Personalisation and recommender systems in digital libraries

Alan F. Smeaton[1]**, Jamie Callan**[2]

[1] Centre for Digital Video Processing, Dublin City University, Glasnevin, Dublin 9, Ireland
e-mail: Alan.Smeaton@dcu.ie
[2] School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213-8213, USA
e-mail: callan+@cs.cmu.ie

Abstract. Widespread use of the Internet has resulted in digital libraries that are increasingly used by diverse communities of users for diverse purposes and in which sharing and collaboration have become important social elements. As such libraries become commonplace, as their contents and services become more varied, and as their patrons become more experienced with computer technology, users will expect more sophisticated services from these libraries. A simple search function, normally an integral part of any digital library, increasingly leads to user frustration as user needs become more complex and as the volume of managed information increases. Proactive digital libraries, where the library evolves from being passive and untailored, are seen as offering great potential for addressing and overcoming these issues and include techniques such as personalisation and recommender systems. In this paper, following on from the DELOS/NSF Working Group on Personalisation and Recommender Systems for Digital Libraries, which met and reported during 2003, we present some background material on the scope of personalisation and recommender systems in digital libraries. We then outline the working group's vision for the evolution of digital libraries and the role that personalisation and recommender systems will play, and we present a series of research challenges and specific recommendations and research priorities for the field.

Keywords: Digital libraries – User interaction – Personalisation – Recommender systems

1 Overview of personalisation and recommender systems

Our definition of digital libraries is that they are collections of information that have associated services which are delivered to user communities using a variety of technologies. The collections of information can be scientific, business, or personal data and can be represented as digital text, image, audio, video, or other media. The information can be digitised paper or born digital material, and the services offered on such information can be varied, ranging from content operations to rights management, and can be offered to individuals or user communities. Good reviews of digital libraries and pointers to digital library resources can be found in [53, 57].

Internet access has resulted in digital libraries that are increasingly used by diverse communities for diverse purposes, and in which sharing and collaboration have become important social elements. As digital libraries become commonplace, as their contents and services become more varied, and as their patrons become more experienced with computer technology, people expect more sophisticated services from their digital libraries. A traditional search function is normally an integral part of any digital library, but user frustrations with this increase as needs become more complex and as the volume of information managed by digital libraries increases. Thus digital libraries must move from being passive, with little adaptation to individual users, to being more proactive in offering and tailoring information for individuals and communities and in supporting community efforts to capture, structure, and share knowledge. Digital libraries that are not personalised for individuals and/or communities will be seen as defaulting on their obligation to offer the best service possible. Just as people patronise stores in which they and their preferences are known and their needs anticipated, so too will they patronise digital libraries that remember them and anticipate their needs.

We define personalisation as the ways in which information and services can be tailored to match the unique and specific needs of an individual or a community. This is achieved by adapting presentation, content,

and/or services based on a person's task, background, history, device, information needs, location, etc., essentially the user's context. Recommender systems are something slightly different and can often be regarded as a particular type of personalisation that learns about a person's or a community's needs and then proactively identifies and recommends information that meets those needs. Recommender systems are especially useful when they identify information a person was previously unaware of. Personalisation can be user-driven, which involves a user directly invoking and supporting the personalisation process by providing explicit input. Examples of this include systems like MyYahoo ! [31, 62], GroupLens [26], the ACM Digital Library [43], PersonalDJ [21], and MovieLens [37], where the user explicitly initiates actions and provides example information in order to control the personalisation. Personalisation can also be completely automatic, where the system observes some user activity and identifies the input used to tailor some aspect of the system in a personalised way [24, 27]. These two examples of user-driven and automatic personalisation are at the extreme ends of the spectrum, and many personalisation tools will have elements of both approaches such as TechLens+ [55] and the FÍSCHLÁR System [50].

Although recommender systems can be viewed as just another type of personalisation, they are a particularly visible form of personalisation that has attracted a distinct research community. Throughout this report we use the phrase *personalisation and recommender systems* to indicate a wide range of individual and community-based personalisation techniques.[1]

Personalisation systems have had great success in areas besides digital libraries. For example, in the area of targeted advertising we see tailored advertisements in the output pages from almost all Web search engines. When using online retail systems such as Amazon.com we are given suggestions for additional complementary services and products. When using a WAP handheld device, the presentation arrangement of menu options is personalised and tailored for different users. Adaptive hypermedia systems [1, 10, 12] demonstrate how personalisation can be used to assist a person in navigating through large, online systems such as the Web or closed collections like courseware.

If we consider the history of digital libraries, we should be conscious that substantial digital libraries have been in place and operating since long before the term digital library became popular in the early to mid 1990s. These include commercial systems as well as university and government systems, but thus far personalisation has had only a limited impact on digital libraries. The initial focus in digital library research was on increasing the availability of digital content and on creating and rolling out basic digital library services e.g. the Alexandra project on

geospatially referenced multimedia material, which ran from 1994 to 1999 [19, 51]. Much digital library research and development activity to date has concentrated on the complete digitisation process, covering things like metadata standards such as Dublin Core [38, 59], on automatic extraction of metadata [23], interoperability [35, 39, 52], rights management ([46] for some older work and [15, 45] for more recent work), and object identification [4]. Many applications of personalisation in digital libraries such as MyLibrary [29], the ACM Digital Library [43], and teaching/learning applications [33, 36] have thus far focused on applying basic personalisation and rudimentary recommender systems in a reasonably straightforward way, and to date such applications of personalisation have not really added much value to the digital library and certainly not done much to bring the digital library up to the next level.

Many fields are contributing to the development of personalistion in the area of digital libraries including information retrieval [6], human-computer interaction [18], computer-supported collaborative work [58], machine learning, user modelling (especially in the context of information seeking [7, 25, 48, 61]), hypermedia, and information science, to name a few. To date, these fields have not had a great history of collaboration or of working together. This needs to change, as in order to realise the potential of digital libraries, we need to incorporate personalisation in a major way, and in order to develop research in the area of personalisation, we need to bring together the many multidisciplinary fields which contribute to its development.

The work reported in this paper is the culmination of work done by a working group set up to examine the role and future prospect for personalisation and recommender systems in digital libraries. The working group was jointly funded by the DELOS EU FP5 Network and the National Science Foundation International Digital Libraries Initiative, and in this paper we report the findings of that group. This paper is structured in the following way. In this section we have presented some background on our scope and definition of what we mean by personalisation and recommender systems. We now follow with an outline of our vision for the evolution of digital libraries and our perspective on personalisation in such libraries. This is followed by a series of research challenges which we believe need to be met in order to realise the true potential of digital libraries. Finally, we conclude with a series of specific recommendations and research priorities.

2 Vision of personalisation and digital libraries

A digital library is defined as a set of collections, services, user community, and supporting technology. Formal large-scale research programmes on digital libraries began about a decade ago with the initial NSF/DARPA Digital Libraries research programme [19], and a number

---

[1] This phrase has its roots in a workshop that we chaired in 2001 [14] and the DELOS Working Group that we chaired in 2003.

of digital libraries were created as a result in domains like scientific data management [34], geospatial data management [51], and television news [16, 54]. Much of the research during the initial stages was on digitising existing sources, creating large-scale collections, devising technological solutions, and providing simple forms of access. The first generation of digital libraries derived from this research provided a small set of services to relatively well-prepared and knowledgeable user communities.

The emerging generation of digital libraries is more heterogeneous along several dimensions. The collections themselves are becoming more heterogeneous, in terms of their creators, content, media, and communities served. The range of library types is expanding to include long-term personal digital libraries [22], as well as digital libraries that serve specific organizations, educational needs, and cultural heritage and that vary in their reliability, authority, recency, and quality. The user communities are becoming heterogeneous in terms of their interests, backgrounds, and skill levels, ranging from novices to experts in a specific subject area. The growing diversity of digital libraries, the communities accessing them, and the use to which information is put requires that the next generation of digital libraries be more effective at providing information that is tailored to a person's background knowledge, skills, tasks, and intended use of the information.

As computers have become common business, educational, and personal tools, long-term personal digital libraries are becoming commonplace. Children begin using computers regularly for education and entertainment by the time they are 10 and will continue to use them for an increasingly complex set of tasks throughout their lives. The information that a person accumulates during a lifetime of computer usage is a personal digital library, which makes *everyone* both a user and a creator of digital libraries. Examples of systems which try to do this are MyLifeBits from Microsoft [22] and Haystack from MIT [2]. People will want to save their pictures, music, educational and professional materials, and personal and other information throughout their lives, but their needs, abilities, and computing platforms will change. People will need personal digital libraries that help integrate information gathered and organised by the 10-year-old with information gathered and organised by that same person at 20, 30, 40, and beyond. Long-term modelling of a person's evolving interests, preferences, knowledge, goals, and social networks will be required to help people manage their personal digital libraries during a lifetime of use, and this information must transcend specific systems, which will change often. The information that a person acquires during a lifetime, how the person organises it, the tasks for which it is used, and the people with whom it is shared – all paint a detailed picture of a person, but little is known today about how to use this information effectively. Interpreting the trails of a lifetime of computer usage, across the many different tools and resources involved, is an extraordinarily challenging and complex problem.

Digital libraries will also be affected by a trend towards mobile devices that have computational power similar to that of desktop machines, are wirelessly net-connected, and have built-in positioning systems such as GPS. In ubiquitous computing (ubicomp) (see [60] for an early vision), access to information sources such as digital libraries is not only possible in many locations but is dynamically adaptive with regard to a person's location as well as the artefacts and interaction devices and other people in those locations. For example, when visiting a museum, one might automatically be presented with information selected from the museum's digital library as one moves through the galleries. The user's motion towards an exhibit would be taken as an implicit cue for retrieval and recommendation of information, based on the exhibit and the previous activity and preferences of the user. Moving out of the museum to explore the surrounding city might trigger access to other information resources in a way that is responsive to the user's recent museum visit. Ubicomps adaptation to the shifting context of use, in order to personalise and contextualise information access, has a side effect of binding together multiple digital libraries. Adaptive ubiquitous computing, which includes mobile digital libraries [32], demands the integration of information that is heterogeneous with regard to ownership by or containment in different digital libraries.

The ubiquity of computing and telecommunications devices also means that communication and sharing of information is afforded to a large proportion of the public, and hence that digital libraries can become less centralized and controlled. Greater community sharing of and interaction through digital libraries is another trend that must be addressed, and this trend makes its presence felt at both technological and social levels. With regard to technology, sharing and access can be carried out not just through institutional digital libraries but through large and dynamically changing peer-to-peer networks. Current digital libraries have just begun to address the issues of provenance, subjectivity, and consistency that come to the fore here. These issues have both negative and positive aspects. For example, having many such sources of information may lead to complex heterogeneity and inconsistency, but also innovation and personal contribution to the shared information resource. With regard to social aspects of sharing, there is great potential for community building and interpersonal interaction.

Both institutional and community digital libraries can serve as meeting places where people can communicate with each other through the documents, annotations, and logs they make available to each other, and through the conversation and discussion around this shared information. Again, there are both negative and positive aspects to consider, balancing invasiveness and privacy with sharing and collaboration. As with all forms of social com-

munication, the same contribution may be considered as useful and novel by one person and as annoying and offensive by another. Since personalisation involves not just the isolated individual but the individual as a social actor, digital library research will have to be both socially mature and technically innovative as it steps up to play its part in the wider environment of public discourse, community, and culture.

2.1 The future of personalisation in digital libraries

First-generation digital libraries were created for people whose information needs were well defined and well matched to the digital resources they contain. They assume relatively homogenous and possibly well-informed users and relatively accurate descriptions of information needs. These characteristics limit their impact on the wider society.

Personalisation is required to make an increasingly heterogeneous population of digital libraries accessible to an increasingly heterogeneous population of users. It is no longer realistic to expect every user to adapt to every digital library. If a person must be an anthropologist to use an anthropological digital library, the library is available to only a limited community; but if the library can tailor its services and materials for a wider range of users, the impact and utility of the library is magnified greatly. The next generation of digital libraries must provide a wide range of personalised services that support the activities of a wide range of users.

Early research on digital library personalisation used simple models of user interests to make individual recommendations. Future digital libraries need to feature broad user models, including a person's background, knowledge, tasks, social activity, and preferences, in a manner similar to that in which this is being addressed in information retrieval [25]. Moreover, ubiquitous computing requires digital libraries to adapt to various parameters related to the context of a person's work. Finally, the need to support communities of users requires extending individual user models with group and community models.

2.2 A wide range of personalisations

Digital libraries can be personalised in many different ways to support many different purposes and types of people and many types of tasks. As illustrated in Fig. 1, the personalisation can be based on different types of characteristics such as characteristics of: a person as both an individual or member of a group (e.g. knowledge or motivitation); the resources or information or documents (e.g. genre of the materials, age, or authenticity; and/or perceived outcome (e.g. novelty or accuracy), all of which are related to the media or channel used (e.g. PDA vs. a cell phone or computer), the task being performed, and the environment in which the user is immersed, namely
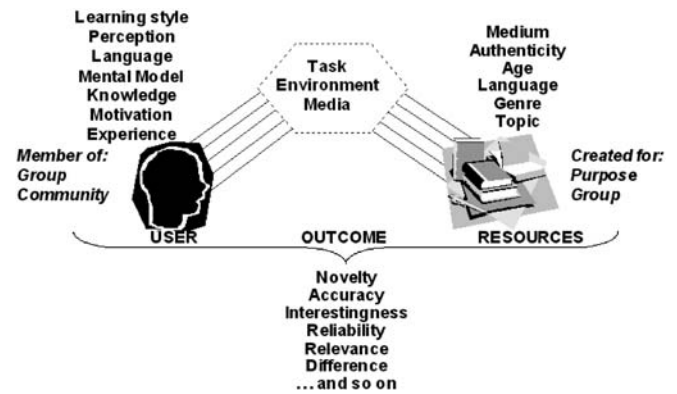

Fig. 1. A wide range of personalisations

the context [8, 9, 44]. This list of characteristics emphasizes short-term personalisation and is not comprehensive, but it highlights the relationships between the most important components – namely people, resources, and perceived outcomes – and serves as a guide to illustrate the rich types of data that are available and need to be manipulated to personalise and/or recommend.

2.3 Potential applications of personalisation and recommender systems

Digital libraries that support a broader range of information-seeking activities, build detailed models of users and user communities, and can tailor information for a wide range of uses will enable new types of software applications designed to support a variety of information-seeking, building, and sharing activities. One justification for this drive is that information-seeking activities need to extend well beyond the classic ad hoc search that is the main access method in the current generation of digital libraries. A few recent examples show that information services can adaptively support diverse information-seeking activities: writing aids automatically suggest related and supporting materials from personal or external digital libraries such as "Writer's Aid" [5], peer-help systems that use information about the tasks and knowledge of individuals to suggest collaborators with specific skills, and adaptive hypermedia systems guide students to the most relevant items in an educational digital library [33, 36].

In the future it will be routine for applications to draw upon and integrate materials from multiple digital libraries and to use long-term user histories to help personalise this material; such systems are beginning to emerge, but the difficulty of integrating material from multiple sources and utilizing long-term user history for effective personalisation makes them expensive to build. Digital libraries that explicitly tailor delivered information for specific uses will simplify such integration as well as support new types of applications. A tutoring system for a second language or that provides reading practice designed to address specific reading comprehension problems based on a long-term model of a user's language

learning will need digital libraries that can provide materials that satisfy very specific and detailed user requirements. Integrating government digital libraries with citizen discussion groups will support more informed debate about public policy, especially if the evolution of that debate is incorporated into the personalisation. To be effective, government digital libraries will also need to bridge the gap between the language of administrators and bureaucrats and the language of ordinary citizens. Lifelong learning services will take a specific information need, interpret it in the context of a person's user model as it changes over time, and create a personalised learning plan that spans multiple digital libraries. These, however, are just examples, and there are many other potential examples of cross-digital library personalisation.

3 Research challenges

In order to realise the vision of digital libraries described to date, with all of its inherent personalisation, there are many research challenges to be addressed and overcome, some of which demand research into new directions. In this section we outline some of those challenges.

3.1 Modelling users

To date personalisation has been inhibited by limited user modelling that reflects an overly simplistic representation of users and their information-seeking behaviour. Current user models draw on a limited set of parameters, but people, jobs, and workplaces are much more complex. More realistic user models should take into account the overall information space and the context [8, 9, 44] including:

– Cognitive abilities, e.g. learning styles, perception;
– Individual differences, e.g. experience, education, age, gender;
– Individual and group behaviour patterns and history;
– Subject domains, e.g. engineering, arts, health;
– Work tasks, e.g. writing an essay, choosing a movie, planning a holiday;
– Work environments, e.g. university, hospital, business office, home; and
– How all of the above change over time.

Information seeking encompasses elements of all the above, and for personalisation some or all of these elements will come into play, and this would need to extend the current approaches to building complex user models taken in other areas besides digital libraries [30, 40]. Furthermore, individuals are members of different types of social groups, forming information communities, which adds to the complexity of model building.

Currently both explicit and implicit methods of learning about users have been used including explicit questionnaires and implicit transaction logs, but new techniques for data collection and analysis need to be developed for building more-useful long-term user models.

The challenges in building user models are multidimensional. Fundamental questions need to be addressed, such as:

– What data can and should be collected?
– How can the data be captured?
– How should the data be analysed?
– What parameters need to be set?
– How are anomalous data recognized and filtered out?
– How are data weighted appropriately over time?

Once these questions have been addressed, more-meaningful and appropriate user models must be developed to better inform the application of personalisation to information tasks and environments relating to digital libraries. In addition, the user models will need to be flexible and dynamic because the information elements listed above will change with respect to time and space and thus the models will need to evolve accordingly. Finally, user models must include features of the community aspect of human behaviour and preferences, in which memberships in social and work groups can influence a person's needs and requirements. This will certainly need to extend even current capabilities in this area [3].

A particular challenge for personalisation research is that long-term user models must encompass a timespan that is defined in terms of a human lifetime [22]. This need defines a type of research and experimental evaluation that has not been done before in computer science and related disciplines.

Finally, it should be stated that personalisation and recommender systems are not the only research areas that will benefit from advances in user modelling as any kind of information navigation, including retrieval, browsing, summarisation, automatic linking, and so on, in almost all domains, would show progress.

3.2 Making recommendations and doing personalisation

Current personalisation and recommendation techniques are based on relatively simple models. Pervasive personalisation and recommendations in digital libraries require research on a range of topics that current systems only begin to address, for example making distinctions between ephemeral and persistent characteristics and requirements to support both long-term and short-duration personalisation and recommendations. Incremental improvement in existing algorithms will not achieve these goals, and fundamental research is required on algorithms for personalisation and recommendation that go beyond current similarity-based accuracy to address issues such as confidence, privacy, resistance to shilling attacks,[2] [28] authority, reputation, trust, novelty, recency, and utility.

---

[2] The term *shilling attacks* refers to malicious users who provide false recommendations, often *many* false recommendations, in order to artificially influence recommender system behaviour. It is common in peer-to-peer networks.

Recent research on recommender systems focuses on server-based systems that make recommendations based on the activities and preferences of large groups of people. Server-based personalisation is natural for building models of user groups and communities and can be tightly integrated with the content and services a digital library offers. Server-based personalisation is also sometimes preferred in commercial environments because it can be used to bind customers to the service, and switching services may mean losing one's personalisations or user models. Client-based personalisation [20, 42], on the other hand, is natural for building a detailed model of an individual over a variety of tasks and transactions and over a lifetime of use, and it gives people greater control over how and what personal information is revealed. The training data available at the client differ significantly from those available at the server, for example requiring the client to understand much more of the semantics of user interaction with various information services and resources. Server-based and client-based personalisation use different techniques, rely on different amounts and types of data, and may be studied by different research communities. One of the important challenges in this research area is bridging the gap between these two extremes, to develop portable server-based user models and hybrid models.

The balance between user-specific and community-based personalisation [47] for an individual and a particular resource or task will vary. The first time a person encounters a digital library (*cold start*), personalisation can be accomplished by relying on commonalities between the individual's library-independent model and that of similar individuals that have interacted with the digital library in the past. Over time, as the individual gains more experience with the digital library, the balance between user-specific and community-based personalisation will shift and the contribution that the individual makes to the community and the community model will increase.

Sophisticated personalisation requires more-sophisticated control strategies. Time-sensitive personalisation requires an understanding and adaptation to the timespan of an individual's information needs (short-term, long-term) and appropriate convergence mechanisms. Convergence must be balanced by an ability to adapt to an individual's changing preferences, knowledge, and abilities over time.

Traditionally many of these topics have been studied using online experiments in operational digital libraries because they require interaction with large user communities and detailed information about user preferences or histories. This research methodology is effective, but it is also expensive and a barrier to entry for new research groups. There is a strong need for greatly improved simulation and modelling capabilities, to reduce research costs and enlarge the community of people who can study these topics. This is a major research topic in itself, and significant progress must be made in understanding users and their information-seeking behaviours before it can be addressed in this context.

3.3 User interaction

Unlike the traditional interpretation of user/system interaction, where each party has a fixed role to play, in personalisation the system should promote a more flexible mixed initiative approach that would allow for the integration of human and automated reasoning for more in-depth interaction. To date most recommender systems have been designed to implement a very simple model of human-machine interaction. Mixed-initiative systems adopt more-flexible approaches to recommendation and feedback. In addition to communication, interaction is concerned with the presentation and representation of information in all of its forms. The research challenge is how to design and facilitate personalised user interaction.

Interaction should accommodate the variability in the way users undertake different tasks within a myriad of work situations. Personalisation should allow systems to adapt to users whilst enabling an appropriate degree of user control. Hence a second research challenge in applying personalisation to user interaction is achieving this balance. Neither challenge can be met without drawing on robust user models.

3.4 Evaluation

Personalisation raises new evaluation issues, and standard traditional approaches are inadequate. First, personalisation needs to be assessed from the perspective of the individual, the individual within a group as well as the group or community as a whole, and this is necessary in order to effectively assess the many different kinds of personalisation we expect to evolve. Second, because of space and time dimensions, longitudinal studies must be conducted, and some of these studies will need to span very long timespans, which will require long-term funding commitments. This is necessary as we believe in the importance of very long-term user models where personalisation will be used. Third, user-centred quantitative and qualitative evaluations will need to be undertaken in both live and laboratory settings depending on the research objectives. Finally, the design of evaluative studies will need to identify appropriate criteria and metrics for defining success that extend beyond and complement current measures of performance as found in [49].

A central research challenge in the evaluation of personalisation, particularly in short- to medium-term horizons, is to build a suitable platform for evaluating personalised information seeking. This would contain rich datasets for training and comparative testing, standard tasks and scenarios, open-source software for applying standard algorithms, and services for conducting both laboratory and live evaluations. This will arise as we expect the current performance measures to be inadequate

for assessing the kinds of personalisation and recommending we envisage in this paper.

3.5 Social effects

As digital libraries become more commonplace and numerous, and as digital library support for social communication and sharing of information becomes routine, digital libraries both affect and are affected by social interaction, and thus the dynamics of social settings must be taken into account.

Personalisation has a role here in that each person's experience of the digital library will vary from that of others. The library will produce different experiences, and possibly different answers, for each individual, thus reducing the common experiences and shared references that bind the community together, and increasing problems associated with transparency, divergent interpretations, and training. Even without direct social interaction, a social effect can thus arise from personalisation. More-direct communication, such as recommendation and annotation, also has social effects. Sharing information with others creates possibilities for discovery, reinterpretation, and discourse. An individual may contribute to a digital library and its community not only traditionally, as an author, but as a source of recommendations and annotations. Recalling the role of the pathfinders in Vannevar Bush's *As We May Think* [13], a digital library user may become increasingly significant to others as his or her personalised interaction with the digital library is made persistent or public. Other people may not wish to take on such a role, preferring to have less information about them made available to others.

A central and complex challenge for digital library research is the balance between privacy and collaboration, an issue that is familiar in other fields such as security. The past treatment of privacy in both traditional academic, public, and research libraries and in many experimental and commercial digital library systems has been overly simplistic. Most have chosen to protect privacy at all costs, not even offering users choices that might allow the benefits of sharing information with other users. A few have promoted sharing using privacy methods such as pseudonyms that are easily compromised.

Privacy protection does not mean imposing crude barriers that prevent an individual from interacting with people he or she might derive some benefit from. The public has demonstrated repeatedly, in settings as diverse as online commerce services and supermarket value cards, that it is often willing to give up a degree of privacy in exchange for a specific benefit. The challenge for digital library developers and researchers is to protect people's most essential privacy while also ensuring that desirable social effects are supported. Privacy solutions must allow people to shape and control how they present themselves to others, which requires that the solutions be comprehensible and based on informed consent. It also means

helping people understand that any sharing of information can bring benefits and losses.

Privacy protection is a societal issue that spans many fields. One example is research in computer-supported collaborative work on design for privacy in ubiquitous computing environments that offers basic guidelines for control over what information is released about an individual as well as feedback about who has accessed what information [17, 41, 56]. Feedback affords better understanding of how to adjust the controls over one's presentation to others and how to adjust one's behaviour given the available controls. Privacy research in digital libraries will necessarily be influenced by privacy research in other fields. In summary, the challenge here is to look beyond purely technical approaches to privacy and collaboration, and beyond purely social ones as well. Instead, we should start from the assumption that users will adapt to and control digital library systems just as much as digital libraries adapt to and influence their users, and we should look for ways to support this larger system of control, feedback, and adaptation.

4 Recommendations and priorities for future research

Our recommendations and priorities for future research follow the research challenges identified earlier and are divided into five major areas.

4.1 User modelling

In order to enable a greater range of personalisations that range over more heterogeneous data, over the short and long term, and that cover input from multiple digital libraries including personal digital libraries, more needs to be known about users, user communities, and their tasks. Greater emphasis needs to be placed on investigating methods for building more-robust, flexible, and portable models of the complexity of users, tasks, and contexts to inform the diverse possibilities for personalisation in digital libraries. This implies a need to support interdisciplinary collaborative research from the different research communities, including HCI, CSCW, IR, and others. Targets for this work include being able to develop implicit rather than explicit methods for learning user preferences that form the user models and developing user models that are portable across applications, devices, and digital libraries. Perhaps the biggest challenge in this area will involve the development of user models that will drive personalisation and recommender systems, that are rich enough to capture as much of the user's task environment (context, task, situation), history, contribution to communities, and individual preferences as possible while conforming to a person's privacy choices. Such models will need to exploit and use the rich data made available from, for example, personal digital libraries.

4.2 Personalisation

The development of more-sophisticated and complete models of user needs and behaviours opens up an opportunity for the development of more-elaborate and sophisticated techniques for personalisation and recommender systems that more accurately capture the demands of the real world and that range over external as well as personal digital libraries. Aspects of personalisation and recommender systems that need to be researched, developed, and tested include the differences between ephemeral and persistent needs, long-term vs. short-duration requirements, hybrid client-server personalisation architectures, and supporting a balance between community-based and user-specific personalisation. Basic research is required to go beyond the current similarity-based measurement as the building block for personalisation and recommender systems. Finally, the personalisation process itself should be open and transparent to users and form part of their model of what the system is supposed to do.

4.3 User interaction

In order to provide effective and diverse forms of personalisation, the focus must be on the design of the interaction per se as an integral part of the whole system. In much of the research in digital-library related areas, user interaction has been seen as an afterthought, or something that is bolted onto a system at the end. There is a need to develop multimodal mixed-initiative interfaces that draw on a range of user information-seeking models, those same models that we earlier indicated would need to be enriched. The requirement is thus for research to develop theories of interaction that underpin the design of applications and vice versa and that go beyond issues of elicitation, presentation, and feedback.

4.4 Evaluation

User-centred evaluation must become an inherent part of system design and the evaluation of new technologies. New methods and evaluation criteria are required to assess personalisation systems in a cost-effective way. The imperative is to develop evaluation methodologies and make standard resources and tools more readily available for system developers. Evaluation based solely on quantitative measurements of system performance will remain important but should not be as dominant as it is now.

A serious challenge to research is the need for a large existing infrastructure of software, content, and committed users with which to perform evaluations. Research progress would be improved considerably by a set of large-scale operational digital libraries in which any qualified researcher could conduct experiments in the real world. Such shared research infrastructure might be viewed as the equivalent of the particle accelerators used for physics research. In practice, a shared research infrastructure might be digital libraries created for the dual purposes of research while also serving some community, or it might be some form of access to existing commercial or non-profit digital libraries. A set of shared digital libraries would dramatically lower the barrier to entry in this research area, and their costs would be amortised across a larger research community.

Research on personal digital libraries, which is nascent, must accelerate quickly. In less than a decade many projections suggest the average home computer user will have sufficient disk space to store full motion video of every moment of a person's life, from cradle to grave. Many people no longer delete e-mail; soon they won't delete anything else, either. There is a long list of interesting research to be done on personal digital libraries, but perhaps the biggest challenge will be evaluation. Studies of how a person's use of a personal digital library evolves over time will need to be very long term and very multidisciplinary.

4.5 Social effects

Social interaction is a feature of large-scale digital libraries that distinguishes them from most other computing environments. Explicit and implicit recommendations and sharing of information, preferences, and experiences exposes a range of social issues that are rarely faced in computer science. The most serious among these is privacy, a problem whose solutions are as much a matter of social policy as technology. A particular challenge is to develop stronger and more varied forms of privacy protection while supporting the collaboration and sharing of information that has come to characterise many popular digital libraries.

Computers were once viewed as isolating people from people. Now many digital libraries play an important social role in forming and strengthening communities of people. Recommendation systems, which are based on sharing of information, and personalisation, which recognises an individual's specific needs, clearly play an important role in community development. However, little research has been done on the social dynamics of such communities, the roles people play within them, how their members interact, and how they evolve over time.

Digital libraries are a forum in which to study a wide range of technology, social, and policy issues at the intersection of computer science and the social sciences. Progress on the issues described in this report requires collaboration among researchers from a variety of disciplines.

References

1. Adaptive hypertext and hypermedia (2004) http://wwwis.win.tue.nl/ah/. Last visited 16 June 2004
2. Adar E, Kargar D, Stein LA (1999) Haystack: per-user information environments. In: Proceedings of the 8th international conference on information and knowledge management (CIKM), Kansas City, MO, pp 413–422
3. Almeida RB, Almeida VAF (2004) Reputation networks: a community-aware search engine. In: Proceedings of the 13th conference on the World Wide Web, New York, pp 413–421
4. Arms WY (2001) Uniform resource names: handles, PURLs and digital object identifiers. Commun ACM 44(5):68
5. Babaian T, Grosz BJ, Shieber SM (2002) A writer's collaborative assistant. In: Proceedings of the 7th international conference on intelligent user interfaces
6. Baeza-Yates R, Ribeiro-Neto B (1999) Modern information retrieval. Addison-Wesley, Reading, MA
7. Belkin NJ, Cool C, Stin A, Thiel U (1995) Cases, scripts and information seeking strategies: On the design of interactive information retrieval systems. Expert Syst Appl 9(3):379–395
8. Brown PJ, Jones GFJ (2001) Context-aware retrieval: exploring a new environment for information retrieval and information filtering. Personal Ubiquitous Comput 5(4):253–263
9. Brown PJ, Jones GFJ (2002) Information access and retrieval: exploiting contextual change in context-aware retrieval. In: Proceedings of the ACM symposium on applied computing, Madrid, pp 650–656
10. Brusilovsky P (199) Methods and techniques of adaptive hypermedia. User Modell User-Adapt Interact 6:87–129
11. Brusilovsky P, Kobsa A, Vassileva J (2002) Adaptive hypertext and hypermedia. Kluwer, Dordrecht
12. Brusilovsky P, Maybury M (eds) (2002) From adaptive hypermedia to the adaptive web. Commun ACM 45(5):30–33
13. Bush V (1945) As we may think. Atlantic Monthly, July 1945, online version. http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm. Last visited 4 Feb 2004
14. Callan J, Smeaton AF (eds) (2001) Proceedings of the 2nd DELOS network of excellence workshop on personalisation and recommender systems in digital libraries. ERCIM Workshop Proceedings – No. 01/W03. Dublin City University, Ireland, June 2001. http://www.ercim.org/publication/ws-proceedings/DelNoe02/
15. Camp LJ (2002) Emerging applications: DRM: doesn't really mean digital copyright management. In: Proceedings of the 9th ACM conference on computer and communications security, pp 78–87
16. Christel M, Kanade T, Mauldin M, Reddy R, Sirbu M, Stevens S, Wactlar H (1995) Informedia digital video. Commun ACM 38(4):57–58
17. Clement A (1993) Privacy considerations in CSCW: report on the CSCW'92 workshop. ACM SIGCHI Bull 25(4):34–39
18. Dix AJ, Finlay JE, Abowd GD, Beale R (1998) Human-computer interaction, 2nd edn. Prentice-Hall, Upper Saddle River, NJ. (3rd edition available at http://www.hcibook.com/e3/)
19. DLI (1996) The DLI testbeds: today and tomorrow. D-Lib Mag. http://www.dlib.org/dlib/july96/07contents.html
20. Fenstermacher KD, Ginsburg M (2002) Mining client-side activity for personalization. In: Proceedings of the 4rth IEEE international workshop on advanced issues of e-commerce and Web-based information systems (WECWIS'02), Newport Beach, CA, p 205
21. Field A, Hartel P, Mooij W (2001) Personal DJ: an architecture for personalised content delivery. In: Proceedings of the 10th international conference on the World Wide Web, Hong Kong, pp 1–7
22. Gemmel J, Lueder R, Bell G (2003) The MyLifeBits lifetime store (demo description). In: ACM SIGMM 2003 workshop on experiential telepresence, Berkeley, CA
23. Han H, Giles CL, Manavoglu E, Zha H, Zhang Z, Fox EA (2003) Automatic document metadata extraction using support vector machines. In: Proceedings of the 3rd ACM/IEEE-CS joint conference on digital libraries, Houston, TX
24. Ho Ha S (2002) Helping online customers decide through web personalization. IEEE Intell Syst 17(6):34–43
25. Järvelin K, Wilson TD (2003) On conceptual models for information seeking and retrieval research. Inf Res 9(1). http://InformationR.net/ir/9-1/paper163.html
26. Konstan JA, Miller BN, Maltz D, Herlocker JL, Gordon LR, Riedl J (1997) GroupLens: applying collaborative filtering to Usenet news. Commun ACM 40(3):77–87
27. Kumar R, Raghavan P, Rajagopalan S, Tomkins A (1998) Recommendation systems: a probabilistic analysis. In: Proceedings of the 39th annual symposium on foundations of computer science, Palo Alto, CA, pp 664–673
28. Lam SK, Riedl J (2004) Shilling recommender systems for fun and profit. In: Proceedings of the 13th conference on the World Wide Web, New York, pp 393–402
29. LANL (2004) MyLibrary @LANL. http://lib-www.lanl.gov/lww/mylibweb.htm. Last visited 4 Feb 2004
30. Liu H, Maes P (2004) User modelling I: what would they think? A computational model of attitudes. In: Proceedings of the 9th international conference on intelligent user interfaces
31. Manber U, Patel A, Robison J (2000) Experience with personalization of Yahoo! Commun ACM 43(8):35–39
32. Marshall CC, Golovchinsky G, Price M (2001) Digital libraries and mobility. Commun ACM 44(5):55–56
33. McArthur D, Giersch S, Graves B, Ward C, Dillaman R, Herman R, Lugo G, Reeves J, Vetter R, Knox D, Owen S (2001) Towards a sharable digital library of reusable teaching resources. Commun ACM 44(5):79
34. McGrath R, Futrelle J, Plante R, Guillaume D (1999) Digital library technology for locating and accessing scientific data. In: Proceedings of the 4th ACM conference on digital libraries, Berkeley, CA, pp 188–194
35. Moen WE (2001) Mapping the interoperability landscape for networked information retrieval. In: Proceedings of the 1st ACM/IEEE-CS joint conference on digital libraries, Roanoke, VA, pp 50–51
36. Moriarty C, Kushmerick N, Smyth B (2001) Personalised intelligent tutoring for digital libraries. In [14]
37. MovieLens (2004) The MovieLens home page. http://movielens.umn.edu/. Last visited 4 Feb 2004
38. OCLC (2004) The Dublin core: a simple content description for electronic resources. http://dublincore.org/Model. Last visited 15 June 2004
39. Paepcke A, Chang C-C, Winograd T, Garcia-Molina H (2001) Interoperability for digital libraries worldwide. Commun ACM 41(3):33–42
40. Ramakrishnan N (2000) PIPE: Web personalization by partial evaluation. IEEE Internet Comput 4(6):21–31
41. Ramakrishnan N, Keller BJ, Mirza BJ, Grama AY, Karypis G (2001) Privacy risks in recommender systems. IEEE Internet Comput 5(6):54–62
42. Rossi G, Schwabe D, Danculovic J, Miaton L (2001) Patterns for personalized web applications. In: Rüping A, Eckstein J, Schwanninger C (eds) Proceedings of EuroPlop-01
43. Rous B (2001) The ACM digital library. Commun ACM 44(5):90–91

44. Ruthven I, Lalmas M, van Rijsbergen CJ (2002) Combining and selecting characteristics of information use. J Am Soc Inf Sci Technol 53(5):378–396
45. Samuelson P (2003) Digital rights management and fair use by design: DRM { and, or, vs.} the law. Commun ACM 46(4):41–45
46. Samuelson P, Glushko RJ (1991) Intellectual property rights for digital library and hypertext publishing systems: an analysis of Xanadu. In: Proceedings of the 3rd annual ACM conference on hypertext
47. Sarwar B, Karypis G, Konstan J, Reidl J (2000) Analysis of recommendation algorithms for eCommerce. In: Proceedings of ACM'00 conference on electronic commerce, pp 158–167
48. Savolainen R, Kari J (2003) Toward a contextual model of information seeking on the web. New Rev Inf Behav Res 4:155–175
49. Smeaton AF, Callan J (2001) Workshop reports: joint DELOS-NSF workshop on personalisation and recommender systems in digital libraries. ACM SIGIR Forum http://www.acm.org/sigir/forum/S2001-TOC.html
50. Smeaton AF, Murphy N, O'Connor NE, Marlow S, Lee H, McDonald K, Browne P, Ye J (2001) The Físchlár digital video system: a digital library of broadcast TV programmes. In: Proceedings of the 1st ACM/IEEE-CS joint conference on digital libraries, Roanoke, VA
51. Smith TR, Frew J (1995) Alexandria Digital Library. Commun ACM 38(4):61–62
52. Stanford University (2004) Stanford annotated interoperability bibliography. http://www-diglib.stanford.edu/diglib/pub/interopbib.html. Last visited 14 June 2004
53. Stanford University (2004) Digital library resources. http://www-diglib-stanford.edu/diglib/pub/Resources.shtml. Last visited 17 June 2004
54. Stevens S, Christel M, Wactlar H (1994) Informedia: improving access to digital video. Interactions 1(4):67–71
55. Torres R, McNee S, Abel M, Kinstan J, Riedl J (2004) Collaboration and group work: enhancing digital libraries with TechLens+. In: Proceedings of the 2004 joint ACM/IEEE-CS conference on digital libraries
56. University of California, Berkeley (2004) Information privacy in ubiquitous computing – background readings. http://guir.berkeley.edu/groups/privacy/readings.html. Last updated 1 April 2004
57. Various (2001) Special issue on the theme digital libraries. Commun ACM 44(5)
58. Vertegaal R (ed) (1997) Readings in CSCW. Centre of Telematics and Information Technology (CTIT), Twente University, Enschede, The Netherlands
59. W3C (2004) Metadata activity statement. http://www.w3.org/Metadata/Activity.html. Last visited 15 June 2004
60. Weiser M (1993) Ubiquitous computing. IEEE Comput 26(10):71
61. Wilson TD (1999) Models in information behaviour research. J Document 55(3):249–270
62. Yahoo, Inc (2004) The My Yahoo ! home page. http://my.yahoo.com/?myHome. Last visited 4 Feb 2004