive network library interface: a historical overview and interim report, *Library Hi-Tech*, **11**(3), 1993, pp. 81-92.
12. King, G., Kung, H.T., Grosz, B., Verba, S., Flecker, D. and Flecker, B. The Harvard self-enriching library facilities (SELF) project, *Proceedings of Digital Libraries '94*, College Station, TX, 1994, 134-8.
13. Koenig, M.E.D. Linking library users: a culture change in librarianship, *American Libraries*, **21**(9), 1990, pp. 844-9.
14. Lim, J.-G. Using Coollists to index HTML documents in the web, *Computer Networks and ISDN Systems*, **28**(1-2), 1995, pp. 147-54.
15. Lin, X., Liebscher, P. and Marchionni, G. Graphical representations of electronic search patterns, *Journal of the American Society for Information Science*, **42**(7), 1991, pp. 469-78.
16. Malone, T.W., Grant, K.R., Turbak, F.A., Brobst, S.A. and Cohen, M.D. Intelligent information sharing systems, *Communications of the ACM*, **30**(5), 1987, pp. 390-402.
17. Maltz, D.A. and Erlich, K. Pointing the way: active collaborative filtering, *Proceedings of the Conference on Human Factors in Computing Systems. CHI'95*, Denver, CO, ACM, 1995, 202-9.
18. Maltz, D.A. *Distributing Information for Collaborative Filtering on Usenet Net News*, MSc Thesis, Department of Electrical Engineering and Computer Science, MIT, 1994.
19. Park, T. and Chon, K. Collaborative indexing over networked information resources by distributed agents, *Distributed Systems Engineering*, **1**(6), 1994, pp. 362-74.
20. Rao, R., Pedersen, J.O., Hearst, M.A., Mackinlay, J.D., Card, S.K., Masinter, L., Halvorsen, P.-K. and Robertson, G.G. Rich interaction in the digital library, *Communications of the ACM*, **38**(4), 1995, pp. 29-39.
21. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. GroupLens: an open architecture for collaborative filtering of netnews, *Proceedings of Conference on Computer Supported Cooperative Work (CSCW'94)*, Chapel Hill, NC, ACM Press, 1994, 175-86.

22. Rodden, T. A survey of CSCW systems, *Interacting with Computers*, **3**(3), 1991, pp. 319-54.

23. Röscheisen, M., Morgensen, C. and Winograd, T. Beyond browsing: shared comments, SOAPs, trails, and online-communities, *Computer Networks and ISDN Systems*, **27**(6), 1995, pp. 739-49.

24. Shardanand, U. and Maes, P. Social information filtering: algorithms for automating "word of mouth", *Proceedings of the Conference on Human Factors in Computing Systems (CHI'95)*, Denver, CO, ACM, 1995, 210-7.

25. Twidale, M.B. and Nichols, D.M. *Situated Observations of Library Users*, Internal Report, Computing Department, Lancaster University, 1995.

26. Twidale, M.B., Nichols, D.M., Mariani, J.A., Rodden, T. and Sawyer, P. Supporting the active learning of collaborative database browsing techniques, *Association for Learning Technology Journal*, **3**(1), 1995, pp. 75-9.

27. Twidale, M.B., Nichols, D.M. and Paice, C.D. *Browsing is a Collaborative Process*, Technical Report CSEG/1/96, Computing Department, Lancaster University, 1996.

28. Twidale, M.B., Nichols, D.M., Smith, G. and Trevor, J. Supporting collaborative learning during information searching, *Proceedings of Computer Support for Collaborative Learning (CSCL 95)*, Bloomington, Indiana, 1995, 367-74.

29. Whitehead, S.D. Auto-FAQ: an experiment in cyberspace leveraging, *Computer Networks and ISDN Systems*, **28**(1-2), 1995, pp. 137-46.

30. Wittenburg, K., Das, D., Hill, W.C. and Stead, L. Group asynchronous browsing on the World Wide Web, *Proceedings of the Fourth International World Wide Web Conference*, Boston, MA, 1995.

This is the authors' final version of:

Twidale, M.B and Nichols, D.M. (1996) Collaborative browsing and visualisation of the search process, Aslib Proceedings, 48(7-8) 177-182.