LANCASTER
UNIVERSITY

Computing
Department

Recommendation and Usage in the Digital Library

David M Nichols, Michael B Twidale and Chris D Paice

ABSTRACT

The movement from the physical to the digital library offers both dangers and opportunities. Alongside the greater quantity of online material goes the problem of quality assurance: how can be the information searcher be sure of the status of a document. We suggest that this be addressed by supporting *recommendations* and that the key feature that links these recommendations together is that of *usage*. The main use of usage data within information science is currently that of a research tool in the form of transaction log analysis. In a digital library this data, together with other evaluations and recommendations, can enrich the existing information structure. Several approaches t o the integration of usage data are described together with their respective costs and benefits. The social implications of these possibilities are discussed with particular reference to the privacy of usage data.

Recommendation and Usage in the Digital Library

David M. Nichols, Michael B. Twidale and Chris D. Paice

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

The movement from the physical to the digital library offers both dangers and opportunities. Alongside the greater quantity of online material goes the problem of quality assurance: how can be the information searcher be sure of the status of a document. We suggest that this be addressed by supporting *recommendations* and that the key feature that links these recommendations together is that of *usage*. The main use of usage data within information science is currently that of a research tool in the form of transaction log analysis. In a digital library this data, together with other evaluations and recommendations, can enrich the existing information structure. Several approaches to the integration of usage data are described together with their respective costs and benefits. The social implications of these possibilities are discussed with particular reference to the privacy of usage data.

Introduction

The movement from the physical to the digital library offers both dangers and opportunities. Alongside the greater quantity o f online material go the twin problems of coping with this far greater amount o f material and of dealing with quality assurance: how can be the information searcher be sure of the status of a document (or multimedia resource)? Users are in danger of being overwhelmed both by the quantity of information and by its variety of form and quality. There are many existing solutions to this problem i n the physical world that can inspire our designs for addressing the problem in the digital world. One broad method is that of *recommendation*.

In the academic environment there are a wide range of features that can be considered to convey recommending information. These include personal recommendations and book reviews, while in the case o f students, reading lists convey recommendations a s to what the lecturer deems to b e significant. Implicit recommendations can be inferred from such features as the provenance o f an item (say being in a peer-reviewed journal) o r even its presence in a specialist library. A library embodies the history of the purchase recommendations o f the librarians and others in its current stock. In the economy a t large a recommendation industry (*Consumer Reports*, *What Car?*, financial advisors, insurance brokers, etc.) has arisen to guide consumers through the myriad options promoted by retailers and advertisers [Aldridge, 1994].

We suggest that the key feature that links these recommendations together i s that of *usage*. The nature of usage can vary greatly depending on the objects under consideration: for a document this can vary from glimpsing the title to regular consultation o f the text. Recommendations often come with an implicit notion of usage: a book review is predicated on the reviewer having read the book. Usage data can embody a recommendation with more credibility, hence the emphasis placed on testing by consumer magazines.

In physical libraries, examples of usage are the well thumbed volume, the book that falls open at an often referred to page, and the often checked out book with a large number o f recent return stamps. Recording more detailed usage is not easy except at the point o f borrowing a volume: by contrast usage in digital libraries could be recorded in great detail. The main use of usage data within information science is currently that of a research tool in the form of transaction log analysis, or online monitoring (e.g. [Borgman, Hirsh and Hiller, 1996]).

In this paper we intend to explore the nature of recommendations and the value of usage data. We examine the value o f usage data in combination with recommendations and examine an approach

for generating recommendations from stand-alone usage data. We consider the problems with attempting to incorporate recommending into digital libraries, including the social implications o f these possibilities with particular reference to privacy and the ownership of usage data.

Recommendations

Recommendations are important because they are   a general mechanism for coping with information. They enable information searchers to cope with greater quantities of data and locate desired material more efficiently. Recommendations embody the experience and judgement o f a community and are the means by which members can benefit from the prior experiences o f others [Grosser, 1991].

A recommendation is a communicated judgement on the fitness of an object for a given purpose. It says that this object, or set of objects, should b e considered or b e prioritised above others that may be considered. That is, recommendations address both recall and precision. The information user, the recommendee, may decide to act on the recommendation or not.

Studies of the information search   process have identified several different   sources of recommendations including subject librarians, colleagues, other documents, bibliographies etc. (e.g. [Bates, 1979a; Bates, 1979b; Kuhlthau, 1991; Menzel, 1959]). Academic researchers make substantial use of recommendations from their peers [Menzel, 1959; Stoan, 1984]. Attendance at conferences enables the informal meetings that lead to recommending events. By contrast beginning graduate students lack this rich set of social links that yield personal recommendations and so could benefit most from an automated system [Barry, 1996]. When working on a new field of study the contacts may well break down for all researchers.

Recommendations can clearly take many physical and electronic forms, but there are common elements that can be identified that highlight the different social   roles that recommendations c an play. Table 1 shows several different dimensions along which recommendations can be compared.

Dimensions	Details
Object	
Type	who, or what, is being recommended?
Recommender	
Identity	known, anonymous or aggregated?
Status	is the person an expert on this type of object?
Experience	evidence of knowledge of object (e.g. testing, usage)
Motivation (Bias)	does the recommender benefit, and if so, how?
Recommendation	
Polarity	positive or negative
Medium	printed reviews, word-of-mouth, email, Web page
Aggregation	several judgements may be combined
Truthfulness	is the recommender mistaken or dishonest?
Strength	is the recommendation tentative/ confident/ emphatic?
Input-cost	what are the costs of receiving the recommendation?
Output-cost	what are the costs of making the recommendation?
Recommendee	
Identity(ies)	who receives the recommendation?
Purpose / context	why might the recommendee be interested?

Table 1 Attributes of recommendations

As the variety of recommendation services in the economy indicates, there are recommendations for almost every type o f good and service. Often these are from known individuals: family and friends, professional adviser or public recommendations (such as restaurant reviews). Informal recommendations (or word-of-mouth) from family, friends and colleagues have been identified as a major influence on consumers' purchasing decisions [Walker, 1995]. Users of physical libraries collaborate in their searches despite the lack o f any technological recognition in the system design [Twidale and Nichols, 1996; Twidale, Nichols and Paice, to appear]. Consumers in retail settings interact with strangers in order to obtain evaluations on potential purchases [McGrath and Otnes, 1995].

For many types of potential recommendation we can see analogues of their use in existing non-technological interactions, but their effect can be made much broader by the use of a technological medium. For example, a recommendation to a group of colleagues working on the same corridor is similar to posting the recommendation to an appropriate Usenet newsgroup, but its effect is magnified and involves passing the information on to people the recommender does not know.

An information searcher may be exposed to several recommendations of different strengths (and polarities) for a particular object, in which case she will have to combine them into a single assessment. This aggregation can also be performed by a recommender (such as a consumer testing organisation), who synthesises several other information sources into a single recommendation (e.g. [Hill et al, 1995; Shardanand and Maes, 1995]). When recommendations are aggregated there is a likely loss of identification; the recommender may be referred to a s part o f a group ("ASIS recommends this book") or may be rendered anonymous.

As [Hill and Terveen, 1996] puts it, in describing the recent interest in social filtering:

> a basic thesis of this work … is that personal relationships are not necessary to social filtering. In fact, social filtering and personal relationships can be teased apart and put back together in interesting new ways. For instance, the communication of quality judgements can occur through less personal, or even impersonal relationships, with Usenet news being example.

Existing sources of recommendations are often impersonal but when they can be made personalised and adaptive new applications become possible. Print-based recommendations abound but, as many are free text (e.g. book reviews), they are difficult (computationally) to combine and personalise.

A recommender may have reasons for deceiving the recommendee and even a truthful source may only be passing on the unreliable evaluation of a third party. The credibility of a recommender is greatly enhanced if they are relating personal experience o f a product . As such it seems worthwhile to examine the ways in which information searchers become aware o f objects, before examining issues of credibility.

'Discovery Model'

A person normally goes through a series of stages i n finding out about an object. These experiences may vary from simply being aware o f an object's existence to using it every day. In general the interaction of a person (P) with an object (X) may be approximated b y the descriptions shown in Table 2. At each stage greater detail is obtained; both from the intrinsic qualities of the object and external recommendations.

Activity		Response
glimpse	P is aware of the existence of X	
		focus or **ignore**
consider	P looks at summary information about X	
		select or **reject**
examine	P looks at detailed information about X	
		adopt or **reject**
use	varies with the nature of X	
assess	P evaluates the experience of using X	
		endorse or **not**

Table 2 Simplified discovery model of a person (P) and an information object (X)

We do not wish to suggest that every discovery process goes through a fixed number of stages; some users may omit stages or only use a subset of the activities. Also the *use* of an object may be a complex activity compared with earlier stages. The main point is that users display increasing *commitment* to the object the further they proceed through the process. In particular, the adoption of an object displays a large increase in the commitment to the object relative to the preceding stages.

The discovery process may be illustrated by online searching in a bibliographic database. First, the user glimpses a one or two line reference to a document among a list of retrievals: maybe just the title and authors. If it seems of possible interest, the user asks to see more information, probably including an abstract. At a further stage, the user may examine the completed document, and if it indeed proves relevant, the document will be 'adopted'. Thus, if the user has a 'real world' goal she may adopt some technique or object described in the document; if she is an academic she may cite the document in a paper. Obviously, during such a search, the user is making discoveries about a whole set of items; many of which are rejected before the adoption stage.

Note that, at any stage during the discovery process, the user's opinion about an object can be communicated to someone else in the form of a recommendation. At the time a user adopts an object we suppose that it indicates that they believe that it is in some way relevant to their information need. This may be further indicated by repeated use or a positive endorsement of the object.

In addition, the level of commitment is an important qualifier to a recommendation. A recommendation that is accompanied by evidence of significant commitment (e.g. repeated usage) is likely to be regarded more highly than others. Although a recommender may be highly committed to an object in the sense of being an experienced user who rates it highly, in comparison to other related objects, she may have other motivations besides the interests of the recommendee.

Credibility of Recommendations

An important and problematic type of recommendations consists of those where the recommender benefits in some way if the recommendee takes up the recommendation. The recommendation of a shop assistant may be treated with caution where it is known or suspected that they have a personal interest in encouraging the purchase of the more expensive option. At least in such a case the direction of bias is known, and this can be factored into the purchaser's assessment of the validity of the advice. A more confusing case arises when the direction of the bias is unknown. A recent example in the UK has been that of financial advice; advisers were paid

commission at different rates on different institutions' financial products, and so had an incentive to recommend some products over others [Waine, 1995]. The purchaser had no way of knowing the direction and extent of any bias in the advice.

In the digital library a blatant example of the problem of bias would be an author recommending their own book. They may have the incentive both of royalties and of the more intangible enhancement of reputation by ensuring that their work is read and referred to more often. Recommenders can be motivated by many social pressures; to demonstrate superior knowledge, to engage in social interaction and personal satisfaction [Dichter, 1966]. Indeed some people engage in recommending activity without any prompting - [McGrath and Otnes, 1995] describe them as 'proactive helpers'; a less intense version of the 'market maven' [Feick and Price, 1987]. The broader social context, especially whether the actors involved expect mutual reciprocity, can affect the flow of word-of-mouth information in complex ways [Frenzen and Nakamoto, 1993] - that are beyond the scope of this paper

In marketing, one method of describing the social influences on consumers is to consider the reference groups to which they belong. Reference group influences are of three types: informational, comparative (aspiring to be like group members) and normative (conforming to group norms) [Kelman, 1961]. In a study to examine reference groups [Park and Lessig, 1977] the key elements reflecting informational influence for a product can be characterised as: people who work with the product, family and friends, professional association, independent testing agency and observation of experts' behaviour. In terms of Table 1, the attributes of the recommender were key elements in the credibility of the recommendation.

Provided that there is no reason to expect bias, the *status* of the recommender is an important factor. Is the recommendation made by an expert or someone of sound judgement? This is not the same as the experience of the recommender: a person may have considerable experience of some object but if they are emotional or impulsive their judgements may have little value.

Costs and Recommendations

Inevitably there are costs associated with recommendation activities: both the input-cost of making the recommendation and the output-cost of accessing the recommendation. As [Grudin, 1994] has noted, a common reason for cooperative computer systems failing is that the costs have been too high relative to the expected benefits to users, and also where those bearing the costs are different from those enjoying the benefits. In particular, a recommendation may not be made if the cost experienced by the recommender is too high (or cannot be recouped). In general, the technical output costs are relatively low: recommending information can be provided as a supplement to a search either as an adjunct to a hit (such as a simple star rating) or by feeding into a broader relevance rating algorithm. However, if the recommendations are viewed as valuable by potential recommendees then a recommender may try to charge a fee for access to them.

To address the input costs, and particularly the effort disparity between the recommender and the beneficiaries, it is important to consider cases (e.g. usage data) where the cost of providing the information is near zero [Hill and Terveen, 1996].

The particular quality of usage data that distinguishes it from evaluative data (such as *Seals Of APproval* [Röscheisen, Morgensen and Winograd, 1995], ratings [Allen, 1990; Resnick et al, 1994; Shardanand and Maes, 1995] and new associations [Bush, 1945; Kantor, 1993; Maltz and Erlich, 1995]) is that the perceived input-cost to the recommender is zero. For example, users of the World Wide Web are often surprised at the level of detail a Web server can record about their activities without any awareness on their part that their usage was even being logged.

Although there are costs associated with users' activities (entering queries, Web surfing) they may be personal benefits. A good example of re-using this personal work is Web page recommendation using a work-group's set of Web browser bookmark files [Wittenburg et al, 1995]. The files are personally structured but the benefits are shared amongst the group.

The costs and benefits of a n information system (or sub-system) can be viewed from several perspectives. Alongside the perspectives of the searcher and the recommender is that of the system owner or information provider. The system provider would need to consider the costs (storage, processing and maintenance) and benefits (improved searching efficiency) of recommendation functionality summed over all the users of the system. As current trends are towards increased computing power and cheaper storage, we may expect systems costs to become less important with the passage of time.

We now consider usage data and how it may be captured and reused in digital libraries.

Usage

Usage is one of the means by which information searchers leave their marks on the systems that they interact with. If recommendations are a mechanism to share experience amongst information searchers then usage data is the raw material that the recommendations are based upon. The capture of usage data is intimately connected to the form of the objects that are used. In a physical library there are severe limits to the information that can feasibly be captured. For example, the unrecorded consultation of books that are never borrowed from physical libraries has produced several research studies to determine its effect (e.g. [Ross, 1983]). Physical libraries are an example o f physical systems in general: they are largely amnesiac. Some physical objects (e.g. car tyres, carpets, books) do retain some information about their use, often via frictional wear. This observation has led to the description o f History Enriched Digital Objects (HEDOs) [Hill and Hollan, 1994] - encoding information about their usage as part o f their structure. However, most physical interactions between objects do not produce (re)usable information:

It is a truism to say that it is difficult to reuse what you cannot record but digital libraries offer such a huge shift i n the potential for recording and reusing usage data that applications never considered before are now feasible. Indeed, the very technology that allows the creation o f digital libraries also allows us to consider storing the large volumes of usage data that will be created by an increasingly networked society.

Applying the HEDO concept to a document produces 'read wear' [Hill and Hollan, 1992] - where the reading of a document is recorded as part of the document. Viewing a digital library as a large HEDO leads to the question: can the usage of a digital library be encoded as part of its structure so as to aid information searchers? To attempt to answer this question we first consider the type of usage data that can be recorded in a digital library. Table 3 shows some of the kinds of usage data, with examples, that may be available to designers of digital libraries. Some of these types of data , such as purchasing, repeated use and explicit referencing, are possible to record in physical systems, although even then it may be exorbitantly expensive. Most, such as examining, considering or querying are not. They can however be captured b y transaction logging systems (e.g. [Borgman, Hirsh and Hiller, 1996; Flaherty, 1993]).

Type of Usage Data	Example
Purchase	buys book
Assess	evaluates or recommends
Repeated Use	multiple check out stamps
Refer	cites or otherwise refers to document
Mark	Add to a 'marked' or 'interesting' list
Examine	looks at whole document
Consider	looks at abstract
Glimpse	sees title in list
Associate	returns in search but never glimpses
Query	association of terms from queries

Table 3 Different forms of usage data that could be captured in a digital library

In terms o f the 'discovery model' described earlier, different forms of usage data can b e (roughly) associated with different levels of commitment to the documents under consideration. [Ingwersen, 1996] describes four cognitive structures of representation:

• Author-generated representations

• Automatically generated representations

• Indexer/domain expertise

• Searcher-generated representations

Each of these produces different forms of structure that can be used in Information Retrieval (IR). In a computerised IR system the searcher-generated representations are reflected in the usage data. [Koenig, 1990] mentions several approaches that exploit searcher-generated representations, such as the "or-group" thesaurus [Reisner, 1966]. Query-based usage data is recorded so that keywords that have been "or-ed" together by previous searchers can be suggested as synonyms to later searchers.

Transaction logging data has largely been used as a research tool [Rice, 1983], to investigate the behaviour of users, identify common problems, improve system interfaces etc. An alternative view is to see usage data as a form of metadata [Böhm and Rakow, 1994] - enriching the database [Hjerppe, 1989]. For example, [Sandore et al, 1993] suggests a social role for the reuse of this data:

> one logical approach is to refine the interface so that it can use transaction log data to modify search arguments to optimize the user's search strategy. An alternative approach might involve using transaction log data to improve retrieval power in the search engine by enriching the "authority" of the database.

We now describe two related scenarios based on the social use of usage data: recommending and matchmaking.

Scenario 1 - Recommendation on the Basis of Usage

	D1	D2	D3	D4	D5	D6	D7	D8
S1	•			•	•		•	•
S2		•	•			•		
S3			•		•	•		•
S4	•		•		•		•	
S5		•		•	•			
S6		•				•		
S7	•			•			•	•
S*		•	•					

Table 4 An abstract representation of database usage

One application of usage data is for an IR system to recommend items to users based on other users' searches. To illustrate how this may work we consider the abstract mini-database shown in Table 4. The database consists of 8 data items (D1 to D8). The database has been used in 7 distinct search sessions (S1 to S7) and is currently being used in a eighth (unfinished) session (S*). To simplify the exposition we consider that there is only one type of usage data: a data item has either been used or not used. In practical terms this notion of use could be regarded as equivalent to 'has been returned in a user-specified search'. Each use of a data item is shown as a bullet; so in the first search session performed on this database the searcher returned items D1, D4, D5, D7 and D8.

In the current search (S*) the searcher has retrieved data items D2 and D3. At this point we could imagine the searcher runs out of ideas and asks the system for help. One approach for the IR system would be to perform the algorithm shown below (specific intermediate results are shown for the database and usage history shown in Table 4):

(1) Compute the similarity between current search session S* and previous search sessions S1 to S8, and generate a rank-list.
(2) Select the highest-ranked previous search session (S^)
(3) Find any data items in S^ which are not in S* and recommend to user.
(4) If more recommendations are needed repeat from (2) for next highest ranked previous search session.

Thus, if the similarity is simply a count of the number of items shared between S* and the other search sessions, S2 is selected as S^ (with count=2). S2 contains D6, which is not in S*; therefore D6 is recommended.

The resulting recommendation is that of a data item (D6) that is connected to the items used in the current session via the previous activities of other users. It is an inferred recommendation rather than an explicit evaluation. This approach is based on a cluster hypothesis: that items found in a search session form a natural cluster. The source of the clustering is irrelevant; items may have found by a keyword search, a citation search, a subject search, a serendipitous find etc. There is evidence that multiple query representations improve retrieval results [Belkin et al, 1995]; we wish to emphasise that different users can be a valuable source of queries that are related to, but not identical, to others. The advantages of this approach are that searchers can retain anonymity

(sessions need not be connected to individuals), the perceived input-cost to the searcher is zero and the output-cost is low.

There are many possible variations of this general algorithm. Different levels of usage (such as those in Table 3) may be recorded and may be combined in different ways. Alternative definitions of a 'session' may be possible (such as a "search cycle" [Penniman, 1975]) and more recent sessions may be treated differently to older ones. In the same way usage within a session is not necessarily equal: as later usage will have been influenced by prior usage [Borgman, Hirsh and Hiller, 1996]. Alternatively, usage information could be used in combination with more conventional approaches such as keyword searching, thesauri and relevance feedback.

[Wittenburg, et al., 1995] describes a restricted example of this scenario, Group Asynchronous Browsing (GAB); in terms of Table 3 it is based on the 'mark' type of usage data. Web browser bookmark files are shared and structured to create a social browsing space for a work group. Only those Web sites which the user has adopted (by inclusion in a bookmark list) are considered, and GAB does not explicitly include a clustering approach; rather it merges users' bookmarks.

[Koenig, 1990] mentions several approaches that are similar in intention to usage-based recommendation, such as the "or-group" thesaurus [Reisner, 1966]: where keywords that have been "or-ed" together by previous searchers are suggested as synonyms. [Chalmers, Ingram and Pfranger, 1996] describes several techniques for including usage data on a 3-D visualisation of a database.

A problem that has been noted by several researchers is that these forms of social recommendation rely on there being a critical mass of data. The ANLI system [Kantor, 1993] has used the hypothesised relationship between items borrowed by the same person to "prime the pump" of its recommendations between books. [Maes, 1994] uses virtual users based on keywords or authors to create an initial base of recommendations. A system that relies on explicit recommendations (with an input-cost to the recommender) starts with no data and is therefore at risk of giving inaccurate recommendations. In addition, users who receive no benefits may conclude the system does not work and stop contributing recommendations. By contrast, a usage-based system (with no input-cost to the user) can collect data unobtrusively for a considerable time before starting to issue recommendations.

An example of usage capture that is invisible to users is the access log of a World Wide Web [Berners-Lee, Cailliau and Groff, 1992] server. User accesses from the log file(s) can be clustered and used to dynamically suggest a Web page on the basis of previous users activities. Furthermore, experimental results indicated that clusters in the user access patterns were not always apparent from the pre-existing structure (the links between the pages) and would not have been discovered without log analysis [Yan et al, 1996].

Scenario 2: Matchmaking on the Basis of Usage

The example above relies on user grouping but does not require the users to be identified. When users can be associated with their usage data additional functionality becomes available. Matchmaking involves the introduction of people who otherwise would not meet: it is a form of recommendation involving people rather than inanimate materials [Foner, 1995]. An example from the academic world would be the introducing (perhaps at a conference) of two acadamics with related interests as determined by a colleague of each.

Introduction agencies that perform matchmaking usually do so on the basis of the characteristics of their clients. These characteristics are usually self-provided [Zhou and Abdullah, 1995] and are analogous to the keyword approach of traditional IR systems. By contrast, consider again the data shown in Table 4. If we examine the seven previous sessions we discover that session 2 and session 7 are highly similar. We may therefore, tentatively, that the users responsible for these two sessions have a lot in common, and therefore (provided they both agree) can be introduced to one another.

The usage representation in Table 4 is extremely crude. A real application would generate very large amounts o f data although there may be scope for combining sessions, both by user a nd session similarity.

Social Implications

We believe it is important to examine in more depth than i s possible here the range o f privacy issues, the options for attempting to address them, people's perceptions of the problems and ways of involving users in the debate about what is acceptable to them. The privacy o f usage data about both physical and digital libraries i s of considerable concern to both librarians and computer scientists [Kurth, 1993; Rosenbaum, 1996; Rotenberg, 1992].

Even if a computer system offers useful functionality, and its interface is carefully designed so that it is usable, it still may not be acceptable in a certain context of use. In the case o f some of the potential forms of recommending described in this paper, a key feature o f acceptability (or organisational usability [Kling and Elliot, 1994]) will be how the system addresses the issues of privacy.

Systems that involve social informatics must address these acceptability issues that relate to most collaborative systems [Bellotti and Sellen, 1993]. We claim that making the use of metadata about usage behaviour of others is a form of implicit collaboration. We can take a general principle o f a trade-off in such systems:

sacrificing privacy permits increased functionality

In the case o f digital libraries w e have seen that if anonymous recording o f usage is permitted, then powerful recommending is possible. If named recording o f usage is permitted, even more powerful recommending is possible. Note that encryption will not solve this particular privacy problem. Encryption can prevent third part eavesdropping, but t o provide the functionalities described the user needs to make her usage information available to the system and to trust that it will not be misused.

The principle also applies in the case of non technological systems. Telling colleagues about what you are working on in the hopes of getting recommending feedback, or even just asking a librarian for help clearly involves a loss of privacy that may be regarded as undesirable and unacceptable i n certain contexts and circumstances.

For those advocating such functionality, a barrier to acceptability is users' (often quite justified) fear that they are signing a blank cheque. In the physical world, the degree of lack of privacy (such as being in an open plan office) i s often blatant. I n the digital world it can be much more subtle. Users do not know whether by assenting t o a seemingly innocuous loss of privacy they are, unbeknownst to them, conceding a far greater loss. If one does not trust developers t o encroach upon the agreed extent o f privacy loss, it makes sense to fear and to refuse to assent to *any* loss even a loss that in itself is acceptable and it is agreed would bring clear benefits to the user.

It is useful to look at current practice to inform our understanding of the issues. The world wide web, as noted above, already enables usage monitoring to be undertaken. The normal monitoring information states the originating machine. This may not serve to identify an individual, since machines are shared and people may use more than one machine. Nevertheless, we think it worthy of note that so many people seem to be unaware of the monitoring that already occurs. In the case of certain commercial websites, particularly electronic news media, the user has to pre-register. The registration process may require the entry of certain person information such as an email address or even a postal address as well as demographic details. Clearly this information can prove of great value to the website managers, allowing tracking not just within a single period of interaction with the site, but across multiple interactions. Usage may be correlated either directly with other users by similarity matches or with other external information about say demographics and purchasing habits. The work on collaborative filtering [Resnick, et al., 1994] is already being developed a s a commercial product to support personalised marketing.

Clearly these are issues that need greater attention. However we wish to recommend that an acceptable system should at least:

1) Make clear to the user what is being done (nature of privacy loss) and why. An elaboration of this might be that if all loss of privacy is perceived as just that, a loss, then it also needs to be made clear how the user is benefiting. In the case of certain commercial web sites maybe the sacrificing of privacy is regarded as acceptable for access to, for example, free high quality journalism.

2) Give the user a degree of control. In information searching, the user must be able to easily switch off and back on the activity recording. They should also be able to retroactively decide that some sequence that has been recorded must be deleted. Likewise users should be able to control the quality of the information recorded such as whether it is named or anonymised. [Hill and Hollan, 1993] point out an important extension to this: that the choices about privacy should themselves be private:

"The absence of history .... [should be] indistinguishable from inactivity"

That is, that one should not be able to find out that another person has chosen to switch on greater privacy.

There are other acceptability issues that do not relate to privacy. Here we consider one to do with loss of control. Libraries, their content and organisation are traditionally controlled by librarians. Mechanisms have been evolved to deal with objectionable (such as offensive or sensitive) information. With the functionality described above, the data available includes what people do with the underlying data. What if this metadata is in some way objectionable to future users? For example: a system with access to metadata could suggest keywords that were associated with those the user has employed so far [Koenig, 1990; Reisner, 1966]. What if a group of pranksters had associated say racial terms with offensive words? Similarly, the common prejudices of many users may build links between books by usage leading to recommendations of 'similar' books where the linkage is offensive to others.

Libraries have considerable experience of dealing with sensitive information. The main differences with the examples here are firstly that it is a new technology and secondly the causes of offence come from usage (the activity of the users) rather than central acquisition or classification decisions both of which are under the more immediate control of the library. The examples have more in common with providing access to email and the web for patrons [Rotenberg, 1992].

To address this problem we need as a minimum to explain to readers what the recommendations etc. actually mean; that they are vague approximations that may be similar in some way. They do not represent an official library view of how books should be categorised nor do they mean that the books present the same argument or even are about the same thing. We may also need to provide means for deleting usage data for these reasons.

Conclusions

Recommendations have always been important, but are expensive to collect. The need for recommendations increases with the growth in size and complexity promised by digital libraries. We have examined some of the issues involved in recommending and have focused on usage as a means of providing recommending information. While there are clear cost advantages including addressing the potential mismatch between the costs to the recommender and the benefits to the recommendee, there are implications for privacy that need careful consideration.

Acknowledgements

References

Aldridge, A. (1994), The construction of rational consumption in *Which?* magazine: the more blobs the better?, *Sociology*, 28(4), 899-912.

Allen, R.B. (1990), User models: theory, method, and practice, *International Journal of Man-Machine Studies*, 32(5), 511-43.

Barry, C. (1996), The Digital Library: the needs of our users, *Paper presented at the International Summer School on the Digital Library*, Tilburg University, The Netherlands, 5th August,

Bates, M.J. (1979a), Idea tactics, *Journal of the American Society for Information Science*, 30(5), 281-89.

Bates, M.J. (1979b), Information search tactics, *Journal of the American Society for Information Science*, 30(4), 205-14.

Belkin, N.J., Kantor, P., Fox, E.A. and Shaw, J.A. (1995), Combining the evidence of multiple query representations for information retrieval, *Information Processing & Management*, 31(3), 431-48.

Bellotti, V. and Sellen, A. (1993), Design for privacy in ubiquitous computing environments, in deMichelis, G., Simone, C. and Schmidt, K. (Ed.), *Proceedings of the Third European Conference on Computer-Supported Cooperative Work - ECSCW'93*, Dordrecht: Kluwer, 77-92.

Berners-Lee, T.J., Cailliau, R. and Groff, J.-F. (1992), The World-Wide Web, *Computer Networks and ISDN Systems*, 25(4-5), 454-9.

Böhm, K. and Rakow, T.C. (1994), Metadata for multimedia documents, *SIGMOD Record*, 23(4), 21-6.

Borgman, C.L., Hirsh, S.G. and Hiller, J. (1996), Rethinking online monitoring nethods for infomation retrieval system: from search product to search process, *Journal of the American Society for Information Science*, 47(7), 568-83.

Bush, V. (1945), As we may think, *The Atlantic Monthly*, July.

Chalmers, M., Ingram, R. and Pfranger, C. (1996), Adding imageability features to information displays, *Proceedings of UIST'96*, Seattle, WA, ACM Press.

Dichter, E. (1966), How word-of-mouth advertising works, *Harvard Business Review*, 44, 148-52.

Feick, L.F. and Price, L.L. (1987), The market maven: a diffuser of marketplace information, *Journal of Marketing*, 51(1), 83-97.

Flaherty, P. (1993), Transaction logging systems: a descriptive summary, *Library Hi Tech*, 11(2), 67-78.

Foner, L.N. (1995), Clustering and information sharing in an ecology of cooperating agents, *AAAI Spring Symposium on Information Gathering from Hereogeneous Distributed Environments*, Stanford, CA.

Frenzen, J. and Nakamoto, K. (1993), Structure, cooperation, and the flow of market information, *Journal of Consumer Research*, 20(3), 360-75.

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.C. and Terveen, L. (1996), Using frequency-of-mention in public conversations for social filtering, *Proceedings of the the ACM Conference on Computer Supported Cooperative Work((CSCW'96)*, Cambridge, MA, ACM PRess, 106-12.

Hill, W.C. and Hollan, J.D. (1992), Edit wear and read wear, *Proceedings of the Conference on Human Factors in Computing Systems. CHI'92*, Monterey, CA, ACM, 3-9.

Hill, W.C. and Hollan, J.D. (1993), History-enriched digital objects, *Third ACM Conference on Computers, Freedom and Privacy*, San Francisco, CA, ACM, 917-20.

Hill, W.C. and Hollan, J.D. (1994), History-enriched digital objects: prototypes and policy issues, *The Information Society*, 10(2), 139-45.

Hill, W.C., Stead, L., Rosenstein, M. and Furnas, G. (1995), Recommending and evaluating choices in a virtual community of use, *Proceedings of the Conference on Human Factors in Computing Systems (CHI'95)*, Denver, CO, ACM, 194-201.

Hjerppe, R. (1989), HYPERCAT at LIBLAB in Sweden: a progress report, in Hildreth, C.R. (Ed.), *The Online Catalogue: developments and directions*, London: The Library Association, 177-209.

Ingwersen, P. (1996), Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory, *Journal of Documentation*, 52(1), 3-50.

Kantor, P.B. (1993), The adaptive network library interface: a historical overview and interim report, *Library Hi Tech*, 11(3), 81-92.

Kelman, H.C. (1961), Processes of opinion change, *Public Opinion Quarterly*, 25, 57-78.

Kling, R. and Elliot, M. (1994), Digital library design for organizational usability, *SIGOIS Bulletin*, 15(2), 59-70.

Koenig, M.E.D. (1990), Linking library users: a culture change in librarianship, *American Libraries*, 21(9), 844-9.

Kuhlthau, C.C. (1991), Inside the search process: information seeking from the user's perspective, *Journal of the American Society for Information Science*, 42(5), 361-71.

Kurth, M. (1993), The limits and limitations of transaction log analysis, *Library Hi Tech*, 11(2), 99-104.

Maes, P. (1994), Agents that reduce work and information overload, *Communications of the ACM*, 37(7), 31-40.

Maltz, D. and Erlich, K. (1995), Pointing the way: active collaborative filtering, *Proceedings of the Conference on Human Factors in Computing Systems. CHI'95*, Denver, CO, ACM, 202-9.

McGrath, M.A. and Otnes, C. (1995), Unacquainted influencers: when strangers interact in the retail setting, *Journal of Business Research*, 32(3), 261-72.

Menzel, H. (1959), Planned and unplanned scientific communication, *Proceedings of the International Conference on Scientific Information*, Washington, DC, National Academy of Sciences, National Research Council, 199-243.

Park, C.W. and Lessig, V.P. (1977), Students and housewives: differences in susceptibility to reference group influence, *Journal of Consumer Research*, 4, 102-10.

Penniman, W.D. (1975), *Rhythms of dialogue in human-computer conversation*, Unpublished PhD Thesis, Ohio State University, Columbus.

Reisner, P. (1966), *Evaluation of a "Growing" Thesaurus*, Research Paper RC-1662, IBM Watson Research Center, Yorktown Heights, N.Y.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994), GroupLens: an open architecture for collaborative filtering of netnews, *Proceedings of Conference on Computer Supported Cooperative Work (CSCW'94)*, Chapel Hill, NC, ACM Press, 175-86.

Rice, R.E. and Borgman, C.L. (1983), The use of computer-monitored data in information science and communication research, *Journal of the American Society for Information Science*, 34(4), 247-56.

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

Rosenbaum, H. (1996), In the trenches o f the digital revolution: intellectual freedom and t he "public" digital library, *Paper presented at ASIS 1996 MidYear Conference: The Digital Revolution*, San Diego, CA.

Ross, J. (1983), Observations of browsing behaviour in an academic library, *College & Research Libraries*, 44(4), 269-76.

Rotenberg, M. (1992), (Chair) Privacy and Intellectual Freedom in the Digital Library, *Panel at the Second Conference o n Computers Freedom and Privacy*, Washington, DC, `http://www.cpsr.org/dox/conferences/cfp92/rotenberg.html`

Sandore, B., Flaherty, P., Kaske, N.K., Kurth, M. and Peters, T. (1993), A manifesto regarding the future of transaction log analysis, *Library Hi Tech*, 11(2), 105-6.

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

Stoan, S.K. (1984), Research and library skills: an analysis and interpretation, *College and Research Libraries*, 45(2), 99-109.

Twidale, M.B. and Nichols, D.M. (1996), Collaborative browsing and visualisation o f the search process, *Aslib Proceedings*, 48(7-8), 177-82.

Twidale, M.B., Nichols, D.M. and Paice, C.D. (to appear), Browsing is a collaborative process, *Information Processing & Management*.

Waine, B. (1995), A disaster foretold? The case o f the personal pension, *Social Policy & Administration*, 29(4), 317-34.

Walker, C. (1995), Word of Mouth, *American Demographics*, 17(7), 38-44.

Wittenburg, K., Das, D., Hill, W.C. and Stead, L. (1995), Group asynchronous browsing on the World Wide Web, *Proceedings of the Fourth International World Wide Web Conference*, Boston, MA, O'Reilly & Associates, 51-62.

Yan, T.W., Jacobsen, M., Garcia-Molina, H. and Dayal, U. (1996), From user access patterns to dynamic hypertext linking, *Computer Networks and ISDN Systems*, 28(7-11), 1007-14.

Zhou, N. and Abdullah, Z. (1995), Canadian matchmaker advertisements: the more things change, the more they stay the same, *International Journal of Advertising*, 14(4), 334-48.