

Fiction selection - an AI at work?

Keywords: AI, expert system, fiction selection

Alan MacLennan

School of Information and Media

Robert Gordon University

Aberdeen

Abstract

This paper briefly considers the historical dimension of the treatment of fiction in libraries and concludes that, if fiction is regarded as a resource, library users are entitled to enhanced access to it. It is noted that enhancing access to fiction is widely regarded in the profession as unacceptably labour-intensive, but suggests that readers are favourably impressed by an expert system which requires relatively little input effort on the part of the library staff. The idea of a "user profile" is described, as is its restricted implementation in a prototype system. A recent study in Aberdeen supported this thesis. Plans for further development of the prototype system are described.

Introduction

"It is certainly not the function of the public library to foster the mind-weakening habit of novel-reading among the very classes - the uneducated, busy, or idle - whom it is the duty of the public library to lift to a higher plane of thinking" W M Stevenson

The above quote, from the 1896 Annual Report of the Carnegie Free Library of Allegheny, Pennsylvania, is quoted in her 1991 article by Catherine Sheldrick Ross(1), who notes the resurgence in the USA of Reader's Advisory Services, after a period of lesser popularity in the past 20 to 30 years. Mary Chelton(2), writing in 1993, notes the same phenomenon. Ross's article is based on over 100 open-ended interviews, in which the readers quoted display great enthusiasm for fiction. She quotes a British survey by Taylor and Johnston which showed that 63.7% of public library users responding stated that their reason for use was "general recreation or leisure".

As may be surmised from Stevenson's quote, the advisory service was originally intended to "improve" the content of the reading material used by the public, ie to move away from fiction and towards "serious", "solid", educationally improving non-fiction. Ross theorises that the renewed popularity of the advisory services is because of "new thinking about popular culture and pleasure-reading". She cites evidence and opinion that reading fiction can be educational, therapeutic and can help people to cope with situations they encounter in their lives. She was struck by the intensity of response to the question "What is the importance of reading for you?" from her respondents. Typical answers referred to life and death, compulsion and physical need. Women, in particular, related experiences of deriving psychological support from reading fiction. It seems, then, that the public library service can serve an important function to its users in this regard, and if such is the case, there may be grounds for suggesting that it actually has a responsibility so to do. A reader's advisor, using interview techniques and tools such as Fiction Index or journal guides to genres, can help match the book to the reader: "A book that suits a reader in a

particular situation is a good book", says Ross(3), but she also notes the Book Industry Study Group's survey of American library users, in which "readers put 'recommended by a librarian' as nineteenth on a list of thirty-eight possible factors"(4) helping them to select a book.

It seems possible that, despite annotations, displays, guides and advisors (where these are available) the reader is still not as well catered for in her quest for "rich resources of comfort and stimulation and renewal"(5) or simple entertainment, as she might be.

The resource implications for improving access to fiction being considerable, and there being as yet no clear favourite in the schemes proposed by, for example, Baker(6), Wells(7), and Walker(8) (although Beghtol's(9) fully faceted 1994 scheme looks promising, there may not be the widespread commitment to apply it, due to its considerable complexity), perhaps the way opens for a different approach. Instead of, effectively, asking the reader what he wants, and then telling him where to find it, why do we not get to know him, then suggest something that he might like?

GRUNDY - user profiling

Rich's(10) GRUNDY system attempts to construct rapidly a profile of a user as an individual, by modifying a series of stereotypes on the basis of its interactions with the user.

This technique is modelled on psychological descriptions of human thought processes, and results in the system effectively classifying the user, by applying a series of facet-value combinations to her. "Stereotypes are simply collections of facet-value combinations that describe groups of system users". Stereotypes are triggered by user input, so that, for example, the use of an advanced search strategy might be the trigger which "fires" an "expert-user" stereotype, which is then used to modify the internal representation of the user, which the system is constructing throughout the interaction.

Such observation of strategy used might be considered as a trigger implicit in the interaction, but GRUNDY can also elicit information explicitly, in order to clear up uncertainties such as the precise reason why a recommended book was rejected.

The system holds each piece of information with an associated weighting, a calculated probability that it is true, which may be modified by subsequent observations, or further "firing" of stereotypes. More than one stereotype will normally fire for each individual, so that the fact that no stereotype contains all the facets with which the system is capable of dealing, is not a disadvantage.

An example Rich uses is the stereotype for SPORTS-PERSON, which contains high values for Interests - Sports, Thrill and Tolerate-violence, all with a fairly high degree of certainty as to their applicability to the canonical SPORTS-PERSON. These are facets missing from the FEMINIST stereotype, which in its turn has facets which the other lacks. However, a feminist sportsperson user would have values for all the facets, so that the system's picture would be rounded out.

When a new stereotype is activated, the current user synopsis, is updated accordingly, the system calculating how much the new stereotype

should influence existing facet values, on the basis of the rating of the stereotype (how likely it is considered to be that the stereotype will apply), and the strength with which it predicts the value of a given facet, in combination with the existing value for the facet.

Having constructed a User Synopsis, GRUNDY proceeds to recommend novels it considers appropriate. It selects the user's salient facets, those having a higher than average probability for a "non-middle-of-the-road" value. By selecting one of these facets at random, it builds a set of books considered to match the facet value. The example Rich uses is that of a user with a high value in the EDUCATION facet "generating" a set of books which score highly on PHILOSOPHY and LITERARY MERIT. Each book in the set is compared with the User Synopsis, and the book most closely matching the profile, facet by facet, is selected.

GRUNDY again uses the Synopsis in order to decide which of the book's features to mention in its recommendation to the user, since it holds more information about the book than is necessary for any one recommendation.

Rich examines areas in which GRUNDY could be improved, but it appears to have performed satisfactorily in the limited number of trials mentioned in the paper. There is considerably more complexity in the structure of stereotypes, triggers and User Synopses than it is appropriate to discuss here. However, one question which arises immediately is that of how the system acquires its extensive knowledge of the books in its database. Although the system is apparently rather good at acquiring and deducing knowledge about its users, which it appears to classify in a faceted, "fuzzy" manner, that is, using measures of certainty to modify absolute judgements, the author does not mention how a classification is applied to the books on the database. It seems reasonable to infer that the process must result in a profile similar to the completed profile for a user, in that the final choice of a book to recommend from the set selected is made on the basis of closeness of match of its facets with the user profile. It would, however, be very interesting to know by what process the book's facet values are calculated. Perhaps an arbitrary measure against collection average scores would tend to be most successful, but this leaves room for a large element of subjectivity to enter the process.

Rich claims that computers can make more effective use of stereotypes than can human beings, because, for one thing, they have no "emotional attachment" to a stereotype which they have previously applied. This can certainly be seen as an advantage, the computer has no self-image to endanger by admitting that it was wrong, so to say, but it seems that, although User Synopses, and perhaps the stereotypes and their weightings, are modifiable by the system, as it gains more experience and learns the applications of more new terms through conducting its "user interviews", the attributes of the books on the database are a "given".

BROWSER - a knowledge based system

Perhaps there is room for a "cheap and cheerful" AI, something which might take on the role of a helpful fiction librarian, knowing the patrons, with no extensive knowledge of systems for fiction classification, but able to recommend for each reader, something that there is a reasonable chance they might like - a "fuzzy librarian", who

knows your tastes, and knows the library stock, especially new material with which you might not be familiar.

The concept of BROWSER was originally that of a system which would be capable of recommendations on a par with those of an experienced librarian - the hypothetical "expert" - who was familiar with the reader's tastes in fiction, as might be the librarian in a small public library. Alasdair Smith (11) points out that the term "expert system" is an unfortunate one, in that it creates expectations of "rather more expertise" than has actually been achieved. He prefers to use the term Knowledge Based System, or KBS, while acknowledging that expert system is the more widely used. He lists the preferred areas for KBS's thus:

What factors distinguish problems that are amenable to the KBS approach? KBS's are best applied to expertise which:

- relates to narrow and clearly defined tasks.
- is based on knowledge, can be described with facts and rules, rather than common sense.
- relates to tasks that take between a few minutes and a few days.
- is available from articulate, cooperative experts
- has consensus about solutions.
- an expert can use to describe over the telephone how to solve the problem
- does not require human skills that are difficult to computerise
- is not trivial and possessed by most people

It is debatable to what extent the "problem environment" matches these guidelines. The task is clearly defined - select a new book which may be to a reader's liking, given previous choices she has made. Is this skill based on knowledge, rather than common sense? An implementation of BROWSER would be based on the accumulated experience of librarians who work with fiction, and the public who read it - if the next selection was a matter of "common sense", why not determine it by asking the "man in the street"? The time element seems correct, one would hope to find "articulate, cooperative experts" in the field, although they might not be able to express these skills over the telephone! The question of consensus will have to await further research, as will the question of how difficult the skills are to computerise. Finally, if the skill were trivial and possessed by most people, there would not be such frequent mention in the literature of librarians receiving requests of the type "Could you give me a good book (like.....)".

In conclusion, although some of Smith's criteria are as yet undetermined for this system, there do not appear to be any which would exclude the area as suitable for a KBS. The system would keep a record of issues to a reader, who would, on returning an item, be prompted to give it a "score", indicating how well it had matched her requirements. The system would then attempt to select, from new additions to stock, a book which matched this "reader profile". A very simple system on these lines was written for a research proposal, using the PROLOG programming language, which is well suited to developing applications of this type. The demonstration program simply held a list of books read and scores awarded them by a small number of fictitious "readers". It then attempted to find, in the list of new books, a book from a genre, or by an author, previously awarded a high score (arbitrarily set at 7 out of 10). Although this system sufficed for demonstration of the idea for

the purposes of the research proposal, it became apparent that it had shortcomings which would make it inadequate for the survey discussed here. There were only a few "readers" and the lists of books read were artificial. Details held of each book were restricted to author, title and a one-word "genre" type of classification. It was decided, therefore, that a substantial rewrite of the system was required for this survey, in order to give the demonstration a more realistic dimension. A longer list of "old" books was required, representing a wide variety of fiction, with a sufficient number of books of each type to make any clear preferences as to type apparent. It was also decided that an attempt would be made to reflect the respondent's own tastes in fiction, an approach which was "riskier" to implement, but which, it was hoped, would be more convincing, if successful.

It became apparent at a very early stage in the project that the system would have to use some form of fiction classification. This has been seen as a significant drawback by some of the professional librarians with whom the idea has been discussed, but it seems that there are three relevant arguments demonstrating that this is not the case.

First, a primary aim of this type of system is that of encouraging greater exploitation of the library's fiction collection, by drawing the attention of readers to material which they might not otherwise consider reading. Although there are publications, such as the Fiction Index, which attempt to provide access to fiction by subject, period and nationality of author, these do not appear to be used commonly by readers, who may well be unaware of their existence, or disinclined to use them by their simple Subject: Author, Title presentation. A system of this type has great possibilities for enrichment of the searching process. It would be relatively easy to link a brief description of a book, such as that currently given by the demonstration system, to a graphical representation of its cover, publisher's "blurb", or contents page, if appropriate. This is merely making use of descriptive material which is already provided, at some expense, by the publisher, but is made relatively inaccessible to the browsing reader, except when a selection of books is presented in a display.

Second, it is exactly this type of information which would be used in what might be termed a "first level" classification, or in guides of the "Now read on..." variety. It is not suggested that a librarian should read through every new work of fiction in order satisfactorily to classify it, but that the information needed is deliberately and explicitly presented on the covers and flyleaf of most books. It is not the intention of the publishing industry to obscure the type of publication, although its openness with respect to quality may be more suspect. It might even be supposed that publishers themselves would be prepared to give a subject description in accordance with an agreed thesaurus as part of the Cataloguing-in-publication data, for inclusion in the MARC record. The question of quality will arise again in a historical context.

Third, there is both a historical and a current precedent for classification of fiction. As discussed earlier, schemes for the classification of fiction have abounded in the literature, and it is common practice in libraries to subdivide the more obviously genre-targeted fiction stock into sections labelled "Romance", "Western", "Crime" and so on. Fiction has, however, remained the "poor relation"

of library stock as regards its subject access. A more extensive classification system was also required, in order that the choices made by the system be more accurately reflective of what a reader might actually be expected to choose from the list of "new" books.

It must be made clear that this classification system, while probably sharing some of the features of the system which might eventually be used in a working product, is by no means as complex as the finished system would be required to be. As with all details of the current system, it is for demonstration purposes only, both to give an idea of what the finished system might look like, and to suggest new ideas for development. The classification used might be thought of as using genre divisions, in the same sort of way as Baker, combining the facets in the same sort of way as Walker, but including elements other than subject, with a bow in the direction of Pejtersen(12). The complexity possible in the combination of subject terms has been demonstrated in Farradane's(13) elegant scheme for relational indexing, in which nine categories of relationships are identified by operators in such a way that complex subjects may be specified by a 'chain' of terms and operators. However, as Farradane points out, "This paper concerns the classification of knowledge as a whole, and not only the relatively simple classification of limited groups of subjects"(14). It is probable that a scheme so general in applicability could be used here, but only at the expense of excessive time spent in indexing. The topics and modes of fiction are not (or not only) concrete scientific topics and the scientific mode, but include mood, fantasy, emotion, irony, metaphor - concepts that are better expressed in terms of genre, where one label can "place" the reader in the correct conceptual space, and further specifications can pinpoint the desired item more closely. Consider the following (imaginary) conversation:

A: "I like spy stories, adventure, thriller, that sort of thing"

B: "Do you mean, something like John Buchan, Ian Fleming's Bond books, John le Carre, or something else?"

A: "Well, Buchan's a bit too chauvinist and 'gung-ho', Fleming's out of date since the end of the Cold War, and a bit too silly. I do like le Carre, but I've read all his"

B: "Yes, le Carre's very much of his own style, but how about Ted Allebuery or Adam Hall?"

This demonstrates what an "active" system might be able to do (as B). Of course, it implies handling of natural language at a sophisticated level, but the search strategy is simple Boolean. The next step is to identify the requirement without asking the questions - more prone to error, but less demanding of the reader.

The system need not concern itself with shelf order, or the combination of facets, but simply with the selection of an item appropriate to the user's requirements, and not even necessarily the most appropriate in the collection, although that would be the ideal. A simplified set of facets was chosen, which might be described as: Author, Title, Genre, Date of publication, Nationality of author, Treatment. Foci from these facets were used both to decide an appropriate selection of item and to describe it to the user. Thus, a work held on record as :

newbook(099,William_Gibson, Neuromancer, science_fiction, 20th_century, American, cyberpunk)

would be presented by the system thus:

Neuromancer by William Gibson
a science fiction novel
by a 20th century American author
with a cyberpunk treatment

The new version of the software was based on the idea that the user would select, from a list of about fifty books, five books which they had read and enjoyed. The system would then make an appropriate selection from a shorter list of "new" books. The requirement for a "score" had been abandoned by this time, partly because it was felt that it might require more thought to be given to input than a typical user would be prepared to give when using a runtime system; partly because making meaningful use of scores would add excessive complexity to the demonstration system; and partly because it had not been decided whether a simple score out of ten, for example, could reflect in a useful way the necessarily complex evaluation which a reader makes of a work of fiction. Rather than elicit a statement from the user which said, effectively, "These are the books I've read: I liked these ones and I did not like these ones", it was decided to attempt to build only a positive profile of the favoured books. This is, again, a "riskier" method, but it was felt that it would pay off in terms of processing time and greater ease of user input - it is perhaps more difficult to evaluate how much one did not like something, than how much one did like something.

The system attempted to create "profiles" of books that the reader might enjoy, by assigning and accumulating weighted scores to several internal templates. For example, it was decided that the fact that a reader had enjoyed a book by an author would probably dispose her favourably to another book by the same author, the more so if she had enjoyed more than one book by the author. Citing a book thus produced a fairly heavily weighted profile, containing only the author's name, the other fields being blank. Although it was reasonably safe to assume a continuation of nationality and period (roughly) for an author, this is not necessarily the case with genre, so a subsequent choice of a book by the same author and in the same genre as a previous choice would add correspondingly greater weighting to the "Author, Genre, blank, blank, blank" profile. Title was not useful for any purpose other than display. Genre, period, treatment, on the other hand, would score well, if repeated, the program working on the principle "If there is no profile like this, create one and give it a score of x. If there is already such a profile, add to its weighting." This gave great flexibility as to how the program made its final selections. Profiles which accumulated scores less rapidly were such as; "Blank, Blank, 20th century English author, blank", which, if they were the only profiles fulfilled by the new books, would probably be fulfilled by any, but equally might be significant if the reader had developed a taste for an historical period or the literature of a country producing less material.

This version, then, required a list of about fifty books, of which it could reasonably be assumed most people would have read five. To aid in the compilation of this list, contributions were solicited from staff in the Robert Gordon University library, and these were combined into a list. The system and questionnaire were then piloted on these and other

colleagues, with reasonable success, though allowing the input of one or more books would be preferable. After due consideration, the software was modified accordingly, and was also amended to offer further selections from the "new books" database, if there were any more appropriate books. In this matter, it should be explained that the system does not rank its selections exhaustively; it attempts to find a "good" match for one of the reader's profiles, then a "medium" match, then a "poor" match. The system merely indicated a "good" match by phrasing its recommendation, "You ought to like...", a "medium" match was phrased "You should like...", and a "poor" match as, "You might like...". This was done partly to convey the system's "confidence" in its choice to the user at a non-obtrusive level, and partly to indicate to the operator which level of match the system was making, in order that its choice could be further explained to the user, if necessary. The books in the demonstration system were selected so that no entry of a single book could fail to find at least one poor match, so that all entries were, as far as possible, assured of success. The omission of ranking, a feature which would be most desirable in an operational system, was a decision taken to keep processing time to an acceptable level.

Methodology

The system was demonstrated in several environments: to staff of an academic library, to students at a school of Librarianship and Information Studies, in public libraries and in city centre commercial premises. Respondents were "self-selecting" - only those who volunteered when approached were included in any count. Reactions seemed generally positive, however, although the public libraries yielded fewer respondents than had been expected. Some librarianship students had the system left with them to experiment with, rather than its being demonstrated, and some ten copies of the system were sent to those who responded by e-mail to a request for help posted to the lis-link newsgroup, accessed via JANET. An answer to this request was received from Ireland, one from Canada, and one from New Zealand, as well as several from the UK, demonstrating some international interest of the library community in the topic.

Questionnaire responses

In answer to the question, "Would you be interested in using a system of the type just demonstrated, if one were available in your public library?", Out of 63 respondents, 61 (96.8%) replied "Yes", and the response to the system was generally very favourable indeed. 31.7% had read and enjoyed one of the books recommended, 1.5% said they had read, but not enjoyed, 63.4% had not read the selection, but would consider doing so, 14.7% had not read the selection, and would not consider doing so. Some precision was lacking in the question as posed, in that users of the later version were able to have more than one recommendation, if more than one suitable "new" book appeared on the database. In response to the question, "Which of the following features would you like to see/use in such a system?", 76% of respondents favoured option (a), being able to find new books of a type, or by an author, they had read before; 73% option (b), any books in the stock of a type, or by an author, they had read before; 52% option (c), the "top 10" most borrowed of a specific type; 31% option (d), the "top 10" of a type, but only those currently available; 60% option (e), the location of books - which shelf?; and almost 81% option (f), books on related topics of non-fiction, if they read non-fiction

31% indicated that they would prefer the system to be operated by a member of staff on their behalf. 92% would give a score as feedback to help the system, and this correlates quite well with the fact that, 3% "would mind if records of loans were kept", and 9% "would mind if a 'user profile' was kept" However, 77% "wouldn't mind if records of loans were kept" and, 65% "wouldn't mind if a 'user profile' was kept".

90% replied that they would appreciate a quick "personalised" recommendation from the system, if they were in too much of a hurry to browse.

Unsurprisingly, almost 97 % read fiction, the most popular of the genres suggested being humour (39%), science-fiction (38%), classics (36%), historical (34%), thriller (34%), and crime (30%).

When asked in to suggest types of fiction which they read, other than those named, 3 respondents named "modern or 20th century", 2 "female" or "feminist", 2 "horror", and 1 each of "travel", "popular", "Scottish", "racy", and "cyberpunk".

Other descriptive words supplied in response to the question are loosely grouped, with number of respondents in brackets:

female author/publisher, feminist, modern female (3)

exciting cliffhanger, suspense, action (3)

mystery, suspense (2)

"satirical amusing thought-provoking escapist, stimulating" and

"original, engrossing, thought-provoking, realistic" (2)

happy (1)

first person (1)

not western (1)

Victorian (1)

good author (1)

good story (1)

biographies (sic)(1)

One question asked which factor the respondent felt influenced them most when selecting a fiction book. Of the 63 respondents, 7 gave two answers to this question, one of the answers in each of those cases being the seventh option, "familiar author". It seems reasonable to suppose that this counts as an "afterthought" in these cases, the respondent having previously selected from higher on the list before noticing this option. This leads to the percentages totalling to over 100% for this question, and may indicate that it should have been better worded, but the responses are interesting, nonetheless. 40% of respondents chose "word-of-mouth recommendation", 30% chose "familiar author", 17% chose "published review", 14% chose "cover 'blurb'", 13% chose "book in a section such as 'Romance' 'Western', 'Crime', etc", and one respondent felt most influenced by an "attractive cover". No respondents cited advertising as their primary influence, however, although it is possible that this is due to peoples' reluctance to admit that they are influenced by advertising.

It emerged that just over 52% of the sample were female, 95% had some experience in using a computer terminal and 73% had experience of using a computerised library catalogue. 96% replied yes, they would like to be able to find fiction by its subject, using a computerised catalogue. 42% of the respondents were in "library - related" occupations, either as library staff or as students of Librarianship. The self-selecting nature of the sample can be seen to have an influence here, in that those people who were interested to see the computer used for the survey were also those who agreed to take part, and it might be supposed that experience of using computers would tend to make people more interested to see one used in an unfamiliar context. There was a degree of "technophobia" evident in some of those approached, particularly the older age groups, but as these people also refused to take part in the survey, this is only a subjective impression, as no record of "failed" approaches was kept.

51% were professionally employed, 29% were students, 13% were non-professionally employed and 6% were retired. One person was not currently in employment. As regards age, 35% were between 20 and 30, 29% between 30 and 40, 25% between 40 and 50, and 6% between 50 and 60. It was obvious from the people in the area while the survey was being conducted that the sample responding was not representative of those present, nor of the public at large. The best that can be said is that the sample may be representative of those prepared to answer it, and is skewed towards students and workers in the field of librarianship.

Data Protection Act 1984 - Considerations

The situation as regards the Data Protection Act 1984 (15) would vary according to the level of system implemented. If, for example, the system was fully integrated with an existing integrated automated library system which incorporates circulation functions, the Act would be applicable. It would be reasonable to suppose that the existing system would hold "personal data" within the meaning of the Act, i.e. "data consisting of information which relates to a living individual who can be identified from that information (or from that and other information in the possession of the data user) including any expression of opinion about the individual" .

Information held as to a borrower's "history" or "profile" would be held "in a form in which it can be processed by equipment operating automatically in response to instructions given for that purpose". A library operating an automated system holding personal data would require under Section 4 of the Act, to be registered as a "data user" but, because of the change in the nature and description of the data held, would be required to apply to have its entry in the Register changed, in order to comply with Sections 5 and 6. The library would also be obliged, in response to a written request, and "on payment of such fee (not exceeding the prescribed maximum)" as the data user (i.e. the library) "may require", to inform an individual whether the individual is the subject of any data held, and to supply the individual with a copy of any such information, including an explanation of any information "expressed in terms which are not intelligible without explanation".

Other requirements laid on the data user concern non-disclosure of personal data, except to the individual to whom it applies, maintenance of data security, accuracy of data held, and that data should not be kept longer than is necessary for the purpose for which it is held. These points are covered in the Data Protection Principles which comprise Part 1 of Schedule 1 of the Act. Data may be held indefinitely for historical, statistical, or research purposes if they are "not used in such a way that damage or distress is, or is liable to be, caused to any data subject" (Schedule 1, Part 2, 7). The obligations are to register and specify the purpose of data collection, and to conform to the Principles.

It is not difficult to imagine circumstances in which, without subscribing to any grandiose conspiracy theory, a reader might simply not care to have the staff of her local public library be able to access easily her reading list for the preceding period. Davies(16) says "There is a view held, based on experience of cases and practices in other countries, that a record relating to a loan and return of an item should be destroyed or discarded immediately, or soon after the cycle of transactions has been completed. Fearing the potential of building up dossiers on people's material borrowing and consultation habits...some observers advocate the yearly destruction of material capable of being applied in this way". Although the system could comply with the Act, then, the user may feel insufficiently protected by the Act's provisions. It is even conceivable that the knowledge that such a system was in use might discourage people from using a particular

library, although it would in any case be desirable to ask users whether they wished to participate in that aspect of the system, and, under the terms of the Act, it would be a legal obligation to inform those who enquired what information was held regarding them. Davies puts this latter point well, when he says "Informing data subjects of data protection arrangements is not only common courtesy but fulfils many of the objectives of the law and the Principles".

Conclusion

The primary question which must be addressed is, how appropriate is an "expert system" type of program for this application? The system demonstrated displays the features of an expert, or knowledge-based, system, in that it implements a set of rules, or heuristics, which are obtainable with relative ease from a human "expert", or which may at least be derived by a process of trial and error from a series of postulated rules which can be refined, by the adjustment of weightings, to reproduce the operation of such rules. The responses to the questionnaire indicated that there does exist a demand for a system of this type, offering subject access to fiction, although respondents not involved in the library professions are less prepared to exchange the anonymity of current systems for the implications of data storage and the perceived effort of providing the feedback necessary to make it more efficiently interactive. The success of the system's operation, as perceived by respondents when operating with a comparatively simple set of rules, is encouraging, in that it indicates the potential of which a more intensively developed system might be capable. If the input and maintenance aspects of the system were enhanced, it offers a facility which users indicated they would value, at a cost in workload to the library which is rather less than that which might be feared by professional librarians concerned by the spectre of "classifying fiction". For the system to function, at a higher level than demonstrated, it would require only the creation of a carefully thought-out list of "known books", and the regular maintenance of the list of "new books", which would need classified at a relatively superficial level. Obviously, the more effort that could be expended in such classification, the greater would be the value of the system, but even at the superficial level demonstrated, the questionnaire results appear to show that it intrigued and impressed those users who were prepared to try it out.

Further plans

The language in which the program is written must be changed. PROLOG, especially the "cut-down" Public Domain version which was used, runs too slowly on the machines, typically IBM PC- compatible, which might be expected to be found in the "live" environment. However, now that the basic logic of the program has been worked out, it ought to be a much easier matter to develop a faster and visually superior version in a language such as C or C++. A change in language also opens the door to another exciting potential development, the possibility of including hypertext links to other documents sharing some attributes with those selected, either by the user or by the program itself. This facility could be imagined as being in some ways similar to the "Navigate" option offered by some OPAC systems, in that it would allow the user to jump the display immediately to other relevant documents, but in which she, not the OPAC supplier, would determine which factors might constitute relevance in her search.

Work is currently in hand to rewrite the system in the "C" language, with the attendant benefits which have been indicated above, and it is hoped that a small "stand-alone" system for the smaller library may be produced as the next step in its development.

References

1. Ross, C S: Readers' advisory service: new directions. RQ 43 (Spring 1991) pp 85 - 98
2. Chelton, M K: Read any good books lately?: helping patrons find what they want. Library Journal, May 1 1993 pp 33-35
3. Ross, p 85
4. Ibid, p 91
5. Ibid. p91
6. Baker, E A. The classification of fiction. Library World (1) 11, March 1899 pp 198-200, 216-221.
7. Wells, A J: Classification of fiction. Library World 37 (1934)
8. Walker, R S. Problem child : some observations on fiction, with a sketch of a new system of classification. Librarian and book world (47) 2, 1958 pp 21 - 28
9. Beghtol, C. Access to fiction : a problem in classification theory and practice. Part II. International Classification 17 (1990) pp 21-27
10. Rich, E. User modelling via stereotypes. Cognitive Science 3 (1979) pp 329-354
11. Smith, Alastair : CyberReference? - An "expert" system for New Zealand reference work. Paper for NZLA conference, 25 September 1991.
12. Pejtersen, A M and Austin, J. Evaluation of a search system based on users' value criteria (Part 1). Journal of Documentation (39) 4, 1983 pp 230-246
13. Farradane, J E L. A Scientific theory of classification and indexing and its practical applications. Journal of documentation (6) 2, 1950 pp 83-92
14. Ibid, p83
15. Great Britain Laws, Statutes: Data Protection Act 1984. London, HMSO 1984. (Public General Acts Series: 1984 Chapter 35)
16. Davies, J E: Data protection - a guide for library and information management. Oxford, Elsevier 1984