

**LANCASTER
UNIVERSITY**

**Computing
Department**



Issues in Collaborative Database Browsing

M.B. Twidale, D. Nichols, J.A. Mariani, T. Rodden and P. Sawyer

Centre for Research in CSCW

Research report : CSCW/14/1994

©University of Lancaster 1994. Copying without fee is permitted provided that the copies are not made or distributed for direct commercial advantage and credit to the source is given. For other copying, write for permission to:-

Computing Department, Lancaster University, LANCASTER, LA1 4YR, UK. Phone: +44-1524-593041; Fax: +44-1524-593608; E-Mail: cscw-info@comp.lancs.ac.uk

Issues in Collaborative Database Browsing

M.B. Twidale, D. Nichols, J.A. Mariani, T. Rodden and P. Sawyer

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

[mbt,dmn,jam,tam,sawyer]@comp.lancs.ac.uk

13/11/94

1. Introduction

Interfaces to databases have traditionally been designed as single-user systems. The existence of other users has implicitly been assumed to be an attribute of the system that should be hidden from end-users. In recent years the emergence of the field of CSCW (Computer Supported Cooperative Work) has highlighted the importance of collaborative approaches in many diverse activities. This report examines the possibilities for extending the CSCW approach to searching and browsing in computerised databases. In particular we are concerned with supporting the learning of browsing techniques.

In the Higher Education sector the most visible data resource is that of the university library; most of which have computerised their stocks with online public access catalogues (OPACs). This resource is supplemented by the provision of databases on CD-ROM of books and journal articles. Increasingly, both the student and the researcher are including remote libraries and databases amongst their data sources [Sack, 1986]. The relentless growth of the Internet and the success of browsers such as Mosaic only add to the enormous variety and number of resources available. The computerised library (or other similar database) is already an integral part of undergraduate courses and the skills to effectively access and utilise can be expected to become increasingly valued, both in academic and commercial environments [Jackson, 1989]. We can expect that in the near future, the ability to effectively browse databases will be one of the transferable skills expected by employers of all graduates regardless of their course of study.

This report takes the OPAC as a context for examining the issues involved in supporting collaborative activities in computerised databases. In addition to conventional searching/browsing behaviour, new types of activity become possible which are not feasible in physical libraries [Rice, 1988]. The report includes a literature review on searching and browsing, an exploratory study of collaborative browsing and a discussion of the mechanisms needed to support collaborative activities in databases. Also discussed are methods to aid the teaching of database searching techniques in the absence of computerised support systems.

2. Literature Review

2.1 Browsing and Searching

Although users are said to ‘search’ online databases much of their activity can be more accurately described as browsing. By browsing we mean that activity when a user does not know in advance the item or items she is looking for, although she does know some of its properties. [Bates, 1989] describes a model that integrates these two descriptions into a *berrypicking / evolving* search. The key feature that distinguishes this process is that the query changes as partial results are retrieved. The variation in queries owes more to the browsing model than to the classic *single query / single output set* model of information science research, e.g. [Robertson, 1977].

Authors on browsing have also noted that the behaviour of users is not easily characterised by a single model of searching. The consequence is that there is no generally accepted definition of browsing [Hanock-Beaulieu, 1989]. However, there is agreement that the term should refer to a spectrum of activity rather than a single narrowly-defined behaviour.

Source:	Browsing Definition	Characterising
[Apted, 1971]	1) General 2) General purposive 3) Specific	examination of sources index search with formal strategy
[Herner, 1970]	1) Semi-directed 2) Directed 3) Undirected	may not know where information is specific goal; no strategy no conscious intent or goal
[Levine, 1969]	1) Random browsing 2) Quasi-random 3) Semi-deterministic	unknown collection previously explored area searching in a limited area
[Cove and Walsh, 1988]	1) Search browsing 2) General purpose 3) Random	known goal consulting sources serendipitous finds

Table 1. Some examples of definitions of browsing

Definitions that do not adopt the typology in Table 1 also emphasise the indeterminacy and informal nature of browsing, e.g.

Let us, then, say that the search without “some object” is a reflection of an awareness that a knowledge gap exists; that there is no way to articulate its character; that the function of filling the gap is the query; that such functional information seeking is “browsing”.

[O'Connor, 1993]

Browsing is an informal or heuristic search through a well connected collection of records in order to find information relevant to one's need. The searcher evaluates the information currently being displayed to determine its value relative to the information he is seeking. Once this evaluation is made, the user then decides what items to select for display and evaluation.

[Thompson and Croft, 1989]

Browsing may be defined as a search, hopefully serendipitous. In connection with a library, one may browse ... through a portion of the library shelves in the hope of finding a text which might contribute the fact or idea needed in some intellectual effort. ... In each case the browser is not certain he will find anything of use to him, but he has hope, and past

Dimension	Example	Values	
purpose	recreational or information	intrinsic	↔ extrinsic
goal	learning or selecting	non-goal-directed	↔ goal directed
content knowledge	physical item or information	non-content-specific	↔ content specific
structure knowledge	physical pathway or meta-information	non-path specific	↔ path-specific
location knowledge	position on a shelf or a list	non-location-specific	↔ location-specific
resource focus	real objects or organising structure	content	↔ search path

Table 2. A dimensional typology of browsing of browsing from [Chang and Rice, 1993]

experience supports that hope.

[Morse, 1970]

The substantial review of [Chang and Rice, 1993] examines browsing from the perspectives of several different disciplines including consumer behaviour, organisational studies, media studies and environmental design. In a similar approach to [O'Connor, 1993] browsing is characterised as an 'iterative movement in a scanning and examining activity' [Chang and Rice, 1993]; Table 2 shows the resulting typology. This typology can be used to characterise various browsing situations; a searcher may have a poor knowledge of structure yet strong knowledge of content, i.e. they know what they are looking for but not how to find it.

[Chang and Rice, 1993] emphasise the dynamic aspects of browsing; that 'serendipitous findings can change a relatively ill-defined goal to a more well-defined one and may intensify one's underlying motivation.' Serendipity is also regarded as an important of feature of browsing that often results in useful 'hits' [Hawkins, 1982; Rice, 1988]. Equally important are the skills to benefit from the serendipitous find and use it to inform future searching.

Historically, browsing has been of two sorts: browsing the indexes and browsing the shelves. Browsing the shelves has several physical constraints on possible finds:

- only items held by the particular physical library are available for browsing
- only items on open access are available
 - e.g. the books on browsing at Lancaster cannot be physically browsed
- only items not borrowed by other lenders are available
- items may be at different places in the library to their stated shelving area
 - i.e. being photocopied, being used for private study
- items are not shelved according to subject groupings but by media, [Rovelstad, 1976].
 - e.g. different media are stored separately: the video room, the audio tape room
- the size of library stock and the consequent time taken [Enright, 1975].
- the composition of library stock
 - the presence of obsolete material has a negative browsing value [Seymour, 1972].
- true undirected browsing can only take place in a randomly-arranged library [Herner, 1970]
- subject divisions hinder 'higher browsing' (that activity whereby a researcher from one discipline is able to inspect the literature in another subject area) [Celoria, 1968].
- subject divisions are important: purposeful browsing occurs in areas where reader has strong interests, law of diminishing returns applies [Morse, 1970]
- shelving characteristics influence browsability

- books on middle shelves borrowed more often [Hyman, 1980]
- item characteristics influence browsability
 - dust jackets, bright covers, blurbs etc.

Computerised browsing has greatly changed the browsing process; instead of physically moving round shelves the browser now examines a VDU. OPACs have removed several physical constraints from the browser:

- the potential browsable stock is now equivalent to the entire library's stock
 - including closed access collections
 - including items borrowed by other lenders
 - items elsewhere in the library
- remote libraries are now browsable
 - even if the items are not immediately available for consultation
- shelving effects are eliminated
- item characteristic effects are eliminated

Potential losses from computerisation

Computerised library databases offer features not available in traditional library information structures (such as title and keyword searching and the ability to work remotely), but they may also lose some useful features. These include the spatial nature of the way information is presented in a library: the ability to walk round the bookstacks and to exploit our powerful spatial memory abilities [Beheshti, 1992].

The organisation of bookstacks can also facilitate browsing: a searched-for book on a shelf will have related books near to it that may be equally or even more relevant to the user's vague and constantly evolving information needs. An electronic view of information is remarkably impoverished compared with the vast amount of peripheral information of real books. These include: the ease of seeing whether a book is brand new, well thumbed, borrowed a lot, or ancient but never consulted [Mitev, 1989]. The context of the book on its shelf can also be informative. The number of similar or books nearby gives a sense of the size of the related available information.

Although providing useful additional information, its nature means that it is easy to choose whether to notice it or not. If the same information was converted to textual form and made available in an OPAC it would require far more effort to read it than glancing at the shelves. The peripheral information embodied in the book gives powerful cues as to whether it is worth glancing at the book's contents. Furthermore of course the book itself is accessible: having seen it one can examine it. Not only can one examine the contents, but the cover, particularly the blurb on the back can be a very useful source of information [Spenceley, 1980; Spiller, 1979].

People also have well developed spatial awareness and memory skills that can be exploited in navigating a physical library and in supporting the retrieval of information by its position [Chang and Rice, 1993]. However not everything is easily accessible in a physical library; items may have been borrowed or moved. For example at Lancaster the librarianship section is not on open access. So you can't physically browse the books on browsing! Even the old-fashioned card-index file can convey peripheral information from nearby cards, the colour and degree of ageing of the card and even how dog-eared it is.

Some improvements available with computerised browsing include:

- random stimulation: the system adds in random items into search results to stimulate creativity and provoke serendipity. The user can specify the degree of randomness the system uses [O'Connor, 1993].
- biomorph approach (based on the work of biologist Richard Dawkins [Dawkins, 1986]): the system generates items and the user selects whichever is nearer to the type of item he is looking for. The system then takes that as a central item and generates further surrounding items and the process is repeated.
- libraries can electronically alert users when relevant items are acquired
- integration of library systems with personal bibliographic software (e.g. Endnote).

- see section 9 for further examples.

Summary

The key features of searching/browsing behaviour are that:

- much of what is described as searching is in fact browsing
- the goal changes during the browse
- partial results can affect the goal
- the goal itself may be ill-defined and partially specified
- the searcher may not be aware of all the methods with which to approach the goal
- serendipity plays an integral part
- computerising the catalogue has drastically changed the browsing process

2.2 Social Aspects

Reviews of browsing such as [Ayris, 1986; Chang and Rice, 1993] show that the

Tactic Name	Description of Tactic
CONSULT <i>Idea tactic</i>	To ask a colleague for suggestions or information in dealing with a search. Comments by practising librarians indicate that this is a valuable and much-used tactic.
WANDER <i>Idea tactic</i>	To move among one's resources, being receptive to alternative sources and new search ideas triggered by the materials that come into view. In our field ... one may hypothesise that to WANDER promotes serendipity and enables useful sources that would not otherwise be discovered.
BIBBLE <i>Information search tactic</i>	One way to cope with the file structure is to find a way to do without it altogether. ... BIBBLE is based on the abbreviation "bibl" for "bibliography." To BIBBLE is to look for a bibliography already prepared, before launching oneself into the effort of preparing one. More generally, to BIBBLE is to check to see if the search work one plans has already been done in a usable form by someone else.

Table 3. Tactics with social aspects from [Bates, 1979a; Bates, 1979b]

predominant scenario is one of *individual* searchers accessing either physical books or electronic records. Although the social aspects of browsing behaviour appear to have been downplayed there are some references.

Bates has written two articles that enumerate methods to aid searchers [Bates, 1979a; Bates, 1979b]. *Idea tactics* [Bates, 1979a] are those designed to generate new ideas or solutions. *Information search tactics* [Bates, 1979b] are those designed to help in the search process and to help teach novice searchers. Table 3 shows the tactics that have explicit social aspects.

To CONSULT is described as asking a colleague for help although it really covers two related scenarios: asking a colleague and asking a member of the library staff¹. In the future it is possible that this tactic could also include asking an intelligent computer-based system.

To WANDER can be viewed in a similar way; as the resources available to a searcher are not limited to physical items but can include people and computerised systems.

¹ For students there are three different types of people to consult: other students, academic staff and library staff.

To BIBBLE is to take advantage of searches that have been done in the past and not waste time and resources 're-inventing the wheel'; it is a call for *search re-use*. A bibliography is a structured version of the results of a past search. However the results of most searches are not published as bibliographies but are private and local to the searchers. This means that many searches that are conducted fail to BIBBLE properly; they fail to take advantage of previous results because there is no mechanism to support the sharing of this information.

The use of bibliographies has been suggested as an alternative to browsing [Urquhart, 1976]. Browsing, it is argued, is an inefficient means of finding material when the stock to be browsed is large and much better results could be obtained by using pre-existing bibliographies. Indeed, it is suggested that browsing survives only through 'bibliographical laziness or ignorance' [Urquhart, 1976]. Many of Urquhart's arguments are disputed by other researchers (e.g. [Ayris, 1986]) but most of them are reliant on a model of physical browsing around bookstacks and are irrelevant to electronic systems that offer greatly superior facilities for browsing [Rice, 1988].

Most OPACs and databases do not provide mechanisms to support social activities. This can be attributed to an implicit general belief on the part of system designers: *you can only browse for inanimate objects*. We believe that browsing for people, their electronic representations or representations of their activities, is a neglected and important area. [Chang and Rice, 1993] mention the social aspects of browsing only in informal situations such as 'hallway chatting or after-meeting discussions.' This 'social browsing' [Root, 1988] is held to be an important part of knowledge creation and collaborative work.

Physical libraries are not environments that are designed to be conducive to the social aspects of browsing; they limit social activities in several ways:

- people must be in the same physical library (co-located activities dominate)
- people must be present at the same time (synchronous activities dominate)
- people must be in the same part of the library
- people must either:
 - already know each other, or
 - be prepared to approach strangers
- by providing an environment that does not encourage dialogue
 - e.g. keep quiet in the library
- accidental nature of meetings means that time is not scheduled
 - e.g. 'I've got to be in a meeting in five minutes.'
- limiting information that people carry with them
 - e.g. 'I got something interesting back in the office.'
 - this increases the acquisition costs of the item

All of these factors contribute to a lack of mutual knowledge amongst browsers; which in turn contributes to a lack of information exchange [Krauss and Fussell, 1990] and numerous missed opportunities for effective collaboration.

3. Informal Interviews of Involved Parties

An initial part of our investigation was to undertake a series of informal interviews. The chief aim was to inform our design intuitions of the requirements of users, as well as to give ourselves examples to compare with the results of the large scale systematic studies described in the literature. Two groups were targeted: subject librarians and relatively unsophisticated library users. It should be noted that these interviews were small scale and opportunistic. They did though serve to confirm the findings in the literature.

The subject librarians described their activities with undergraduate and postgraduate students. This is a complex and subtle interaction in that they are undertaking two different roles: both aiding the search/browsing process by acting as an information consultant or intermediary as well as acting as an educator / empowerer to enable the student to acquire the skills to undertake the action independently. This dual activity leads to decisions about how much to teach a client so as not to overwhelm her, and how much to do for her. The clearest example of this dual role came from the law librarian and his use of the LEXIS database [Allan, 1993; Bosworth, 1993]. For some clients, he would undertake the whole interaction

and give them the results, with others he would sit alongside offering help or be nearby to help out when things got confusing, whilst other clients he would leave to their own devices to explore and learn about the system.

We particularly asked subject librarians about their activity as educators. They regard this as an important activity but one made complex by lack of resources. The class-based approach although economic was not considered particularly effective or rewarding (similar to those of [Warmkessel and Carothers, 1993; Wielhorski, 1994]). There did seem to be a danger of 'losing' less technically adept or confident students.

We also interviewed a selection of undergraduates on their experiences of using the Lancaster OPAC system. What was surprising to us as computer scientists was how contented they seemed to be with the system. We had expected that some would still prefer to use the card index system. However on the occasions when we saw people using that, it transpired that this was because they were looking for some obscure item that had not yet been added to the computerised catalogue rather than by preference. Most students seemed happy with the OPAC system, claimed it served its purpose and compared it favourably with that of others they had used. According to one of the subject librarians who had also observed this, he believed that this was due to the feature of the system of allowing a student to browse the stacks electronically, by returning a hit in its context, rather than just the lone result. By contrast, the opinion of people back at the Computing Department was that the OPAC system was dreadful because it had such a primitive interface, restricted functionality and a few bizarre inconsistencies. It seems that the general users without much computing background had adapted to the system, and not being aware of the potential of having anything better could live with it, although (we suspect) not using much of its functionality.

4. Studies of Collaborative Browsing

Building on the information gleaned from the literature and our interviews, we created an exploratory environment to study collaborative browsing by combining existing computational tools.

Using a telnet connection to the Lancaster University Library, the BIDS² Service in Bath and the MELVYL library system of the University of California, the Unix talk program, X Windows and a keypress recording program we were able to build a prototype that allowed remote synchronous collaborative browsing to be performed by a searcher and an experienced searcher. The system recorded the actions of the searcher and the sessions were also videotaped.

Table 4 shows a partial transcript of one of these sessions with a human expert searcher sitting next to a subject. The transcript shows several interesting features:

- the use of pointing at the screen to establish a common referent.
- the occurrence of several 'common errors' [Tenopir, 1984]
 - spelling mistakes (e.g. ecofinism)
 - case mistakes (e.g. duhb instead of DUHB)
- the subject twice entered the classmark KDQWJ, however he wasn't aware that the second item from the subject index had the same classmark as his previous selection.
- the subject asked whether keyword searching could be done with more than one word. The example screen had just shown him an example ('acid rain') with more than one word moments earlier but because the screen had changed he had forgotten it.
- browsing of large numbers of results (multiple 'forward screen' commands -'f').

Recordings of the exploratory collaborative browsing environments exist but have not yet been analysed in the same detail as Table 4. However some general points that were noted about these sessions:

- there were several timing problems. The expert would attempt to correct some action and the searcher would misinterpret the response as referring to another concept. This then led to a further problem as the intervention would be applied to the wrong aspect of

² Bath ISI (Institute for Scientific Information) Data System holds many indexes of articles.

the search.

- one subject wished to be able to exclude items, namely dictionaries, which was unavailable on the system
- ambiguity of messages. As with timing problems messages were interpreted as referring to items other than their intended ones.
- separable expert connection. It was found to be extremely helpful for the expert to have a separate connection to the database so that he could perform independent searches.
- results could be pasted into the talk window using X windows. This sharing of results was effective and quick although the formatting of the results sometimes cause problems.

<CR> is a carriage return
 is the delete key
 S is the subject of the experiment
 E is the experimenter
 Time in HH:MM

Time	Keypress	Notes
00:00		Description of facilities and experimental procedure
00:01	a	Personal Bibliography option
	2	Create new Personal Bibliography (PB)
	c	Create option
	0010580<CR>	Library card number
	annecy<CR>	User supplied password (doesn't work)
	{my password}	
00:04	test<CR>	Bibliography code
	<CR>	No title for PB
	u	Unrestricted option for PB
	f	File new PB
	g	Any key to continue
	o	Invalid key
00:06	m	Main menu
	8	Subject search
	ecofeminism<CR>	Enters subject; entry does not appear S: "That's a good start isn't it." (irony) S: "Lets go for a new subject."
00:07	m	Main menu
	8	Subject search
	feminism<CR>	Enters subject
00:09	b	Back a page
	f	Forward a page
		S scans subject index
	9	Selects 'FEMINIST: PERSPECTIVE: WOMEN:SOCIOLOGY'
		Enters the classmark search section at KDQWJ
00:10	fffff...f	Forward several pages E: "So is this a part of the library you've not really explored?" S: "Yes."
	b	Back a page
00:12	r	Returns to subject index
	10	Selects 'FEMINIST:THEORY:WOMEN:SOCIOLOGY'
		Enters the classmark search section at KDQWJ
		This is exactly the same place as before!
00:13	m	Main menu

	<p>S: "I was actually wanting a specific class of books on ecofeminism but there wasn't one. They're all over the place. So it wasn't very helpful for me in finding the kinds of books I wanted. Because I would have had to go through the whole lot."</p> <p>E: "So there are books in the library on that?"</p> <p>S: "They were in another classification, or minor classification."</p> <p>S: "So I would to have to have gone through about 3 different sets of classifications."</p> <p>S: "It wasn't worth going on."</p> <p>E: "Ok."</p> <p>E: "These are the sort of things we're interested in – decisions about when to abandon things or not."</p> <p>E: "So that was where you just getting too much, and is it that its fairly diffuse?"</p> <p>S: "Yes. I know about – there's a list of about 60 or 70 books and about 6 of them were ones that I could recognise in the area that I wanted."</p> <p>E: "Ok."</p> <p>S: "I realised that would be repeated whatever item I selected from that list – it would have been too tedious to go on."</p> <p>E: "So there are certain key books that you already know about?"</p> <p>S: "Yes – then it would have been really hard to find – I think."</p> <p>E: Right.</p> <p>S: "Some of these titles are obviously ecofeminism – that I know are because I know them."</p> <p>E: "Ok."</p> <p>E: "Perhaps try the same thing using the keyword search because before you did it by 8 – this one."</p> <p>E points at the screen</p> <p>S: "Ok."</p> <p>E: "I warn you that this is a new feature, it has some bugs in it so it might go wrong."</p> <p>S: "Ok."</p>
k	Selects keyword searching
ecofeminism<CR>	Enters search term
	4 hits as set 1
m	
	E: "If you–"
c	
	E: "Try using a delete key."
	
dualism<CR>	11 hits as set 2
	S: "Does a keyword have to be 1 word?"
	E: "No."
domination of nature	
<CR>	0 hits; treated as 'Domination' and 'of' search

	E: "I don't know if you saw it flash up 'maximum of 2 words' – error message."
	S: "No."
	E: "It's very quick. So you can have 2 keywords together but not 3."
3 	S: "How can I get rid of 3?"
	E: "I think this is a mistake with it – I don't think you can get rid of 3."
	S: "Ok, well I'll abandon it."
	S: "I'll search for ... 1."
1	E: "Return to just bring it up."
	S: "Yes."
<CR>	list of 4 author titles is displayed
n	New search on keyword
ecofinism	
...	E: "It's a bit slow – erratic."
ecofeminism<CR>	4 hits
domination nature	
<CR>	1 hit
2<CR>	1 author/title display
	S: "Only 1 I've got."
p	S: "Previous – that's what I should have done before."
33	mistakes
m	S: "Now I want to find out what's at DUHB."
4	Classmark search
uhb<CR>	5 items displayed in classmark display
	E: "Can you see what has happened?"
	S: "It should be in upper case."
	E: "Oh, but its uhb." (points at screen)
	E: "But you wanted DUHB."
	S: "Can't read my own writing."
m4duhb<CR>	repeats mistake
m4	
<CAPSLOCK>	
DUHB<CR>	S: "That's better."

Table 4. Partial transcript of experiment browsing the Lancaster Library OPAC

Although collaborative browsing is possible using pre-existing tools it is clear that in order to fully take advantage of the potential benefits it offers specifically-designed software is required. Our experiments indicate the removal of ambiguity should be a key feature of such software.

5. Collaborative Activities

Computerised systems have undergone a series of evolutionary steps from single-user systems, to multi-user systems, to networked systems and to CSCW systems. Many diverse collaborative activities can now be supported by computers including those for which the this technology is a *necessary* pre-requisite e.g. [Brewer and Johnson, 1993; O'Neill and Gomez, 1994].

In addition to their collection of items, libraries also contain people: lenders, reference

users, enquiry staff, technical support staff, bindery staff etc. This social aspect of library activity has so far been ignored in the literature and in the implementation of computerised catalogues. In addition to serendipitous encounters (section 2.2) there are other classes of person-person interactions in libraries:

- collaborative work
 - a group of users who decide to work collaboratively, e.g. for a group project
- collaborative learning
 - asking a colleague for help
 - leaning over the shoulder of a neighbour to ask how they did something
- enquiries
 - a user cannot achieve the desired functionality from the library and asks for help
- administration
 - e.g. clarifying an inter-library loan request, new library card, paying fines
- borrowing an item
- user tours
 - a guided tour round the library to demonstrate its facilities
- user education
 - training users in using the systems: e.g. OPAC, CD-ROMs, BIDS, short loan etc.
 - formal courses or sessions
 - consultations about a task that turn into educational interactions

All of these interactions require the sharing of information and it can be expected that support mechanisms to facilitate collaborative browsing will *also* support many of these other activities.

Collaboration issues

It is not just the inanimate contents of a traditional library that convey useful information. The people can also be useful to a browser. One can observe and learn from the browsing techniques of others (both at the bookstacks and at the OPAC terminals), discuss issues with co-learners or with subject experts, and also be aware of the activities of others that may be of interest and relevance to one's own work. For example, upon seeing a colleague in an unexpected part of the library, you might choose to ask what she has found there. Similarly, upon seeing someone in 'your' area, you may decide to introduce yourself as someone also interested in that field. A computerised library that is accessed remotely will lack these advantages unless we take steps to re-introduce them into the system.

Research in Computer Supported Cooperative Work employs a useful classification of collaboration [Rodden, 1991]. Collaboration may be remote or co-located, as well as being synchronous or asynchronous. In conventional libraries, we can consider most cooperation to be co-located and synchronous, but the computerisation process makes the other permutations possible, while offering new opportunities for the first.

We envisage a number of scenarios in which the system being developed might be used in an educational context. In all cases the interaction may be synchronous (participants working at the same time) or asynchronous (participants leaving messages for each other):

1) Expert consultancy. A learner is browsing the database and decides that she is not making the progress she would like. If in a physical library she might go and talk to the subject librarian. For the database use, she may wish to communicate by telephone or email (or even ultimately by video link). The expert calls up a representation of the student's browsing history and composes suitable advice. This advice can be specific, general and remedial. It can help to solve the current task, explain a generic browsing technique and also correct any apparent misconceptions. As part of the explanation, the expert passes on an annotated browsing procedure which the student can view and even use on her terminal.

2) A 'language laboratory' for database skills. A number of individual learners are browsing a database as part of a practical class on database browsing skills. They may all be in the same teaching room, or working remotely on their own terminals elsewhere. The expert can observe the browsing activities of the students in turn, and offer advice as necessary.

3) Collaborative browsing for learners. Small groups of learners who have a similar browsing task make use of the awareness mechanisms to monitor and discuss the progress and activity of each other. This is known to be a useful learning activity both for the questioner (who asks questions such as "Why are you doing that?") and for the respondent who has to reflect on her action in order to generate a suitable explanation.

4) Serendipitous meetings. If the groups in (3) are made larger, database awareness can be a useful tool for supporting serendipitous meetings of the kind that naturally occur in physical libraries where strangers meet by browsing the same bookstack and finding common interests for collaboration, or when finding that a book they want is already being used by someone else, who therefore might be worth talking to. Now one can become aware when others are, or have been, browsing the same parts of the database.

6. Learning Browsing Skills

This section describes some of the issues that need to be considered when wishing to support the learning of skills whether by traditional or computational mechanisms.

The skills themselves can be usefully broken into two groups: strategies and tactics. Tactics can be regarded as the skills required to compose a single query as well as the knowledge of the commands of a given system's interface. Strategy involves the management of a sequence of queries in order to effectively browse the database to obtain the desired (but constantly evolving) result.

The following two lists give examples of strategies and tactics:

Tactics

Composing and refining a query

Use of Boolean operators

Use of keywords

Use of wildcards (*, ?, \$)

Handling low value words

Using word roots

Using a Thesaurus

Using Indexes

Using search categories

(author, title, abstract fields)

Displaying and recording the results

Rejecting noise

(e.g. query on 'browsing' yields papers about the behaviour of reindeer)

Strategy

Choosing between databases

Undertaking tentative queries

Getting a 'feel' for a database by random or unsystematic browsing

Free form browsing

Restricting or widening a search

Managing your goalstack:

Remembering where you are, why you came here and what you want

Adding new subgoals

Refining, deleting or modifying goals in the light of new discoveries

Coping with the disorientation of a large meandering goalstack

Handling tangential working

Exploiting serendipitous discoveries

Deciding between things to investigate now and things to investigate later

The distinction between strategy and tactics is a useful starting point for beginning to address the problem of domain dependence. Given the chaotic state of current database provision and in particular the huge variation in interfaces to different systems, one of the

major problems for all users and particularly for novices is to learn how to use all the different systems. It is a recurrent complaint of subject librarians. We can expect that the situation will improve somewhat with a drive towards standardisation of facilities. However the problem remains, and is particularly acute if one wishes to build a system to support the learning of browsing skills. Will it not be inevitably tied to just a single system? While acknowledging the very real problems that remain to be resolved, we believe that a useful approach is to focus chiefly on the teaching of browsing strategy on the grounds that this will vary less from system to system.

Learner errors

Another useful area to focus on when developing a learning environment is that of the classic errors and their underlying causal misconceptions. We can distinguish between misconceptions, where a student does something wrong but believes it to be the right way of working, and slips where the student accidentally (but perhaps frequently) makes an error that, if brought to her attention she can easily spot and correct. Slips are often the cause of working memory overload but remain undesirable rather than merely annoying as a sequence of them may be too bewildering for a novice to recover from. In the context of browsing strategies, by 'wrong', we mean an action that although it might be perfectly legal, is not one that an expert would undertake. In fact, many of the errors of novices are errors of omission rather than commission: they fail to exploit opportunities for further investigation that an expert would spot and make use of.

Within the set of errors to be collected and analysed are those which may be classified as suboptimal work patterns. This is where the novice has settled into a way of working with the database that yields some results but appears to an expert as incredibly laborious or merely skimming the surface of the potentially available information. We need to treat such patterns with caution: it is a perfectly acceptable learning strategy to learn a minimal (but sub-optimal) subset of the available commands. The choice of when to intervene with further teaching is a difficult one and ideally one that should be initiated by the student. The explanations should not overly downplay the use of the earlier command subset.

Examples of such work patterns observed and cited by the subject librarians as well as noted in the literature [Bates, 1989; Yee, 1991] include:

- Failure to refine a search. The user gets 150 hits and proceeds to read them all.
- Failure to look at all the results when there are a reasonable number
(the reverse of the above)
- Failure to pursue an interesting lead
- Failure to revise a strategy that does not seem to be working
- Failure to revert to the correct goal after pursuing a sub-goal
(the remaining branches of the search are simply ignored)

6.1 The economy of learning

The failure of novices to learn more than the basics is all too easily ascribed to laziness or ignorance of the opportunities for improvement. However there may be rational reasons for the failure to learn that need to be borne in mind when developing a learning environment. If possible the barriers to learning should be lowered.

Why might novices fail to improve?

- There are just too many options that they don't know which to investigate.
- In order to progress to the next level of performance they have to learn about too many extra features.
- They are too focused on the task in hand. The priority is always to do the work and they are in too much of a rush to spend time learning about new features
- They do not believe that the effort of learning will be justified by the improvement in their browsing
- Although able to use the new features, they make so many mistakes that it is safer to stick with the familiar subset where they feel more in control
- Getting something wrong is felt to be too embarrassing

Learner types

We can envisage different types of user of library databases who will have very different learning styles and needs [Wielhorski, 1994]. Implicit in our discussion so far has been our stereotypic novice: an 18 year old undergraduate majoring in arts/humanities with minimal computing background. The subject librarians mention that they are particularly sensitive to the needs of mature students (particularly those taking undergraduate courses alongside 'conventional' students) [Whitlatch, 1983]. Many of the former feel insecure with computer use, believing that their younger co-learners are far more knowledgeable and confident than they are. The following is a first attempt to list some of the types of learner.

- Technophobe control freak
 - Fearful of computers in general
 - Low opinion of their own competence
 - Keen to know details thoroughly before progressing
 - Wanting 'a course'
 - Unwilling to experiment

- Tweaker
 - Happy to experiment
 - Comfortable with areas of ignorance
 - Can live with 'magic words' that somehow achieve the desired effect
 - Want to 'get started now'
 - Although most computer scientists would fit here the group is not exclusively scientific

- Expert novice
 - Expert at using a traditional library
 - Expert at information handling & browsing
 - Novice at using computer systems
 - Will find certain functions that map from traditional library searching (eg author or classmark searching) far more intuitive than new ones (titles and keywords)

- Novice expert
 - Novice at information handling & browsing
 - Expert at using computer systems
 - Keen to learn about the functions available.
 - Less able to use domain information to guide the choice between options and strategies.

7. Sharing the browsing process as well as the product

The main issue to emerge from our study was that 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, who can then offer suitable advice on building on a search. Normally this information is only available by looking over the shoulder of a user as they undertake a search. It is naturally very hard for a novice to accurately recall the steps taken in a prior search in order to ask for advice on other things to try. By contrast, if a record of the history of a search is available, the user can subsequently approach a librarian and say "I did this (pointing) and I didn't seem to get very much and I'm sure that there must be more stuff there: what am I doing wrong?". The librarian can make quite sophisticated use of a search record. It can reveal not only gaps in the user's browsing techniques but also an indication of their degree of searching sophistication which can be used in phrasing an explanation at the appropriate level of detail.

In addition an externalised representation of the search process reduces cognitive load on the user, which in turn reduces the likelihood of slips. It also facilitates reflection – a vital component of learning. Reflection and dialogue with co-learners can be further encouraged by providing facilities for annotation. The user can comment on what she did and why, which strategies were effective and which ineffective, which were 'lucky' and which although a reasonable thing to try, happened not to yield very much this time.

An initial analogy to the process representation would be the Unix history list, but the actual representation will need to be considerably more sophisticated and flexible. A suitable crafted representation can illustrate the techniques of expert browsing by encouraging appropriate actions. The student can be given exemplars both of best search practice and of naive browsing.

8. Teaching Database Browsing

This section describes the aspects of the work that can be used in the teaching of database searching skills to novices.

Interviews with subject librarians who conducted database skills workshops yielded the following comments:

- wait until students have a real need; authentic activities are considerably better than contrived scenarios. Motivation is improved by using ‘real’ relevant data.
- it helps if students can see the keyboard / mouse. This is particularly important with computer novices who may not understand verbal references to key combinations. (Again, authentic rather than contrived)
- it helps if the teacher can point at the screen and be seen by the learners. This enables mutual common knowledge to be easily established. Hands-on work or live demonstration are preferred to slides or overhead transparencies [Wielhorski, 1994].
- teachers prefer same-time same-place one-to-one teaching. This is also called point-of-use instruction and is preferred by learners [Wielhorski, 1994]. This approach has also been successfully used as a means of reducing the disruption caused by introducing an OPAC to a paper-based library [Sager, 1986]. Unfortunately it is a very expensive approach and can be difficult to arrange. Consequently we need to examine lower cost alternatives that maintain many of the advantages.
- a common goal is to get the students to a minimal level of competence where they have a subset of the available skills. This enables the students to work independently and obtain useful results. A study at Hofstra University suggests that students do *not* need to know advanced searching techniques, thesaurus use, anything beyond basic Boolean strategies

[Barbuto and Cevallos, 1991]. Carroll's work on *Minimal Manuals* for word processors provides strong evidence for the effectiveness of this approach [Carroll, 1990]

- each additional layer of sophistication should be accompanied by a demonstrable increase in utility. This makes clear to the student that it is worth the effort of learning about and practising the new features. The means should be closely coupled with the ends; the browsing needs of users.

From our work on systems development, the importance of a record of the search process indicates that even a paper representation may be a powerful tool to support the understanding of the nature of browsing and its associated strategies.

Different visualisations of the search process may be provided. These could be exemplars of best practice which include and emphasise the acceptability of failed searches. They can also include examples of the detection and exploitation of serendipity. By contrast, the novice can also be provided with exemplars of poor searches, including failures to exploit ideas and subgoals never returned to. These visualisations can be annotated to illustrate their main points.

Example searches can also be used to teach syntactic details, especially to those unused to formal grammars. Copying and adapting examples (the case based approach) is recognised as a powerful educational device. Its chief disadvantages are that it cannot cover any but a few common cases. Also it can lead to errors due to inappropriate mapping from the example to the case in hand. However, as few people are willing to read formal manuals, despite these dangers it can be a useful way of providing information.

Given a list of common misconceptions, a traditional way of teaching browsing either lecture-based or by use of handouts can directly address the problem by referring to and refuting these misconceptions, and by offering more suitable analogies to understand the process. For example it is useful to stress that the database search engine is not like a skilled librarian filtering one's requests with a lot of common sense and domain knowledge but rather like a very simple-minded and overly literal assistant who almost seems to try to deliberately misinterpret requests whenever possible.

Where the computer-based environment may provide interactive support for browsing, a static version may be created as a tick list of strategies or tips to consider. This could be laminated to allow for reuse. The student would work through the list, marking those activities relevant to her current task and writing in the gaps available lists of goals and subgoals, other aspects to investigate and issues to consider later.

Table 5 indicates a first attempt at listing the issues that might be represented.

<p>Things to think about when using a database Write here what it is you want to find out about</p>
<p>Come back every so often and consider whether you now have a different or a more precise aim.</p>
<p>Are you currently:</p> <ol style="list-style-type: none"> 1) Trying to get a feel for what is in the database 2) Trying to find out about a very general area 3) Trying to find books about a particular topic 4) Trying to find a particular book
<p>Words to search with List your current key terms here</p> <p>List some alternative terms here</p>
<p>Getting started in a new area Can you think of a classic introductory book? Use this and then try going off at tangents from it</p>
<p>Having done a search</p>
<p>Viewing the results If there are not too many, try having a quick look at them all If there are quite a few, is there a way at rapidly skimming them to spot interesting items If there are too many, try to think about how to restrict a search</p>
<p>Coping with too many results – You need to narrow the search Can you think of some more precise terms? Can you combine terms using AND to make the search more restrictive?</p>
<p>Coping with too few (no) results – You need to widen the search Can you think of some less precise terms? Can you combine terms using OR to make the search less restrictive?</p>
<p>Going off at tangents – When you want to explore related work If there is a co-author, try looking at his/her work Examine the title for keywords that you haven't tried yet If the item has a classification (such as a classmark) try other items with the same classmark If the item has multiple classifications (classmarks, or is filed in a different place), try investigating these alternatives NB Note when you have gone off at a tangent if there still were some conventional items to examine</p>

Table 5. Strategies and tips for searching

9. Support Mechanisms

The previous sections have described several collaborative activities that can be undertaken by searchers and those who collaborate with them. This section details the software support that can facilitate these interactions.

The berrypicking model of [Bates, 1989] expects there to be many partial results during the search process. On some systems (e.g. CD-ROMS) at Lancaster Library there is the facility to generate a hard copy of your search activities and/or results. This recording of

results is clearly an important aspect of an evolving search and software support should make this straightforward. This facility has been provided on the BIDS service by allowing partial results to be emailed to a user-supplied address.

A new feature of the Lancaster OPAC is the provision of bibliographies that users can create and are stored online. These can be made publicly accessible and so the information can be shared. One use is for lecturers to make a course bibliography available. This is a basic mechanism and is not integrated into users' personal bibliographic systems.

All these mechanisms allow means to share the *products* of browsing. So far though, none provide a means of sharing the browsing *process*. Based on the work described earlier we believe that it is the sharing of the process that is crucial for effective collaboration.

This implies that there is a need to make the process an object which can be:

- visualised – the visualisation of a process enables searchers to concentrate on the specific sub-task at hand and removes the cognitive load of maintaining a record of the search history. Cognitive load is held to be an important reason why much user behaviour is inefficient [Rudd and Rudd, 1986]. The visualisation of the process also helps the searcher to detect, repair and reflect on aspects of the process.

- communicated – the search process can be sent and received to different users. The search is a structured object; not just plain text. A support system would also allow this object to be stored, recalled and executed.

- edited – a user should be able to mould the search object so that what is shared is under their control. This enables a user to elide portions of the search that are not relevant to a particular communicative act. This is especially important for educational applications.

- annotated – a user should be able to have notes associated with particular portions of the search object. Again, this is important for educational and online help applications.

A topical example is that of the Universal Resource Locator (URL) from the World Wide Web; a URL is a frozen edited search history which is easily communicated and can be executed by a recipient. The *Hotlist* of Mosaic is also a means of storage and recall of URLs.

Other browsers

One part of sharing the browsing process is to provide information on the browsing activities of other browsers. One simple mechanism would be to mark an item in the database and ask the system to list those other browsers that examine or borrow the item. Such a facility could be expanded to take account of the interests of other browsers and then filter the information so that only those browsers with specific interests were display. For example, those who have shared interests with the browser marking the item. A mock-up interface might look something like that shown in Figure 1. Such a mechanism would be a pre-cursor to sharing the results of searches that others had conducted. This would enable a browser to access those items that similar browsers had accessed in the past but which had not yet been found.

10. Conclusions

Our conclusions can be summarised as:

- the sharing of both the search *product* and the search *process* are important for collaborative activities
- there exists great potential for improving search effectiveness through the re-use of previous searches; this is one mechanism for adding value to existing databases.
- browsing is not restricted to browsing for inanimate objects; browsing for people is also possible and could be a valuable source of information
- searchers of databases need externalised help to reduce their cognitive load during the search process. This can be provided both by traditional paper-based technology and through computerised systems.

Item Annotation Interface	Item: Teach Yourself to Read Author: Janet N. John Type: book Date: 1994 Class: YVW3
--	---

History of Item

Show me a list of people who have:

browsed this item

marked this item

borrowed this item

In the last:

week

month

year

Whose descriptor contains:

Whose descriptor has any term
in common with mine

Who have also

browsed

marked

borrowed

Who has browsed/marked/borrowed:

more than items in my bibliography

more than items that I have br/mrkd/brwd

In the last:

week

month

year

Future of Item

In the future if anyone:

browses this item

marks this item

borrowes this item

Whose descriptor contains:

Whose descriptor has any term
in common with mine

Who has also

browsed

marked

borrowed

Send me a message, summarizing:

daily

weekly

monthly

Figure 1. An example of an item annotation interface.

Acknowledgements

Financial support for this research has been provided by JISC (Joint Information Systems Committee).

References

- Allan, R. (1993), Computerised legal research pioneer: Mead Data Central and the LEXIS service, *The Law Librarian*, 24(3), 131-133.
- Apted, S.M. (1971), General purposive browsing, *Library Association Record*, 73(12), 228-230.
- Ayris, P. (1986), *The Stimulation of Creativity: a Review of the Literature Concerning the Concept of Browsing, 1970-85*, CRUS Working Paper No. 5, Consultancy and Research Unit, Department of Information Studies, University of Sheffield.
- Barbuto, D.M. and Cevallos, E.E. (1991), End-user searching: program review and future prospects, *RQ*, 31, 225.
- Bates, M.J. (1979a), Idea tactics, *Journal of the American Society for Information Science*, 30(5), 281-289.
- Bates, M.J. (1979b), Information search tactics, *Journal of the American Society for Information Science*, 30(4), 205-214.
- Bates, M.J. (1989), The design of browsing and berrypicking techniques for the online search interface, *Online Review*, 13(5), 407-24.
- Beheshti, J. (1992), Browsing through public access catalogs, *Information Technology and Libraries*, 11(3), 220-28.
- Bosworth, K. (1993), In praise of law librarians: LEXIS in the United Kingdom - 1975-1993, *The Law Librarian*, 24(3), 133-136.
- Brewer, R.S. and Johnson, P.M. (1993), *Collaborative Classification and Evaluation of Usenet*, Technical Report CSDL-TR-93-13, Collaborative Software Development Laboratory, Department of Information and Computer Sciences, University of Hawaii.
- Carroll, J.M. (1990), *The Nurnberg Funnel*, Cambridge, MA: MIT Press.
- Celoria, F. (1968), The archaeology of serendip, *Library Association Record*, 70(1), 251-3.
- Chang, S.J. and Rice, R.E. (1993), Browsing – a multidimensional framework, *Annual Review of Information Science and Technology*, 28, 231-76.
- Cove, J.F. and Walsh, B.C. (1988), Online text retrieval via browsing, *Information Processing & Management*, 24(1), 31-37.
- Dawkins, R. (1986), *The Blind Watchmaker*, Harlow, UK: Longman.
- Enright, B.J. (1975), Bibliochothanasia: library hygiene and the librarian, in Barr, K. and Line, M. (Ed.), *Essays on Information and Libraries*, London: Bingley, 61-78.
- Hanock-Beaulieu, M. (1989), Online catalogues: a case for the user, in Hildreth, C.R. (Ed.),

- The Online Catalogue: developments and directions*, London: The Library Association, 25 - 46.
- Hawkins, D.T. (1982), Online bibliographic search strategy development, *Online*, 6?, 12-19.
- Herner, S. (1970), Browsing, in Kent, A. and Lancour, H. (Ed.), *Encyclopedia of Library and Information Science*, 3, New York: Dekker, 408-15.
- Hyman, R.J. (1980) "Shelf classification research: past, present – future?", University of Illinois, Internal Report.
- Jackson, A.H. (Ed.) (1989), *Training and Education for Online*, London: Taylor Graham.
- Krauss, R.M. and Fussell, S.R. (1990), Mutual knowledge and communicative effectiveness, in Galegher, J., Krauss, R.M. and Egidio, C. (Ed.), *Intellectual Teamwork: social and technological foundations of cooperative work*, Hillsdale, N.J.: Lawrence Erlbaum Associates, 111-45.
- Levine, M.M. (1969), An essay on browsing, *RQ*, 9, 35-6.
- Mitev, N.N. (1989), Ease of interaction and retrieval in online catalogues: contributions of human-computer interaction research, in Hildreth, C.R. (Ed.), *The Online Catalogue: developments and directions*, London: The Library Association, 142-76.
- Morse, P.M. (1970), Search theory and browsing, *Library Quarterly*, 40(4), 391-408.
- O'Connor, B. (1993), Browsing: a framework for seeking functional information, *Knowledge: creation, diffusion, utilization*, 15(2), 211-232.
- O'Neill, D.K. and Gomez, L.H. (1994), The collaboratory notebook: a networked knowledge-building environment for project learning, *ED-MEDIA 94 - World Conference on Educational Multimedia and Hypermedia*, Vancouver, BC, Canada, AACE, 416-423.
- Rice, J. (1988), Serendipity and holism: the beauty of opacs, *Library Journal*, 113(3), 138-41.
- Robertson, S.E. (1977), Theories and models in information retrieval, *Journal of Documentation*, 33(2), 126-48.
- Rodden, T. (1991), A survey of CSCW systems, *Interacting with Computers*, 3(3), 319-54.
- Root, R.W. (1988), Design of a multi-media vehicle for social browsing, *Proceedings of the Conference on Computer-supported Cooperative Work*, Portland, OR, ACM, 25-38.
- Rovelstad, M.V. (1976), Open shelves/closed shelves in research libraries, *College & Research Libraries*, 37(5), 457-67.
- Rudd, J. and Rudd, M.J. (1986), Coping with information load: strategies and implications for librarians, *College & Research Libraries*, 47(4), 315-22.
- Sack, J.R. (1986), Open systems for open minds: building the library without walls, *College & Research Libraries*, 47(6), 535-44.
- Sager, H. (1986), Training online catalog assistants: creating a friendly interface, *College & Research Libraries News*, 47(11), 721-23.
- Seymour, C.M. (1972), Weeding the collection: a review of research on identifying obsolete stock. Part I: monographs, *Libri*, 22(2), 137-48.

- Spenceley, N. (1980), *The readership of literary fiction*, M.A. (Information Studies), University of Sheffield.
- Spiller, D. (1979), *The provision of fiction for public libraries*, M.L.S. Thesis, Loughborough University.
- Tenopir, C. (1984), To err is human: seven common searching mistakes, *Library Journal*, 109(6), 635-36.
- Thompson, R.H. and Croft, W.B. (1989), Support for browsing in an intelligent text retrieval system, *International Journal of Man-Machine Studies*, 30(6), 639-668.
- Urquhart, D.J. (1976), National lending/reference libraries or libraries of first resort, *BLL Review*, 4(1), 7-10.
- Warmkessel, M.M. and Carothers, F.M. (1993), Collaborative learning and bibliographic instruction, *Journal of Academic Librarianship*, 19(1), 4-7.
- Whitlatch, J.B. (1983), Library use patterns among full- and part-time faculty and students, *College & Research Libraries*, 44(2), 141-52.
- Wielhorski, K. (1994), Teaching remote users how to use electronic information resources, *The Public-Access Computer Systems Review*, 5(4), 5-20.
- Yee, M.M. (1991), System design and cataloguing meet the user: user interfaces to online public access catalogs, *Journal of the American Society for Information Science*, 42(2), 78-98.