

DEBORA: Developing an Interface to Support Collaboration in a Digital Library

David M. Nichols¹, Duncan Pemberton², Salah Dalhoumi³,
Omar Larouk³, Claire Belisle⁴, and Michael B. Twidale⁵

¹ Department of Computer Science, University of Waikato, Hamilton, New Zealand
daven@cs.waikato.ac.nz

² Computing Department, Lancaster University, LA1 4YR, UK
pemberto@comp.lancs.ac.uk

³ École Nationale Supérieure des Sciences de L'Information et des Bibliothèques,
69623 Villeurbanne Cedex, France
{dalhoumi, larouk}@enssib.fr

⁴ LIRE, Unité Mixte de Recherche 5611, CNRS-Université Lyon 2, France
claire.belisle@ish-lyon.cnrs.fr

⁵ Graduate School of Library and Information Science,
University of Illinois at Urbana-Champaign, IL 61820, USA
twidale@uiuc.edu

Abstract. Interfaces to library systems have largely failed to represent the inherently collaborative nature of information work. This paper describes how collaborative functionality is being implemented as part of the DEBORA project to provide access to digitised Renaissance documents. Work practices of users of Renaissance documents are described and the collaborative features of the client software are outlined. Functionalities discussed include annotation, the creation of virtual books and the inclusion of user-supplied metadata.

Keywords: collaboration, interfaces, work practice, Renaissance books

1 Introduction

This paper describes the development of collaborative functionality for users of digital libraries in the context of the EU *Telematics for Libraries* project DEBORA (Digital Access to Books of the Renaissance).

The aim of the DEBORA project is to make Renaissance books more generally available as digital resources and to examine the potential for novel collaborative functionality. The collection being created within DEBORA consists of digitised images of books from libraries in Lyon, Rome and Coimbra.

The first part of the paper outlines the nature of collaboration in digital libraries. Section 3 describes evidence gained from real life users of Renaissance materials. Section 4 describes the implementation of collaborative functionality in the DEBORA client followed by some initial user studies and a conclusion.

2 Collaboration in Digital Libraries

Digital Libraries offer new opportunities for collaboration and communication that were unfeasible in traditional libraries. Interfaces to information systems (including databases, library catalogues and information retrieval applications) have largely reflected single user stereotypes [27]. That is, the activities of other users have had almost no impact on the experience of any one user.

The technology associated with Digital Libraries allows us to consider how to support new ways of working with library materials [13]. Specifically, for users to work in groups, rather than individually, and for users to contribute to collections, rather than simply reading.

2.1 Overview of Collaborative Work for Digital Libraries

Digital Libraries, in comparison with print-based libraries, more easily support the modification of their contents. Several researchers (e.g. [27, 8, 9, 10]) have recognised the potential for users, rather than librarians, to contribute to the development of a collection through user-supplied data (USD) [10]. Such USD can come in many different forms, although it can be split into two main groups: data automatically collected from users' activities and data explicitly generated by users. Implicit additions to a collection include: search term suggestion [10], ratings [19] and 'read-wear' [7]. There have been many proposals for explicit USD: annotation [30, 14, 2], keyword addition [4, 10], evaluative commentary [9, 10], hypertext links [3, 9, 8, 18, 15, 24, 23], ratings [23, 2], error correction [26] – for a review see [27].

The common thread amongst all of these ideas for collaboration is that the actions of one user can in some way be shared with other users *within the system* [27, 13]. In a traditional paper-based library such sharing is more circuitous – via the publishing of cross-cited works that eventually physically arrive in libraries. The belief amongst researchers is that such collaboration will be more productive for the users [13]: for example, by enhancing retrieval effectiveness through community rating of resources [2]. Although many forms of USD have been suggested the most common example is probably *annotation*.

Annotations. Annotation has been frequently proposed as a technique for users to add content (and so share ideas) within information systems – for a recent review see [22]. However, as Wilensky recently notes: “despite its evident usefulness, digital annotation capabilities are not very widespread” [30]. There are several annotation systems in different contexts [22] but the ‘broad territory’ of annotation [15], from free text to metadata, has not been conducive to the development of an accepted standard for annotations [22]. Marshall [15] characterises annotations along seven dimensions: formal/informal, explicit/tacit, permanent/transient, public/private etc. Wilensky [30] suggests 4 requirements for an annotation system; annotations should be:

- Able to be placed *in situ*
- Expressive, extensible and composable
- Format and platform independent
- Free of permission and registration requirements

That annotations should be placed (and viewed) *in situ* accords with evidence from real world studies [14, 15]. The second and third requirements follow from the variability of users, usage contexts and documents; although they also reflect the generality of the *multi-valent document* approach [21] and the desire to support spontaneous ubiquitous collaboration [30].

The fourth requirement, that annotations should not be dependent on prior registration in a system, illustrates a particular perspective on the usage of annotations. In situations where we wish to collaboratively construct cataloguing information using annotations [4] (at the formal end of Marshall's dimension), then registration of users may be necessary to maintain the authority of the metadata.

The scope of the different interpretations placed on 'annotation' cover many of the forms of explicit USD noted earlier. Consequently they share attributes with other novel collaborative functionalities: they present significant added complexity to designers, there is a lack of accepted standards and their adoption would result in significant changes to users' work practices.

2.2 Designing for New Ways of Working

A collaborative digital library that allowed users to contribute and share information around the objects of a collection would be radical change for many users – especially in a domain such as Renaissance texts. As noted in the next section, and by other researchers [20], humanities scholars do not perceive themselves as working collaboratively. Thus merely asking them about their collaborative work is insufficient; that work must also be observed in order to see the substantial 'invisible collaboration', and the trend to increased amounts of collaboration. By studying existing practice and comparing it with evolving practice in other disciplines (especially in the sciences) that have enjoyed better computational resources over a longer period of time, we can explore the design space and create systems that can 'add value' to existing work activities. Our design challenge then is to support new ways of working. We believe that adding collaborative features is a crucial feature, but it must be incorporated in a thoughtful manner. The system must allow different kinds of use, from the currently conventional solitary forms of work, to supporting more effectively existing kinds of collaborative work, to supporting kinds of collaboration wholly new to this group. To be acceptable, the system needs to support graceful (and if necessary, slow) transitions in use along that scale. In this paper we focus on the use of annotations, which initially can be regarded as solely for conventional personal use, but once incorporated into the system allow new additional collaborative benefits.

3 Work Practice and Design Implications

The availability of Renaissance texts has, so far, generally been limited to acknowledged scholars because of the rarity and fragility of the books. Some well-known works have been republished since their first printing, but most of the books are very difficult to access. There are an increasing number of requests for access to 16th century material coming from a variety of users including: educators and their students, linguists, book historians, social and cultural historians ('histoire des mentalities'), specialists in literary studies, illustrators, wardrobe designers etc. The interests, and uses of the material, vary widely with different users: be it the content of the book, the language used, its structure, the illustrations, etc. [5].

This section summarises a usage study undertaken through observations, interviews and questionnaires with users familiar with Renaissance or other old books. The implications of changing the mode of access to the resources are considered, along with their impact on system design.

3.1 Observations of the Use of Renaissance Documents

Although there can be said to be still many 16th century books around, in fact each one is unique: there are almost no two identical copies that have crossed the centuries; even when we have two or more copies of the same text of the same publisher [16]. Each copy will reveal unique information on where, when and how it was printed, through explicit information in the book or through the book's material composition because no two printing processes were exactly the same [17]. The research work done on such book corpuses is basically comparing, confronting different versions, finding similarities and differences, identifying and tracing originality and influences in the written works under scrutiny [25, 29].

Two main groups of users of Renaissance books can be distinguished: book specialists (who may eventually require access to the physical copy) and the larger group of those interested in the content. This second group, where there are users of the existing digitised material on the Internet [29], has working habits that can be summarised in four main areas:

- They need to be able to identify the specific copy of the book they are accessing and to do a very careful reading of the text or a close study of all illustrations and decorative pieces.
- While doing this, they need to take notes, either handwritten or on their computer if they have been allowed to bring it into the library (and if they are allowed to use local electrical outlets).
- Such scholars will find themselves alone, or almost, in their speciality: for example, the study of the 16th century dialect of Lyon in the writings of Paradon (a French regional bishop). Consequently, they will not be prone to exchange information - often because there is no-one to share it with. This may also be related to a university tradition of individual evaluation based on personal, and not collaborative, publications.
- Each scholar usually maintains a personal system of indexed cards where she stores all the information patiently gathered from many trips to libraries and archives, and searches in manuscripts; together with any personal notes. Some scholars now use computers for this task.

3.2. Current Collaborative Practices

There is an increasing recognition of the collaborative aspects of most forms of work and studies of social interactions are revealing the complex interactions involved in group activity [12, 27, 28]. Our expectation is that the benefits potentially available from networked interaction will increase collaboration and modify the manner in which users perform their work.

In responding to our (ongoing) questionnaire, scholars indicate that they exchange information they feel will be important for other colleagues, but rarely have the occasion to collaborate specifically in their field of interest and research. Electronic mail is used widely although written and verbal modes predominate. These collaborations often involve the passing on of information found by chance (serendipitous finds).

Collaboration in real time (synchronous conferencing) is not seen as an essential function. Answers focus on a more restricted sense of collaboration; finding out how the work of colleagues is going and asking information on certain aspects of their own work. A majority of scholars are willing to share their notes with colleagues and, of those who intend to collaborate, most express a desire for these tools to be integrated into the interface of the access software.

3.3 Metadata

Image collections, such as DEBORA, have a problem with a lack of detailed metadata for resource description. At the level of the book the librarians in the project are supplying typical metadata according to the MARC standard. Beneath this there is the internal structure of a book, specific to 16th century texts: the location of indices, prefaces etc. However for effective retrieval using conventional searching the detail of individual pages (illustrations, decorative elements etc.) is needed to allocate indexing terms. Although some basic structuring, such as the differentiation of illustrations from normal text, can be achieved with image analysis tools, most detail must be contributed by specialists in Renaissance texts. For most image collections, DEBORA included, this amount of effort is infeasible. It is at this level, of detailed page-specific metadata, that collaborative contributions by the users of image collections could prove most valuable.

3.4 Design

Based on the above findings, we focus on annotation features as a mechanism for exploring collaborative functions. Annotation is already a part of existing solitary scholarly practice, and so potential users are most likely to be willing to bother to learn to use the system in order to obtain the benefits of a familiar kind of working. This acts as our ‘Trojan Horse’ for studying new collaborative features that build on the use of those same annotations. The first design task is to explore the kinds of annotations that scholars find useful, by a conventional mechanism of iterative prototyping. Although based on conventional personal annotation, the system architecture is intended to support additional collaborative features.

4 The DEBORA Client

DEBORA is based around a client-server architecture with two distinct types of server. A Z39.50 based server is used for the storage of ‘official’ metadata – including the location of the catalogued images. In common with many other annotation systems a separate server is used to store and retrieve user annotations [22]. The client has two main functions: to provide access to the images of the collection and to support the collaborative functionality of the system.

4.1 Client Interface

The DEBORA client is implemented in Java – as the expected mode of access to the image collections will be via a cross-platform network. The basic image viewing window of the client contains several image viewing tools (magnification, brightness, contrast etc). Fig. 1 shows an annotation attached to a rectangular area of a document – here the text of the annotation is shown as a tool tip. Annotations are currently free-text, as opposed to the structure of thesaurus terms [2]. The personal-public dimension of an annotation (as specified by the author) is indicated by colour and can also be used to restrict viewing to a subset of all of the annotations.

The client also provides facilities for the highlighting of areas of an image in a variety of colours – typical of annotations on paper [14, 15]. User-definable workspaces are provided to allow users with similar interests to structure their collaboration activities. Alternatively a user can restrict their additions to the system to be personal and private.

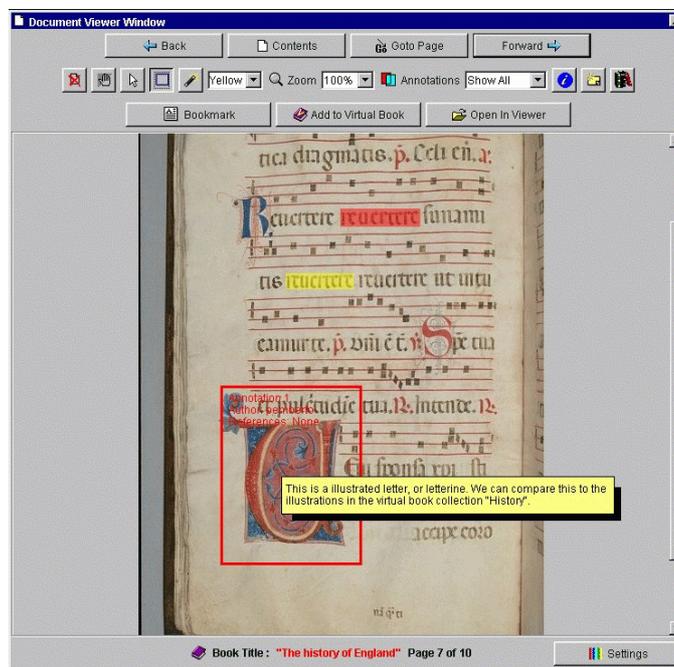


Fig. 1. The DEBORA client interface showing annotation and highlighting

Any set of annotations can be chained together to provide trails, or paths, [3, 24, 23]; following a path may involve moving between any parts of any of the books in the collection. This hyperlinked set of annotations and associated images can be gathered together by a user to create a *virtual book*. We expect that this aggregation will help to reduce the adverse navigational effects of traversing a trail that spans many collection items.

In Fig. 2 the virtual books are shown in the lower left corner, beneath ‘real’ books from the collection. A user creates a virtual book by selecting elements (such as pages) from existing resources and arranging them in ‘virtual chapters’. The resulting composite virtual document [18] can then be represented as part of the collection to other users. A typical scenario for such usage could be a professor tracing the historical development of an artistic style and collating examples into a virtual book for her students (this is typical educational activity [24]).

The client currently displays the virtual books as separate to material in the main collection – however, if we are to take the promise of the digital library seriously then these books should be seen as equal in status to the main collection. Extending the parts of a virtual book to include items from other collections (or the Web) and the issues of generating metadata for such composite documents imply that an extensible notation such as XML should be considered [22].

4.2 Collaboration and Metadata

In addition to shared annotations and virtual books there are at least three other methods for USD to enhance the metadata of the DEBORA collection.

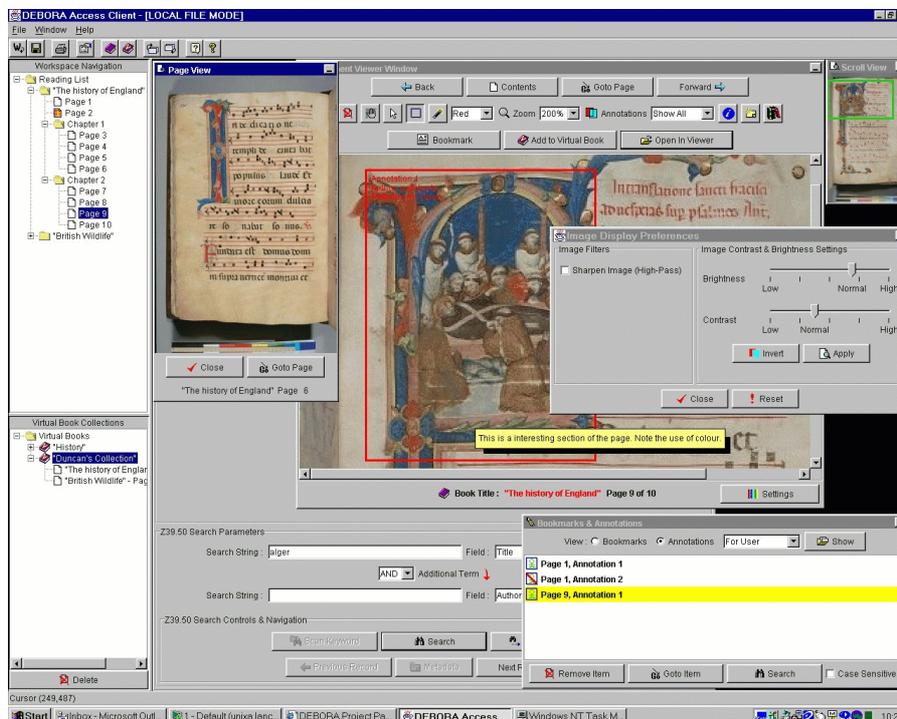


Fig. 2. The DEBORA interface showing the creation of a *virtual book* (bottom left)

Error Correction. Most users of databases have no way to record the presence of errors they detect in the descriptions of items. The client currently supports one-click boolean ‘error-present’ actions and allows users to suggest replacements for descriptions they consider incorrect. With a population of active users any data quality effort could be made more productive by considering those items with most reported errors [26], or *error-wear* by analogy with *edit-wear* [7].

User-Supplied Metadata. In addition to correcting existing metadata, the client can accept new keywords from users. At present these terms are not easily integrated with metadata on the main DEBORA server but this is a small technical problem. A simple interface is all that is required to allow human authorisation of additions to the metadata.

Re-purposing Annotations. The annotations are stored separately and so can be easily searched independently of the collection metadata. Annotation databases represent potentially valuable sources of text [11, 22]; particularly so in the case of image databases such as DEBORA. Golovchinsky et al. [6] use the text identified by freeform annotations as a source of query terms. Conversely, it may be possible to mine user annotations as a source of indexing terms – where otherwise the images would not have any associated text that could be searched. ‘Multimedia annotations ... are simply meta-data associated with multimedia content’ [1].

A major difference between conventional image annotation and this approach is the purpose of the annotation: collaborative annotations are not *intended* to describe the images but to share

information. Although this re-purposing of annotations will generate index terms of lower accuracy than those of an expert cataloguer, we believe it will be better than their complete absence.

Metadata Authority. When user supplied data is used in conjunction with ‘official’ descriptions then search tools need to be aware of the difference in authority that should be attached to the terms [4]. One approach would be to attach less weight to USD terms in matching queries. If and when USD is accepted by collection maintainers then this increased trust could be reflected by increasing the weights of other USD terms from the same user [26].

5 Initial User Studies with the Client Interface

The DEBORA client has by now been scrutinised and explored by several 16th century specialists and has yielded information on its current functionality:

- The virtual document concept where each user creates his or her own document by selecting pieces of one or several books is considered as essential. Resulting from a possible appropriation of one or more digitised documents, the virtual book is perceived as answering important and specific user needs and is seen as very close to the actual way of working with physical documents.
- The possibility of having on the same screen two pages side by side or two illustrations taken from two different copies of the same book facilitates comparison. This functionality is particularly appreciated by those specialists that need to authenticate an incomplete or dubious copy or to restore one of the copies.
- The possibility of reorganising the different pieces of a text is also greatly appreciated, especially since it can be exchanged and used in collaborative work.
- The zoom function by which the content of the different page images can be rapidly discovered or the simultaneous presentation of the different pages is seen as very useful. It allows a rapid assessment of the content of the different pages presented and also is an effective replacement of leafing through a book. The possibility of rapidly going back or forward facilitates memorising and global comprehension of the document on the screen.
- Users working on illustrations or interested in typography suggested having a command (button) showing the actual size (1:1) of the illustration or the fonts. These specialists also appreciated the highlighting facility that augments the contrast and sharpness necessary in their tasks. Some expressed fear that digitising books would result in a clean image that would not be sufficiently faithful to the original.
- Annotation is appreciated as a fundamental functionality allowing collaboration and exchanges between users. The three levels provided, private annotations, annotations shared within a specific group and annotations shared publicly with anyone, all have their specific interest. But users expressed concern about the management of these annotations once they begin to pile up. How will the different types of annotations be recognised? How will they be presented if there are many for the same page?
- Confronted with this new tool, users felt disoriented, particularly by its complexity. There is a need for a simplified version of the interface, at least for first time users and for a more expert version for experienced users.
- As had already been observed during studies for 16th century specialist work, the collaboration dimension goes beyond simple manipulation and working on digitised text. Users see the collaboration field as possibly including *any* communication on the different re-

search paths and results. The interface would then be expected to allow links to other communication tools.

6 Conclusions

The paper considers how collaborative features can be added to a digital library. This creates several problems, not least of which is that the intended users do not perceive themselves as working collaboratively (even when they do), and so are unlikely to see why having additional collaborative features would help them, or why they should bother to learn how to use them. We are exploring the provision of annotation features as a mechanism to support a graceful transition from solitary use to new kinds of collaborative working.

Future work will involve a continuation of the refinement of the basic annotation features for conventional use (starting with the user feedback), along with an exploration of the additional collaborative use of these same annotations – a re-purposing of annotations. Examples include information sharing, recommending and the exploration of the utility of collaborative annotations as index terms.

It is important to explore how the features of a digital library can be more closely integrated into users' work activity. Thus conventionally, rare texts and annotations of those text were separate. In DEBORA we have brought them together. We can consider how to further incorporate other aspects of work, such as the multiple organising and interlinking of those annotations (as in their card index equivalents), the creation of virtual books, and also incorporating the texts that the users of the digital library write themselves, based on their study in the library. Such a richer, more integrated environment will not only promote easier switching between its components, but it will itself afford new forms of collaboration, and in a networked digital library mean that the user's office then becomes as mobile the Renaissance books themselves have become.

Acknowledgements

The DEBORA project (5608) is funded under the EU *Telematics for Libraries* program.

References

1. Barger, D.M., Gupta, A., Grudin, J., Sanocki, E., Li, F.: Asynchronous Collaboration Around Multimedia and its Application to On-Demand Training. Microsoft Research Technical Report 99-66. (1999)
2. Bouthors, V., Dedieu, O.: Pharos, a Collaborative Infrastructure for Web Knowledge Sharing. In Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries. Springer Verlag, Berlin (1999) 215-233
3. Bush, V.: As We May Think. *The Atlantic Monthly*. 6 (1945) 101-108
4. Chandrinou, K., Immerkær, J., Dörr, M., Trahanias, P.: A Visual Technique for Annotating Large-Volume Multi-Media Databases – A Tool for Adding Semantic Value to Improve Information Filtering. In Proceedings of the 5th DELOS Workshop: Filtering and Collaborative Filtering. ERCIM, Le Chesnay, France (1998) 125-129
5. Chartier, R.: *Les Usages de L'Imprimé*. Fayard, Paris (1986)

6. Golovchinsky, G., Price, M.N., Schilit, B.N.: From Reading to Retrieval: Freeform Ink Annotations as Queries. In Proceedings of the 22nd Annual ACM Conference on Research and Development in Information Retrieval (SIGIR'99). ACM, New York (1999) 19-25
7. Hill, W.C., Hollan, J.D.: History-Enriched Digital Objects: Prototypes and Policy Issues. *The Information Society*. 10 (1994) 139-145
8. Kantor, P.B.: The Adaptive Library Network Interface: A Historical Overview and Interim Report. *Library Hi Tech*. 11 (1993) 81-92
9. King, G., Kung, H.T., Grosz, B., Verba, S., Flecker, D., Kahin, B.: The Harvard Self-Enriching Library Facilities (SELF) Project. In Proceedings of Digital Libraries '94 (DL'94). Texas A&M University (1994) 134-138
10. Koenig, M.E.D.: Linking Library Users: a Culture Change in Librarianship. *American Libraries*. 21 (1990) 844-849
11. Lawton, D.T., Smith, I.E.: The Knowledge Weasel Hypermedia Annotation System. In Proceedings of the Fifth ACM Conference on Hypertext. ACM, New York (1993) 106-117
12. Leplat, J.: L'Analyse Psychologique du Travail: Quelques Jalons Historiques. *Le Travail Humain*. 56 (1993) 115-131.
13. Levy, D., Marshall, C.C.: Going Digital: a Look at the Assumptions Underlying Digital Libraries. *Communications of the ACM*. 38 (1995) 77-84
14. Marshall, C.C.: Annotation: from Paper Books to the Digital Library. In Proceedings of the 2nd ACM International Conference on Digital Libraries. ACM, New York (1997) 131-140
15. Marshall, C.C.: Toward an Ecology of Hypertext Annotation. In Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia. ACM, New York (1998) 40-49
16. Martin, H.: Histoire et Pouvoirs de L'Écrit, Bibliothèque de L'Évolution de L'Humanité. Albin Michel, Paris. (1996)
17. Martin, H.: La Naissance du Livre Moderne, XIVE – XVIIe Siècles. Electre, Paris (2000)
18. Myaeng, S.H., Lee, M., Kang, J.: Virtual Documents: a New Architecture for Knowledge Management in Digital Libraries. In Proceedings of the Second Asian Digital Libraries Conference. (1999)
19. Nichols, D.M.: Implicit Ratings and Filtering. In Proceedings of the 5th DELOS Workshop: Filtering and Collaborative Filtering. ERCIM, Le Chesnay, France (1998) 31-36
20. Palmer, C.L., Neumann, L.J.: Exploration and Translation: the Research Work of Interdisciplinary Humanities Scholars. *Library Quarterly* (to appear)
21. Phelps, T.A., Wilensky, R.: Multivalent Annotations. Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries. (1997)
22. Ovsianikov, I.A., Arbib, M.A., McNeil, T.H.: Annotation Technology. *International Journal of Human-Computer Studies*. 50 (1999) 329-362
23. Röscheisen, M., Morgensen, C., Winograd, T.: Beyond Browsing: Shared Comments, SOAPs, Trails, and On-line Communities. *Computer Networks and ISDN Systems*. 27 (1995) 739-749
24. Shipman, F.M., Furuta, R., Brenner, D., Chung, C., Hsieh, H.: Guided Paths through Web-Based Collections: Design, Experiences, and Adaptations. *Journal of the American Society for Information Science*. 51 (2000) 260-272
25. Sordet, Y.: Repérage et Navigation dans L'Espace du Livre Ancien. Communication Présentée au 1er Forum de L'Édition et de la Documentation Spécialisé (Docforum). (1997)
26. Twidale, M.B., Marty, P.F.: An Investigation of Data Quality and Collaboration. Technical Report ISRN UIUCLIS--1999/9+CSCW. Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. (1999)
27. Twidale, M.B., Nichols, D.M.: Computer Supported Cooperative Work in Information Search and Retrieval. In Williams, M.E. (ed.): Annual Review of Information Science and Technology (ARIST), Vol. 33. Information Today, Inc., Medford, NJ (1998) 259-319
28. Twidale, M.B., Nichols, D.M., Paice, C.D.: Browsing is a Collaborative Process. *Information Processing & Management* 33 (1997) 761-783
29. de Ventabert, G.: Représentation et Exploitation Électroniques de Documents Anciens (Textes et Images). *Document Numérique*. 3 (1999) 57-73
30. Wilensky, R.: Digital Library Resources as a Basis for Collaborative Work. *Journal of the American Society for Information Science*. 51 (2000) 228-245

A revised version appears as:

Nichols, D.M., Pemberton, D., Dalhoumi, S., Larouk, O., Belisle, C. and Twidale, M.B. (2000)
DEBORA: Developing an interface to support collaboration in a digital library, *Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2000)*.
Springer-Verlag. 239-248.

http://dx.doi.org/10.1007/3-540-45268-0_22