

Matchmaking and privacy in the digital library: striking the right balance

David M. Nichols and Michael B. Twidale

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

{dmn | mbt}@comp.lancs.ac.uk

Abstract

Although searching for documents is the dominant paradigm in information retrieval it is often more effective to ask another person. A library can serve as a matchmaker that introduces users with common interests to each other. Several methods of matchmaking are described and their implications for privacy are discussed. An approach to prototyping matchmaking functionality is outlined.

Introduction

Although the dominant approach in information retrieval has been that of searching for documents the complementary approach of collaborative browsing (Twidale and Nichols 1996) emphasises the role of people in the information search process. In making the move towards Digital Libraries (DLs) it is easy to concentrate on the technological aspects of scanning, indexing, searching etc. at the expense of the social relationships that libraries support.

People are often the most valuable resource for information searchers - whether as subject librarians who can advise on how to use a particular database or colleagues who can filter the literature and select the 'key paper' (Bates 1979; Grosser 1991; O'Day and Jeffries 1993). These interactions have traditionally taken place on a face-to-face basis and are in danger of becoming rarer as the DL encourages more asynchronous work.

In this paper we examine the role of libraries in supporting the introduction of users to others with similar interests. We consider the computational and social implications of such a system and discuss how to investigate novel DL functionality.

Matchmaking

In both physical and digital libraries, finding library staff is considerably easier than finding other users who may have a common interest in your subject area. Libraries recognise that their users may have difficulties in finding information and clearly indicate the availability of human assistance. Finding people with similar interests, although an activity which is often carried out *via* libraries, is not one that is explicitly *supported* by libraries.

To illustrate our aims consider the following rather utopian scenario. You work in a small research institution/university with a well stocked library. You are browsing the shelves when you spot a colleague from a different department looking at books in 'your' section of the library. You strike up a conversation and discover that this colleague is doing some research in a new area for her. She discovered a reference to a book and was in the process of tracking it down. You discover that you have a mutual interest in the topic, although you are viewing from the perspective of two quite different disciplines. Amusingly you discover that in these two disciplines the topic has two quite different names, so you would not have known about your mutual interests from reading the organisation's staff information guide. You decide to share resources and tackle the problem together.

The utopian nature of the scenario is at least partly due to the constraints of the 'functionality' provided by the physical library and which potentially can be addressed far better in a digital library:

- you both need to be in the same place at the same time in order for the serendipitous meeting to occur.

- you need to be prepared to strike up a conversation with someone in a library: a location where you meant to be quiet.
- if you do not know the person very well, you must be prepared to strike up the conversation in rather odd circumstances, even though such a conversation would be quite normal in other circumstances, such as at a conference reception.
- you may need to have resources to hand (be carrying certain photocopies, copies of things you are writing) in order to illustrate your ideas or provide the other person with some useful references

However in order to realise the potential of such a scenario, we clearly need to provide our digital library with suitable functionality.

Current methods of locating people who can provide us with useful information include reading their work, meetings at conferences, Internet searches, tracking down personal web pages, noticing contributions to mailing lists, recommendations from third-parties etc.

There are several reasons why the introduction of people with similar interests, matchmaking, may be a useful service for a DL:

- people working on similar projects may be unaware of each other (Foner and Crabtree 1996) - particularly if they have a background in different subject areas.
- other users may be useful filters and sources of recommendations (Nichols, Twidale and Paice 1997) that can help to prevent the continuous 're-invention of the wheel'.
- some users, especially those new to a field, are not 'plugged-in' to the word-of-mouth in a subject area (Foner and Crabtree 1996).
- some may not have the resources to engage in traditional forms of matchmaking - such as attending international conferences.

There are also several reasons why unconstrained matchmaking may be an undesirable service - these are discussed later.

It is one of the challenges of research in the field of DLs to investigate how to provide functionality that has not been possible before and yet which meets the needs of users. The difficulty is that we can potentially provide facilities that users have not considered as an option (because they were previously unavailable and infeasible). Such facilities may be based on real-world analogues, or involve quite different ways of working. The analogue for 'matchmaking' is that of a researcher describing her work to a colleague in a common room or at a conference and the colleague replying "That's interesting. It isn't my field, but you really ought to talk to X. There seem to be some interesting similarities between your approaches."

Models of Matchmaking

There are several different models of agent introduction - see (Kuokka and Harada 1995) for a review. Most models have an intermediary, a matchmaker, that facilitates communication between the parties. These are analogous to some of the traditional models of marriage arrangement. In examining the role of formal intermediaries in the marriage market (Ahuvia and Adelman 1992) use a three stage model:

- searching: information acquisition
- matching: bringing together compatible exchange partners
- interacting: the formation or rejection of a match

For example, personal lonelyhearts advertisements are part of the first stage whereas dating agencies encompass both of the first two stages, and traditional Orthodox Jewish matchmakers are involved in all three stages.

The information acquisition stage of the model is where the intermediary (or the other party in the case of direct communication/address routing (Kuokka and Harada 1995)) obtains information about an agent. In most cases this involves a candidate providing a description of themselves or their interests. Although this clearly has a cost for the user, supplying the information to an intermediary is much more efficient than repeatedly contacting numerous individuals.

This self-description is subject to several problems including bias, language and cultural differences, subject-specific terminology and, if it is to be processed computationally, the keyword matching difficulties common to information retrieval systems everywhere. A slightly different method of self-description is to use an example-based registration system - "I wish to

register my interest in this particular book because I might be interested in others who are also interested in it."

An alternative, or complementary, approach is to use ubiquitous user data (Hill and Terveen 1996) – data that would be present anyway, independent of the matchmaking functionality. For example, the Yenta matchmaking system has been tested with large numbers of personal email messages (Foner and Crabtree 1996).

In addition to these data-based approaches there are also metadata-based options. Automated collaborative filtering systems (e.g. (Shardanand and Maes 1995)) use ratings supplied by users to identify users with similar tastes. The approach with the lowest 'perceived cost' (at the interface) to the user is to use ubiquitous automatically generated metadata. For example, regarding a catalogue record as a history enriched digital object (Hill and Hollan 1994) that records its usage by different users (Böhm and Rakow 1994) potentially enables those users to become candidates for matchmaking (Nichols, Twidale and Paice 1997). This usage information can be at a fine level of granularity (this user consulted this particular document – or part of a document) which facilitates particularly accurate 'matches'.

The hypothesis is that the usage of real resources (as opposed to self-description) is a useful indicator of users' interests - either as an isolated function or in combination with other approaches. The same reasoning also applies to the 'extended presence' provided by software agents that autonomously search for information on a user's behalf.

A key advantage of using ubiquitous data, and metadata, is the low cost to the user at the information acquisition stage. This transactional cost is an important factor in the success of a cooperative system (Grudin 1994); however, there are other informational costs to be considered.

Most usage information is currently 'forgotten' by database systems - an indication of its perceived value (Koenig 1990). Some usage information can be re-used anonymously but for any matchmaking purposes searchers' identities must be preserved. It is immediately clear that there are serious implications concerning the privacy, security and ownership of information searchers' interactions.

Privacy and Acceptability

Clearly the provision of a matchmaking functionality raises an important set of issues relating to privacy and acceptability. It is easy to imagine naively developing a system that provides very powerful functionality that users agree would in theory be very useful, but which they are reluctant to use themselves because of concerns about privacy. We have considered some of these issues elsewhere (Nichols, Twidale and Paice 1997).

We take as the base point of our consideration that any amount of collaboration regardless of whether it is computer-mediated necessarily involves some loss of privacy. It is easy to envisage circumstances when a user chooses not to tell even trusted colleagues what she is working on, and is willing to sacrifice the potential gains that might ensue from collaboration because of the costs of privacy loss. A concrete example is the case of Andrew Wiles, prover of Fermat's Last Theorem, who did not tell his closest colleagues of the true nature of his research, partly out of fear of ridicule should he fail (Singh and Lynch 1996).

It is important that users have a clear means of controlling the degree to which they make information about their interests and information use activities public. This must include not only the ability to selectively hide certain activities (and to hide that they have anything to hide), but also to control the degree to which information is made available to different layers of the public space. A user may wish to reveal certain things to her research group, slightly less to her department, less again to her organisation and even less to the world at large.

Some privacy concerns can be addressed by the usage mechanism provided. In a similar way that dating agencies warn about the dangers of social introductions to strangers, an interest-matchmaking system could effect a progressive activity of introducing with both sides provided with opportunities of cancelling the process before it is finalised.

A legitimate fear that users may have of a matchmaking system is that although they assent to the loss of privacy necessary to enable it to work, they fear that the data provided may in the future be put to other, unauthorised uses. Consequently they may refuse to use a system, irrespective of the benefit which they are in perfect agreement it would provide, because of fears of signing a blank cheque. Many of these concerns about personal information derive

from a feeling of lack of control caused by the intangibility, ease-of-duplication and black-box nature of computer storage of personal information.

One possible scenario for the interacting stage, using the example-based registration approach mentioned earlier: when another user also registers their interest in the same way, the DL as intermediary, independently informs each user. If both sides agree, additional information (but not the identity) may be provided about the participants (Foner and Crabtree 1996). If for example, one of the people is an eminent Professor and the other an undergraduate, either side may consider that they don't really want to pursue this into a potentially embarrassing non-meeting of minds. Otherwise, if both sides agree the system arranges a suitable virtual meeting, which may be as simple forwarding two email addresses. Alternative options, such as interest-specific chat areas are also possible - such as in the music recommendation system *Firefly* (Firefly Network, Inc.).

The nature of the records in the database will also have an influence on the privacy issues. The preservation of identities and search interactions in a music-oriented databases may not raise many objections. This is not necessarily the case for activities in a database of financial records, patents, DNA sequences or criminal records.

Prototyping

We are investigating the potential of this introduction functionality for a DL by undertaking a paper-based study in Lancaster University library. A selection of books that may be of interest to users from a range of disciplines have a slip of paper inserted allowing the reader to register their interest by using a tear-off slip, to be posted into a box as they leave the Library. The experimenters then act as the intermediary. The advantage of such a paper-based study (in addition to its speed of set-up and low cost) is that the privacy issues are clearer to users than in computer interactions - where the fear and lack of knowledge about potential uses are much more apparent.

The paper-based prototyping approach (Ehn and Kyng 1991) has been used with considerable success in participatory design. There it has been used as a way of exploring the functionality of advanced technology at the very early stages of the design and requirements capture process. In such circumstances, it is difficult for conventional techniques to ensure a good match between the current practice of real end-users, the hypothetical features that could be designed for the new system and the potential practice of end-users if they had these new features.

Conclusion

Much of the discussion of digital libraries has focused on providing the same functionality as that in physical libraries, but enabling greater access to scarce resources. However in addition to this vital basic functionality, we believe that it is important to consider how quite new functionalities can be provided that were infeasible in a physical library. This paper has considered just one such functionality, by use of low-cost prototyping methods. At this stage we do not know whether the proposed functionality will prove of use to people or indeed whether it is possible to provide sufficient guarantees of controls with respect to privacy to ensure acceptability. The studies we are currently undertaking with a paper-based prototype will provide some evidence for these issues and will help in determining the features of a working system that will in turn need to be tested.

Acknowledgements

This research was funded by the British Library Research and Innovation Centre project *Investigation of collaborative browsing and the consequences for library systems design*.

References

- Ahuvia, A.C. and Adelman, M.B., 1992, Formal intermediaries in the marriage market: a typology and review, *Journal of Marriage and the Family*, **54** (2), 452-463.
- Bates, M.J., 1979, Idea tactics, *Journal of the American Society for Information Science*, **30** (5), 281-289.

- Böhm, K. and Rakow, T.C., 1994, Metadata for multimedia documents, *SIGMOD Record*, **23** (4), 21-26.
- Ehn, P. and Kyng, M., 1991, Cardboard computers: mocking-it-up or hands-on the future. In Greenberg, J. and Kyng, M. (eds.), *Design at Work*, (Lawrence Erlbaum Associates, Hillsdale, NJ), 169-195.
- Firefly Network, Inc. No date, *Firefly*, [online] Available at <http://www.firefly.com/> [Accessed 14th Mar 1997].
- Foner, L. and Crabtree, I.B., 1996, Multi-agent matchmaking, *BT Technology Journal*, **14** (4), 115-123.
- Grosser, K., 1991, Human networks in organizational information processing, *Annual Review of Information Science and Technology*, **26**, 349-402.
- Grudin, J., 1994, Groupware and social dynamics: eight challenges for developers, *Communications of the ACM*, **37** (1), 92-105.
- Hill, W. and Terveen, L., 1996, Using frequency-of-mention in public conversations for social filtering, *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'96)*, Cambridge, MA, (ACM Press), 106-112.
- Hill, W.C. and Hollan, J.D., 1994, History-enriched digital objects: prototypes and policy issues, *The Information Society*, **10** (2), 139-145.
- Koenig, M.E.D., 1990, Linking library users: a culture change in librarianship, *American Libraries*, **21** (9), 844-849.
- Kuokka, D. and Harada, L., 1995, Matchmaking for information agents, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, Vol 1, Montréal, Canada, 672-678.
- Nichols, D.M., Twidale, M.B. and Paice, C.D., 1997, *Recommendation and Usage in the Digital Library*, Technical Report CSEG/2/97, Computing Department, Lancaster University.
- O'Day, V. and Jeffries, R., 1993, Information artisans: patterns of result sharing by information searchers, *Proceedings of Conference on Organisational Computing Systems (COOCS'93)*, Milpitas, CA, (ACM Press), 98-107.
- Shardanand, U. and Maes, P., 1995, Social information filtering: algorithms for automating "word of mouth", *Proceedings of the Conference on Human Factors in Computing Systems (CHI'95)*, Denver, CO, (ACM Press), 210-217.
- Singh, S. and Lynch, J., 1996, "Fermat's Last Theorem", [online]. London, Horizon, BBC. Available from: <http://www.bbc.co.uk/horizon/95-96/960115.html> [Accessed 14th Mar 1996].
- Twidale, M.B. and Nichols, D.M., 1996, Collaborative browsing and visualisation of the search process, *Aslib Proceedings*, **48** (7-8), 177-182.

Cite as:

Nichols, D.M. and Twidale, M.B. (1997) Matchmaking and privacy in the digital library: striking the right balance, *Proceedings of the 4th UK/International Conference on Electronic Library and Visual Information Research (ELVIRA '97)*. Aslib. 31-38.