Pebbles for the Mosaic of Cataloging Expertise: What Do Problems in Expert Systems for Cataloging Reveal about Cataloging Expertise?

Alenka Šauperl and Jerry D. Saye

Twenty-five years of research in expert systems for descriptive cataloging and related areas are reviewed. Researchers who developed prototype expert cataloging systems in the 1980s found that cataloging rules are extremely comprehensive and complicated, but still insufficient to permit these systems accurate cataloging results. Those researchers also identified smaller areas that need to be pursued for successful implementation of expert cataloging systems. Subsequent research has focused on narrower areas. Cataloging rules were studied and more was learned about problems with their logic structure and organization. Another focus was optical reading of bibliographic elements in documents. Categories of responsibility and graphic design continue to pose problems. The visual characteristics of documents were studied to understand more about the automatic recognition of bibliographic elements necessary for the bibliographic description of documents. The systematic study of the cataloging process, necessary for development of expert systems, may result in improvement of manual working procedures and enrich the education of new catalogers.

CATALOGING AND EXPERT SYSTEMS

Cataloging is one of the most time-consuming tasks performed in libraries. It is not surprising then that, since the introduction of computers into library operations, there have been numerous attempts to automate the cataloging process. The early use of computers to create catalog records that were used most often to produce cards for card catalogs has been largely supplanted by the creation of bibliographic records that are used to populate online catalogs. The most challenging aspect of cataloging is to determine the content of the bibliographic record rather than to create the record itself. This is where research in expert systems and related areas plays an important role.

ALENKA ŠAUPERL (saupa@ruby.ils.unc.edu) is a doctoral candidate and JERRY D. SAYE (saye@ils.unc.edu) is Associate Professor, School of Information and Library Science, University of North Carolina at Chapel Hill. Manuscript received March 9, 1998; accepted for publication July 2, 1998.

In expert systems computers are used to organize and preserve the expert knowledge of catalogers and to employ that knowledge in cataloging with minimal or no human intervention. Unfortunately, few of these prototype cataloging expert systems have succeeded in the real world. They have contributed, however, to an understanding of cataloging processes and standards, as well as to an understanding of the documents cataloged. When researchers in this area determined that the entire cataloging process was too complex to allow for full expert system implementation, they then concentrated on some elements of the cataloging process, such as cataloging rules, document characteristics, and the characteristics of names used as access points. These efforts resulted in a better understanding of both the advantages and limitations that computers offer to catalogers. Read from another perspective, however, the problems found in having computers create cataloging might be indicative of the problems catalogers themselves face. This might be particularly true for novice catalogers and students of cataloging.

Cataloging standards are an essential part of the knowledge base for the cataloging process. They state the purpose, procedure, and results of the process. The most important standards used in cataloging in the United States are the *Anglo-American Cataloguing Rules,* 2d ed. *(AACR2)* and the documentation that supports the MAchine-Readable Cataloging format (MARC). These are supplemented by national and local written policies and guidelines as well as by the unwritten practices and policies of libraries, library systems, and bibliographic utilities. Catalogers must learn and master all of these in order to perform their jobs successfully.

The purpose of a catalog is to create a surrogate of a document to support access to that document. A variety of document formats are represented in library catalogs. Document characteristics, in combination with expected retrieval needs, should determine what is contained in the catalog record. Catalogers learn to identify those characteristics, interpret them in accord with the existing rules, and represent them in meaningful, standardized descriptions.

The strength of developments in expert systems relative to cataloging is the process of knowledge acquisition. In this process, specific cataloging knowledge, decisions, and work procedures are systematically studied. The questions motivating the research described in this paper were: What have we learned in the 25 years of research into expert cataloging systems about the characteristics of the cataloging process, the materials cataloged in that process, and records created in the process? If expert systems do not provide the answer to cataloging problems now, can a better understanding of what cataloging is, characteristics of its processes, objects (materials), and results (records) help to improve the quality of work and the training process?

DEFINING EXPERT SYSTEMS

"Expert systems are computer-based systems that use knowledge and reasoning techniques to solve problems that would normally require human expertise" (Morris 1992, 1). The basic elements of expert systems include a knowledge base, an inference engine, an interface, and a general database. Each is described below.

The *knowledge base* is where human expertise is organized and stored. This base contains facts and heuristics of the domain of expertise. In the case of automated cataloging, the knowledge base would contain the current cataloging rules and modifications for local practice as well as the experience of catalogers.

The *inference engine* enables the formalized decision-making process. It makes use of the knowledge base to enable the system to perform like an expert. The engine infers results based on the stored knowledge.

The *interface* supports communication between humans and the system. Usually, the interface has three components. First, a user interface enables communication with the user (the cataloger who uses the system). Second, a developer's interface assists the knowledge

engineer (the person who develops the expert system) and makes the development of the system possible. Third, an external interface provides for data exchange with external sources (e.g., authority files or classification schedules).

The *general database* keeps track of the current problem and records all the relevant data and steps in the process.

Expert systems have shown a great deal of potential in the cataloging arena for several reasons:

• Human catalogers can be reassigned from routine tasks to work on more exciting and creative tasks. Thus, catalogers could leave simpler items to the expert system and work primarily with documents that are more difficult to catalog. Similarly, greater resources might be dedicated to other cataloging tasks such as enhancement of subject access.

• Expertise, which is rare and difficult to acquire, could be archived and saved for the future. When experienced catalogers leave a library, expert systems could be used as consultants or even as sources of training for less-experienced colleagues.

• Expertise could be distributed more widely and used more readily. In contrast to human experts, expert systems could be used 24 hours a day by multiple users at different sites.

• Critical examination of the decision process could be enhanced. The analysis of cataloging tasks necessary for the development of the system would force developers to understand and describe the work process systematically. This might identify errors and illogical steps in the workflow that were not obvious before.

• Full understanding of all parts of the process could help in the development of standardized approaches to problem solving for particular tasks. These standardized procedures might result in more consistent catalogs as well as support the training of novices.

With all their positive qualities, expert systems would seem to be the obvious cataloger's tool of the future. However, the realization of this ambition has

been impeded by a number of challenging obstacles:

• The building of an expert system requires a substantial amount of time and work from the builders of the system (the knowledge engineers) and from the experts in the particular domain (here, experienced catalogers).

• The expert system's domain needs to be carefully chosen within a narrow and well-defined area if the system is to be reliable. Despite original hopes, researchers have found that descriptive cataloging is too broad a domain. Although cataloging rules exist, they have not proven sufficient for accurate cataloging. Researchers have also found that cataloging rules are difficult to transfer into a knowledge base. In studies, it has been indicated that only when cataloging is divided into small subdomains will the resulting expert system be successful.

• Knowledge engineers, who are responsible for preparing the computerized version of the human expert, need to know the computer system well. They also need to be familiar with the expert domain in order for the required knowledge to be represented in the system in a meaningful and useful way.

• Knowledge engineers, who are also responsible for acquiring knowledge from human experts, frequently have been unable to get those experts to express completely what they think and do.

To explore more fully the potential of expert systems for use in cataloging operations, we reviewed 25 years of research. The studies examined yield important information on the fundamental requirements for designing the essential knowledge base for creating a standard cataloging record.

PROTOTYPE EXPERT SYSTEMS

Davies and James (1984) were the first to investigate the feasibility of creating an expert system for cataloging. They tried to develop a system that would provide a complete catalog entry (computer record or catalog card). Their system required a

cataloger to choose options from a menu about the type of publication under consideration and access points. After a series of questions about the publication, the system offered a template, called a frame, for the cataloger to record the remaining data for bibliographic description. Davies and James found their system very complex and difficult to manage. Two main reasons were identified: (1) *AACR2* is very complex; and (2) the general database, which kept track of the current problem, needed more space than their computer systems was able to provide at that time. In essence, the cataloging process was more complex than the existing technology available was able to manage.

Davies (1987) explained further how frames could be used in cataloging to infer and complete particular fields in a MARC record. He believed such a system should be able to infer who the publisher was from the International Standard Book Number (ISBN) and provide the complete publisher element. This could reduce some routine typing and typographical errors. Although not specifically stated by Davies, it is clear that accuracy—particularly in spelling—is vital. Errors in these areas are not trivial for they can have substantial impact on later retrieval. A good expert system would support the cataloging process with editing input.

Cataloging expertise, however, often lies in the choice of appropriate rules for bibliographic description and the choice of access points rather than in the simpler inference of publisher or the consistent spelling of words. The system created by Davies and James was intended to guide catalogers through the construction of the bibliographic record by providing rules relevant to each particular element of the description. Ultimately, this was found to be too complex for computer systems of the early 1980s.

While Davies and James worked on their system, Hjerppe and Olander (1989) built the Expert System for Simple Choice of Access Points for Entries (ESSCAPE). Hjerppe and Olander's system asked questions relevant to access points and the cataloger answered them.

This interaction resulted in the system suggesting the *AACR2* rule number appropriate for given situations. Hjerppe and Olander eventually abandoned further development of this system because they concluded that *AACR2* alone was not sufficient to allow for the automatic decision making in the selecting of access points. They found that human interpretation was essential because the cataloging rules were neither self-contained nor clearly formulated.

Evidently, the experience of catalogers enables them to develop more specific rules to given situations than are provided by the cataloging code for decision making. Varied experience can lead to differences in the interpretation of the information not provided for by the cataloging rules. Different or inconsistent inferences for missing information can result in inconsistent catalog records.

Gibb and Sharif (1988) also created an expert system to assist catalogers in creating full catalog records. By following menu options in the system and using "yes/no" and numbered decision alternatives, a cataloger was led to the appropriate *AACR2* rule. Based on their experience, Gibb and Sharif commented that using only *AACR2* as a knowledge source was appropriate. This would allow the lengthy process of knowledge acquisition from experts to be bypassed and avoid personal bias. Gibb and Sharif suggested that the system was particularly useful for training new catalogers.

Gibb and Sharif's optimistic conclusions contradict those of Hjerppe and Olander. This difference may be due to different expectations. Hjerppe and Olander were looking for a system that used *AACR2* to generate correct access points for the variety of items encountered in a real environment, and concluded that *AACR2* was insufficient for this purpose. Gibb and Sharif, on the other hand, developed a system for the accurate use of *AACR2* and were in general less concerned with cases in which those rules did not give sufficient guidance.

If cataloging is so complex, it seems reasonable to expect that a successful system might better be limited to addressing

specific document types. This would likely result in fewer rules being required to make the system operational. Ercegovac (1990) and Ercegovac and Borko (1992) followed this approach and created a semi-automatic cataloging advisor to assist catalogers in the cataloging of maps. Their system, "Mapper," helped catalogers establish main entry, title statement, statement of responsibility, publisher, place, and year of publication. Their expert system differed from previous efforts because the developers incorporated principles from:

- the system's performance testing;
- user interface design;
- different sources of knowledge for the development of the knowledge base; and
- knowledge elicitation methods to gain the necessary knowledge from map cataloging experts (Ercegovac 1990).

Ercegovac's method for system performance evaluation was to examine the results (the catalog records) and the user interface. Catalog records produced through Mapper were compared to the catalog records for the same maps prepared by the Library of Congress (LC). This standard for comparison was chosen because LC records are often treated as the U.S. national standard for quality cataloging. Although tested on only a small sample of maps, Mapper performed well and its user interface was well accepted by its users. Overall, it was perceived as helpful, easy to learn and fast.

Davies and James (1984) used cataloging rules and local guidelines as sources of expertise while Gibb and Sharif (1988) and Hjerppe and Olander (1989) used different parts of cataloging rules and, very likely, their own personal knowledge, although this is not mentioned in their reports. All commented on the weaknesses of cataloging rules and the adjustments that would be necessary to implement them in the knowledge base of an expert system. Ercegovac (1990) showed how to overcome these weaknesses with additional sources of expertise. First, one must observe common features of the documents. The printed logos, number codes, and the fold of a map can give clues

to its producer. Second, one must interview experienced catalogers about their understanding of documents and the process of map cataloging. Because the rules don't clearly identify the bodies responsible for a map, that information was gathered from catalogers and developed into a definition. Additionally, Ercegovac inquired about the experts' approach to cataloging and the cataloging process.

One area of cataloging that novice catalogers must master is knowledge of the sequence of appropriate cataloging rules to use in particular cases. Each of the expert systems guided the cataloger through a sequence of questions, suggesting the appropriate rule but requiring the cataloger to provide the necessary bibliographic data. This cooperation resulted in the creation of a partial or complete bibliographic record. Davies (1987) later suggested that cataloging systems interfaces could provide assistance with templates by interpreting some data across different fields of the record. Interfaces could also assist by prompting catalogers for required data following an established pattern of rule sequences.

It appears obvious from these research efforts that traditional cataloging is successful because it incorporates printed standards, national and local policies (written and unwritten), and the experience and knowledge of the cataloger. It is very likely that in order for an expert system to be successful it must incorporate all these elements. These also are areas that novice catalogers need to master over time through acquisition of experience. Library and information science programs can usually provide some instruction on rules and document characteristics, but rarely are able to provide the development of an experience base at a level sufficient for quality cataloging.

It may come as a surprise that the development of expert systems essentially ends with these described prototype systems. The basic unit of an expert system for cataloging still has not been resolved—that is, the structure and organization of the cataloging rules appropriate for computer use. Without that fundamental unit in place, the tedious and

expensive process of cataloging knowledge acquisition for developing the knowledge base seems to lead to no usable product.

Although Ercegovac (1990) demonstrated how an expert system could overcome the difficulties of the cataloging rules and incorporate appropriate expert knowledge, no one has reported that they have resumed the development of expert cataloging systems. It appears that catalogers and library managers both perceive other avenues as easier and more reasonable ways to reduce the human effort and thus the cost of cataloging. Efforts such as outsourcing of cataloging and the development of cataloging workstations seem to have taken the place of a fully matured expert system.

Recently, researchers and developers have focused on the improvement of the cataloger's tools, e.g. cataloger's workstation (Brisson 1995a; 1995b). In this approach, advances in computer technology—from dedicated terminals to personal computers—are predominantly used to benefit human work patterns. At first these workstations were only able to handle one task at a time, but eventually were improved to allow the use of multiple applications simultaneously. The major advantage of these workstations is that catalogers can access electronic versions of cataloging resources. Although such workstations are very useful and bring immediate benefits to the catalogers, they do not reduce the intellectual effort in the cataloging process, which continues to be the catalogers' responsibility.

CATALOGING RULES

While some researchers tried to build prototype expert systems and found it difficult to implement cataloging rules in these systems, others were interested in the function and structure of the rules. This is a narrower yet vital part of a comprehensive cataloging expert system development because it is an essential part of a knowledge base. Different chapters of the cataloging rules have been studied from a variety of perspectives, with the common purpose of enhancing their understanding and suggesting how the rules could be improved in further editions.

Svenonius and Molto (1990), Davies (1992), and Weibel (1992) prepared extensive reviews of automated descriptive cataloging and related research. They point to Sandberg-Fox (1972) as the first to work on this topic. Sandberg-Fox (1972) compared the *Anglo-American Cataloguing Rules* (1967) (*AACR*) to the older *A.L.A. Cataloging Rules for Author and Title Entries* (1949) for the computerized assignment of access points. She found that both rules posed problems for implementation in a computer environment. Given the problems that the prototype expert systems encountered and the comments of researchers on their difficulties with implementing *AACR2,* Sandberg-Fox's result suggests that these problems were already present in previous editions of the cataloging rules.

Jeng (1991a) analyzed chapter 1 (General Rules) of the *Anglo-American Cataloguing Rules*, 2d ed., 1988 revision (*AACR2R*), to identify the definition and functions of a rule. For the use of rules in cataloging monographs she created condition/action pairs for rules from chapters 1 and 2 (rule 1.0A1, Sources of Information, to rule 1.1G4, Items Lacking a Collective Title, and rule 2.0A1, Scope, to rule 2.1G2, Relationship Between Statement of Responsibility and Works in an Item that Lacks a Collective Title) into 77 condition/action pairs. This process would be necessary to implement cataloging rules in the knowledge base of an expert system.

Further, in a sample of 50 title pages, Jeng identified which rules were applicable and how frequently they would have been used. She determined only 13 of the 77 condition/action pairs were applied to every title page in the sample, while more than half of the condition/action pairs were never applied. For the Rule 1.0C1, which was divided into 25 condition/action pairs, only 3 pairs were applicable to all the title pages in the study. Although the sample of title pages was small, the results suggest that the rules may not be equally applicable, with some being general and some very specific. We feel that for a much larger sample of documents in a limited domain it could be anticipated

that not all the rules would need to be implemented, thus making the knowledge base of a system somewhat more manageable. While the findings of some researchers agree with Jeng's, others have found the contrary—in some cases more rules would be necessary because of the specialized role of the rarely used rules.

Meador and Wittig (1991) studied how frequently the rules used for choosing access points in *AACR2* (chapter 21), were applied to books in chemistry and economics. In their experiment, only 12 of 143 rules in that chapter were used on a sample of 30 items from each discipline. The authors concluded that, although the core rules were different for each discipline, the same expert system could work for both because: (1) a small number of rules was used in total, and (2) rules used for economics represented a subset of rules used for chemistry. This study, although on a small sample, indicates that some of the rules are general and could be used across different disciplines.

Weiss (1994) reported on a project conducted at the National Library of Medicine (NLM) intended to create an expert system to assist catalogers with personal name authority work. Implementing only *AACR2R* chapter 22 (Headings for Persons) proved insufficient. Weiss found that to construct personal name headings correctly, the expert knowledge of experienced catalogers was required for the correct translation of personal names appearing in documents into personal name headings. He concluded that acquiring that expert knowledge is a time-consuming and expensive task. One might imply from Weiss that the foundation for correct decision making in cataloging requires not only a sound knowledge of cataloging rules but also an extensive experience in the application of these rules.

The results of Jeng's (1991a) and Meador and Wittig's (1991) studies point in the same direction. Because the entire set of cataloging rules is difficult to implement in a knowledge base, using only a subset of the rules may be a better approach. If the rules used more frequently are also the most important rules and ex-

clusion of other rules does not lead to problems, perhaps only that smaller set of important rules needs to be implemented in an expert system. Weiss's (1994) findings do not appear to support this conclusion, however. These contrasting results show that using a subset of rules in a cataloging expert system needs to be studied more fully. One way to address this problem might be to develop a knowledge base using the subset of important rules to identify problems that arise. Judging from the results of prototype expert systems, it is clear that there is some subjective interpretation of cataloging rules because of missing information or insufficient guidance in the rules. Moreover, in the studies cited, the issue of researcher bias is not addressed, and its effect is undetermined.

Clearly, there are some difficulties with the logical structure of cataloging rules. Hjerppe and Olander (1989), Ercegovac (1990), Jeng (1991a), and Weiss (1994) all commented on problems with inconsistent, contradictory, insufficient, and redundant information contained in the rules. Fidel and Crandall (1988) examined *AACR2* from a generalized database approach, using the entity–relationship model. They categorized rules into six types: (1) content; (2) establishing entities, relationships, or attributes; (3) authorized sources; (4) domain; (5) format; and (6) access points. They demonstrated that the current arrangement of *AACR2* scatters rules of the same category in different parts of the text and also mixed different categories into the same rule. They believed that this adds to the confusion of what are and what are not conceptual rules. They suggested that systematic investigation of the structure and function of the rules would be possible with the entity–relationship model.

Taniguchi (1996) built a prototype expert system for analyzing *AACR2R*. This system was designed to analyze the internal structure of the rules and the relationships among them. The first phase of the analysis in this system, a manual one, was a transformation of cataloging rules from natural language into a more formal,

structured language that a computer system could process. The next phase, a computerized approach, involved the normalization of rules, i.e., converting a rule from its still relatively natural, but already structured form, into an entirely mathematical, logical form (condition/action pairs). The third phase of Taniguchi's system involved testing of a rule with rule templates. These rule templates were skeletal schemes that provided the very basic structures of rules. They consisted of the characteristics, i.e., the variable information of condition/action pairs, necessary for logical functioning and verification of information for all the necessary components. Three types of templates were developed: (1) for descriptive rules, (2) for definition rules, and (3) for organization rules.

It seems reasonable to expect that Taniguchi's first phase was rather subjective, given that rules can be interpreted differently by different people. As a consequence, the second and subsequent phases could then have different results as well. Yet, these differences might illustrate the missing information in the rules, which might lead to improvements in rule structure.

There are different views on the categorization of cataloging rules. While Taniguchi's three categories of rules were description, definition, and organization, Jeng (1991a) listed five categories: definition, description, organization, identification of the source of information, and transcription. Fidel and Crandall (1988) specified six categories: content; establishing entities, relationships or attributes; authorized sources; domain; format; and access points. Taniguchi (1996) expressed a need for consensus on this issue as a condition for success of any analysis and suggested that the proposed system offered an appropriate method that could help with developing less ambiguous and more consistent rules.

Although one might argue that human reasoning is different from computer processing, the computerization of a process offers a way for people to organize their thoughts and produce a more systematic assessment of their own work. The system may not be perfect or be able to perform analysis independent of human experts, but it offers a way to gain a deeper and more systematic understanding of human experts' own reasoning, including their strengths and limitations.

In these studies of *AACR2*, researchers have indicated how, and why, it is difficult to "teach" a computer how to catalog. This should come as no surprise to instructors of cataloging, who encounter the need to convey document patterns and the intricacy of what is stated in the rules, and (equally as important) what is not stated in the rules to their students. Complicating all these instructional challenges are the different life experiences that the students bring to cataloging process. Much of the "artistry" of teaching cataloging seems to lie in teaching how to search for clues that provide a basis for the creation of an appropriate bibliographic description. Such clues often are found in the combination and context of the bibliographic elements in a document and the cataloging rules and their interpretations appropriate to the situation. The question arises of how this artistry can be formalized in an expert system. For the majority of "typical" documents, expert systems using knowledge base likely could work independently, while for more complicated documents, these systems would need to interact with human experts.

CHOICE AND FORM OF ACCESS POINTS

Access points are necessary because they offer a way to access catalog records systematically and to arrange them for display. Thus, both their selection and their formatting are essential components of existing catalog codes. Access points have been a special area of inquiry from two perspectives. First, they have formed a subset of studies of cataloging rules. Second, access points have been studied with a view toward better control of what would be one part of a comprehensive expert cataloging system. Such a system would incorporate automatic reading of parts of documents to produce data for the creation of access points.

Svenonius, Baughman, and Molto (1986) reported studies of automatic recognition of names and titles on the title page. Their purpose was to identify the possibilities of simplifying the determination of name access points. Their work made the assumption that optical character recognition (OCR) was advanced enough to identify text on title pages accurately. They investigated responsibility statements in a sample of 400 monographs from an academic library. All the documents were in English. Author names most frequently appeared on title pages, preliminaries, and tables of contents (38% of the 1,310 author names in the sample). Tables of contents contained the largest share of all names (66% of the total 2,536 names). However, the researchers further found that many of the personal and corporate names that appeared on title pages, preliminaries and tables of contents were neither authors nor people or bodies whose function was related to authorship.

Svenonius and Molto (1990) investigated the possibility of automatic recognition of type of responsibility from machine readable title pages of monographs. They also examined whether title page data were adequate for identification of name access points. The success rate was measured by the number of correctly identified access points (all names), and in terms of precision (the number of correctly identified name access points automatically identified by their algorithm compared to the records created by LC or NLM). Their system correctly identified 90% of the names that appeared on the title pages. Of the correctly identified names, 95% were also chosen as access points in LC and NLM cataloging records. Svenonius and Molto estimated that approximately 88% of the access points selected by LC and NLM could be automatically derived from title page data.

Molto and Svenonius (1991) also investigated automatic recognition of names appearing on samples of title pages from monographs in an academic library and a public library. Their method correctly identified 86% of corporate names and 85% of personal names. Only 65% of the names identified as personal names in the sample were also chosen by LC as personal name access points. Approximately 14% of corporate names and 15% of personal names were either not recognized or were incorrectly recognized. Molto and Svenonius suggested that their rules for establishing name access points and the graphic presentation in the documents contributed to the incorrect or failed recognition.

The state of OCR technology at the time of these tests could not reliably interpret text for cataloging purposes. Problems arose due to the function of names and the artistic design employed in the documents. Systems have difficulties identifying the function of some authors. While illustrators and translators are usually clearly credited, the roles of editors, compilers, and other contributors can be harder to determine. In some cases these functions had to be inferred. Electronic files used to produce printed texts might be more amenable for computer interpretation. Perhaps cataloging rules need to be more flexible to allow the use of this new technology.

Excluding these obstacles, two issues emerged from the research on the automatic recognition of access points. First, it was not possible to anticipate all the possible personal and corporate names to be included in name authority files. These files are a necessary component of expert systems. Words scanned in a document need to be compared to the names in the authority file by an expert system, in order to be accepted as a name, verified against other names, and accepted as an access point. Second, even if it were possible to identify all the appropriate names correctly and their current form automatically, the ability to distinguish different functions of authorship remains a challenge. A mechanism is needed to distinguish among writers, translators, editors, and numerous other people who appear on the document and are responsible in different ways and degrees for its creation.

Both of these are major impediments to building expert cataloging systems as well as a challenge for the everyday work of catalogers. The identification

and interpretation of bibliographic data in a document is the underlying skill. The "fuzzy" use of language, which poses such difficulties for computer systems, is less problematic for people. Catalogers must judge the importance of names and titles appearing on the document and decide which should be chosen as access points and which should be disregarded. This is where experience plays a crucial role. Here the cataloger finds the organization and layout of the document helpful, whereas an expert cataloging system finds them problematic.

DOCUMENT CHARACTERISTICS

There are linguistic characteristics most people use without much thought, but an expert system must be programmed to recognize them. Many catalogers know that terms like "Associate Professor of Clinical Medicine" appearing under a name is not a title of the book but rather a person's affiliation. People familiar with American culture know that "Washington" can refer to a city, a U.S. state, or a personal or corporate name. These people can easily distinguish through context which "Washington" is meant. The challenge to the expert systems is to interpret that context properly.

Studies of visual and linguistic characteristics of documents are extremely important in building expert systems. Thirty years ago, Kilgour (1969) suggested the automatic extraction of bibliographic data from documents. Researchers of this operation still encounter difficulties of technological limitations similar to those researching prototype expert systems. Although the optical reading of catalog cards for their conversion into database records has had good results, OCR equipment is not yet sufficiently developed to create catalog records from original documents. Two characteristics of documents add to the task's difficulty. One is the indeterminate nature of language, which poses relatively little problem to speakers of the language, but is difficult to formalize for a computer system. The other is the graphic design of documents. Although design can be diverse, catalogers expect that certain types of data will usually appear in certain places. Understanding the visual and linguistic structure of documents may, therefore, be necessary in the development of cataloging rules, if they are to be used effectively in an expert system. Rules derived by generalizing characteristics observed in documents can ensure more homogeneous interpretation of data in documents.

Jeng (1986) suggested that title pages tend to have structure, which could be used for automated data recognition. She defined a "block" (one or more words separated from other blocks by blank vertical space) as a basic unit of analysis. She found that blocks have two characteristics: physical appearance and content. Physical appearance is described with type face and size. Content is the meaning of the text. Jeng (1987, 1988) later presented a study of linguistic characteristics on a sample of 203 title pages in which she identified the occurrence of words, phrases, and punctuation marks. In her sample, an average of 23 words appeared on a title page. The most frequent types of words used were proper names (36%) and common nouns (26%). Other syntactic elements that appeared were prepositions (11%), initials (initials in names or abbreviations) (7%), articles (6%), adjectives (5%), conjunctions (4%), and verbs (4%). She examined the occurrence of these elements in different blocks on title pages (e.g., titles, other title information, statement of responsibility, publisher).

Jeng (1991b) continued her research on title pages, concentrating on their visual characteristics. On the basis of frequency of bibliographic elements, she proposed a three-level prototype title page, with an appropriate location and sequence of bibliographic elements on the page. The first level contained the title proper, author and publisher information. The second level added other title information and place of publication, with the third level adding author affiliation and year of publication. Based on her findings, Jeng concluded that because cataloging rules were not derived from a systematic knowledge of title pages, the interpretation of title page information is

more dependent upon a cataloger's interpretation and experience. She suggested that her research into the composition of title pages has implications for both the future development of cataloging rules and expert cataloging systems.

Weibel, Oskins, and Vizine-Goetz (1989) also investigated the possibility of automated recognition of data on the title page and tested Jeng's findings (Jeng 1986, 1988). They simulated OCR techniques using locally produced electronic versions of title pages. This simulated OCR technique was necessary because contemporary OCR devices were not able to perform at the level that the researchers required. Their goal was to identify automatically the following data from the title page: title, other title information, statement of responsibility, edition statement, publisher, place, and date of publication. Their basic unit of data was represented by a "token," which was a space-delimited character string such as name or title (what Jeng called a "block"). Tokens had attributes representing position, size, typeface, and the case class of each character string. Case class reported upper- and lower-case characters, numbers, punctuation, symbols, and blank spaces. After the system identified the data, a confidence level was calculated, and the data were assigned to appropriate fields.

In their sample, 45 records of English-language monographs were used as training data to adjust and improve upon the theoretic assumptions for the system. Forty-six similar records were used to test the system. Their results showed that identification of publisher, place and date of publication was far better than the identification of other bibliographic elements. For these bibliographic data they achieved 80% correct identifications. This success was attributable to being able to compare publication related data to external files. For example, publisher names could be compared to an authority file of publishers. They found that it was much more difficult to identify titles, other title information, and statements of responsibility correctly.

It seems likely that an expert system for optical recognition of bibliographic data will require human verification of the results. Catalogers will continue to need to master the identification and interpretation of bibliographic data from documents. Places where this information is likely to appear are frequently, but not always, prescribed in the rules. However, often document designers or publishers don't follow standard practices in the placement of data.

Another suggestion derived from the development of prototype expert systems is the study of catalogers' perception and understanding of document layout. Ercegovac (1990) addressed this in her attempt to identify the characteristics of documents for identification of access points the producer of the map. The use of document characteristics has been found to be very useful by catalogers. That use, however, has not been incorporated in the cataloging rules. How do students and novice catalogers learn to use these clues? Most likely, experienced instructors and experienced catalogers impart this information on using the clues as well as the use of the cataloger's general knowledge to create a complete catalog record.

CATALOGING EXPERTISE

These expert cataloging system research efforts illustrate some of the difficulties of capturing the skill and knowledge required to catalog. Cataloging has been found to be so complex that it needs to be studied as a range of different subtasks. The details of these varied tasks need to be put together as pebbles in the mosaic of cataloging expertise. A similar conclusion was made by Jeng (1992a). She asserted that the understanding of the role of cataloging expertise (the general and personal knowledge of expert catalogers) needs to be explored. In addition, a larger knowledge base (with transformed rules) and better user interfaces must be developed.

Ercegovac (1990) was the first to discuss the incorporation of personal knowledge from expert catalogers and specific document characteristics into an expert

system for descriptive cataloging. For Mapper, she interviewed experts to devise additional rules for the knowledge base to supplement situations in cases where *AACR2R* provided insufficient detail. She studied the characteristics of printed maps and interviewed experts about common characteristics of these maps in order to create a knowledge base.

In a later paper, Ercegovac (1992) discussed observation techniques she used in the development of Mapper. Interviews and surveys were used in addition to the careful study of the characteristics of maps. The first two approaches were designed to elicit details about the data necessary for map cataloging. For instance, questions were asked about definitions of authorship and publishing data, and a ranking of specific responsibility functions. This enabled her to understand this narrow field in detail and successfully imitate expert catalogers in a computerized system.

Jeng (1992b, 184) defined expertise as "the high degree of skills, dexterity, or knowledge of a specific subject area." In that study, in which she studied cataloging expertise and the transfer of that expertise, she interviewed the head of the cataloging department and three professional catalogers at National Agricultural Library (NAL). Documentation related to the cataloging process and job descriptions was also gathered. Jeng concluded that people doing the lowest ranked tasks have the least expertise, that transferring expertise is well formalized in NAL, and that informal cooperation among catalogers was also common. Jeng emphasized that learning from questions that arose during regular workflow is the most common method for transferring expertise. This method would likely prove to be an obstacle in building expert systems.

Ercegovac (1992) found that a focused interview with a detailed schedule was a better source of information about the cataloging process than verbal reports and associated protocol analysis. The focused interview is less biased by the weaknesses of human memory. In her work, exact questions could be asked about authorship and the characteristics of spe-

cific maps. One wonders how difficult it would be to answer in an exact way the questions about the cognitive processing of how and why someone decided to follow a detail on a map for determining authorship when that detail is not mentioned in cataloging rules.

Jeng (1997) found verbal reports and protocol analysis for understanding cataloging expertise to be a successful and appropriate method for acquiring an understanding of cataloging expertise. In that project, she studied the kinds of knowledge and skills catalogers needed, what tasks were involved in cataloging, and what strategies catalogers used to solve specific problems. She used multiple methods: (1) a questionnaire, (2) a self-administered verbal report of cataloging sessions, (3) verbal protocols of sessions in which catalogers were asked to "think aloud" during the process and the researcher observed the process, and (4) verbal protocols of training sessions in which the researcher was taught by a senior cataloger. Jeng did not present the result of this research, but she illustrated her research methods with examples of the work process of two catalogers. Workflow was found to be influenced by the workscreen. One cataloger's problem-solving strategy was to leave a difficult decision until the end of the task, while the other solved the problem when it occurred. Little explanation was provided in her report about the cognitive process of decision making. The two catalogers used standard cataloging knowledge for their work and problem solving techniques. The report of the catalogers' work did not mention the use of knowledge beyond cataloging rules.

Ercegovac (1992) and Jeng (1997) suggest that knowledge of cataloging rules and procedures—as well as knowledge of cataloging systems—are necessary. Some specific knowledge, which is not written in cataloging rules or policies but is part of an oral tradition among catalogers in the same environment, is also necessary. This knowledge needs to be supplemented with general knowledge. The unwritten rules and general knowledge that assist in problem solving needs further research.

Different strategies of employing this knowledge in solving cataloging problems are used by different catalogers.

The different aspects of cataloging expertise likely will require different research approaches. While self-administered reports might be useful in studying the mechanics of cataloging procedures, observations and teaching sessions might be more useful for obtaining reasons for particular steps in the procedures. While questionnaires might be useful for learning about certain concepts in the cataloging process, they are useless if catalogers are not aware of those concepts. Different documents are examined and described in bibliographic records containing different data. In the same way, different kinds of processes and categories of knowledge need to be examined and described differently. Catalogers need to employ their long tradition of categorization of data, enriched with related disciplines, on the cataloging process to develop mechanisms for systematic representation of their expert knowledge.

SUMMARIZED IMPLICATIONS

In this paper, we have reviewed 25 years of work and thinking in the area of the use of computers to create cataloging records. Particular attention was paid to the types of knowledge necessary for these systems and to the problems with contents of the knowledge base. Have we found any clues about the missing pebbles for the mosaic of the cataloging expertise? We believe we have. Here are some clues found in the literature we reviewed:

- Comprehensive cataloging systems (i.e., those that cover all cataloging operations and all document formats) are too *complex* to develop either in research projects or for work environment.
- Reporting successes and failures in the *knowledge acquisition* for these prototype systems has been as important for the development of the discipline as successes and failures in actual development and implementation of the expert systems themselves.

- Some cataloging tasks seem to be more amenable to automation than others. Such tasks seem to have *routine procedures*. Standardization of such procedures could aid in the development of expert systems for the automatic generation of the content of the cataloging records. Such standardization would also be beneficial for the education of the students and new catalogers.
- *AACR2* has been found to be problematic primarily because: (1) of the inconsistent logical structure of the rules; and (2) the information required to make cataloging decisions is missing in the rules. Difficulties arising from those problems are troublesome not only for the expert systems, but also for human experts.
- *Characteristics of documents* are important in developing expert systems that could automatically derive bibliographic data from the document. These same characteristics are useful to human catalogers. Researchers have found some patterns in the characteristics that would help both.
- Systems and people are fairly successful in recognizing personal names. The difficulty lies in deciding about the *role of a particular name* in the generation and production of the document.

We wish to add some more reflection on each of the six clues for the missing pebbles by returning to each individually.

Initial optimism that theoretical research in building expert systems would help with the complex task of descriptive cataloging dissolved quickly under the disappointing performance of prototype systems. These complex and comprehensive systems were generally not successful. Researchers then focused on smaller, narrower areas, in which more understanding and less variability existed. This resulted in some success.

Few researchers who built prototype expert systems have reported on actually eliciting knowledge from human experts, but they have noted that rules can be appropriately interpreted only by expert catalogers. Information contained in

cataloging rules has been shown to be insufficient by themselves for making cataloging decisions. Rather, catalogers base many decisions on information acquired thorough experience. Because this information is subjective and varies at different levels of expertise, cataloging results also vary. This subjectivity may result in inconsistent catalog records and calls for greater standardization.

One approach toward improving consistency could be the standardization of cataloging procedures for particular types of documents. It seems reasonable to believe that some types of documents consistently require the use of certain procedures. For example, some items require the use of certain sequences of rules to be consistently applied. These procedures may involve use of some external knowledge. If these cases could have the external knowledge captured, they might be amenable to formalization in a knowledge base. Several benefits could result. First, more consistent decisions and more homogeneous catalog records could be provided. Second, new catalogers could study and understand in detail all the steps and details in the process. A third benefit would be the storage and documentation of the expertise.

Problems using *AACR2* as the source of knowledge have lead to the investigation of cataloging rules. The structure of rules, their relationships, and the function and use of individual rules and chapters have been studied by a number of researchers. Each has concluded that cataloging rules were not particularly amenable for immediate computerization but could, and should, be made more adaptable for implementation in a knowledge base. This becomes particularly important as the use of computers to enhance productivity and accessibility increases and demands to reduce processing costs rise.

The need to implement cataloging rules in a knowledge base is not the only reason to improve the rules. It is clear from prototype system development that there is some subjective interpretation of cataloging rules because of missing information. More systematic and consistent rules could result in easier work for cata-

logers, who could spend less effort on inferring the missing or contradictory information. More consistent catalog records could result if there are fewer individual decisions. These are also areas that would require more emphasis in the education of new catalogers. Instruction in the use of cataloging rules should be complemented with an examination of inconsistencies and gaps in the rules. Strategies commonly used by expert catalogers could be used to illustrate problems and help novices understand the kind of external knowledge they need to employ in certain situations. Library and information science teachers might then better prepare students to catalog in real environments.

Some researchers have studied the characteristics of documents and experimented with automatic recognition of data from the document. In their work, they all made the assumption that future OCR technology will be able to correctly read title pages regardless of graphic design variations. Some also suggested that the electronic version of the printed title page should be accepted as the primary source of information. Results in this area have been quite encouraging. More than half of personal names, corporate names, and title access points could be automatically assigned in the English-language monographs that were tested.

Two major problems exist with the accurate automatic identification of names: (1) names need to be included in the knowledge base and (2) types of responsibilities need to be distinguished. These two areas also need to be mastered by beginning catalogers. While knowledge of names mostly comes from the cataloger's personal knowledge, types of responsibilities are learned in cataloging courses and through experience. Knowledge of which names are associated with the creation of a document and where they appear in a document is one of the basic skills catalogers need to master. It is one closely associated with the knowledge of visual characteristics of documents.

Understanding visual document characteristics might help with cataloging when catalogers are not familiar with the culture and language of the document, provided

that these visual clues are not culturally based. Identification of these characteristics could help catalogers find and interpret the necessary bibliographic data.

The research we have reviewed here has covered cataloging expertise and the cataloging process. Understanding the decision process is as important as understanding sources, rules, or building cataloger-friendly interfaces on workstations. Unfortunately, these researchers seem to have contributed little to the real working environments and procedures of catalogers. While significant progress has been made in the development of cataloging workstations, which make multiple cataloging tools available for simultaneous consultation on the computer screens, they continue to leave the challenging and difficult intellectual work entirely to the cataloger.

CONCLUSIONS

From the studies presented one can conclude that catalogers invest significant intellectual effort in the interpretation of cataloging rules and in the interpretation of document characteristics. They learn from their own experience and from experienced colleagues how to master these two tasks. The interpretation of the rules likely depends both on the general knowledge and the expertise of the cataloger. Expert knowledge comes from formal education and interaction with colleagues. A culture of sharing expertise among colleagues has great value here. The general knowledge of catalogers is the knowledge that is not directly connected to their profession. It is gained through education and social interactions in the community. Although these aspects were not studied, a systematic approach to the individual interpretations of the rules might result in easier work and in more consistent decisions. Consistency and clarity of catalog records is, after all, the main goal of all catalogers, as they try to provide the most relevant and accurate information about sources to the users.

What must a student interested in entering the cataloging specialty learn to begin the journey to becoming an expert cataloger? The research in prototype expert systems has demonstrated that knowledge of the cataloging rules is not sufficient. It was shown that there are several missing pebbles from the mosaic of cataloging expertise. The ability to interpret cataloging rules is one important area. An example of this form of interpretation is represented to a limited degree by *Library of Congress Rule Interpretations (LCRI)*. In addition, a high degree of experience and common sense has to be employed to make the necessary judgments called for by the rules.

Another area to be addressed is the sequence of the rules in the system. Certainly there are many documents that follow the same rule pattern. Some of these patterns are taught in cataloging classes. Others, less common, are only mastered with experience, while a few always require innovative approaches to rule application.

Cataloging workstations currently utilize relatively rudimentary knowledge technology. For documents, which follow certain patterns, these workstations could assist by providing templates, such as inferring the publisher from the ISBN, as suggested by Davies (1987). Molto and Svenonius (1998) proposed an interface for the online version of *AACR2R* which would make the use of this basic tool easier. Perhaps the electronic version of *AACR2R* will begin realize this hope. Certainly linking the *LCRI* and *AACR2R* within these workstations would be a useful tool for catalogers. Again, the Cataloger's Desktop from LC offer some hope for optimism.

Access points are essential in the efficient retrieval of documents. The identification of names, titles, and functions as well as ranking their importance is necessary. Experiments have demonstrated that many access points and their proper headings for personal names could be automatically generated. These decision criteria do not come easily. Instead they are the result of multiple encounters with cataloging rules. Some criteria are suggested in *AACR2R* and *LCRI*, while others are not.

Essentially, students aspiring to the cataloging specialty cannot reasonably be expected to be taught all the solutions to

problems they may encounter. Rather, their instruction, in addition to providing the basic building blocks of cataloging (i.e., rules), must concentrate on developing problem solving skills which will allow them to enhance their cataloging experience base.

Catalogers have demonstrated that they can, and want to, use computers in the best ways possible. At the moment, however, some aspects of their work are more amenable than others to the employment of this tool. This does not mean, however, that we should not continue to try to understand and formalize the cataloging process further. Knowledge acquisition methods offer a systematic inquiry into the cataloging expertise. A thorough understanding of the mental processes involved in cataloging offers not only a tool for developing expert systems, but also for rationalization of our own work. Systematic processes are easier to learn and the transition from novice to an expert is therefore smoother. Such approach should appeal to the catalogers, who are traditionally regarded as systematic organizers of materials and information, and to all those who are concerned with the quality of library catalogs.

Considering the research and developments to date, it is apparent that the expert knowledge base continues—and will likely continue—to reside with the cataloger. The new tools that might reduce the cataloger's other burdens are not likely to reduce the need for human expert intervention in the cataloging process. This expert knowledge continues to be the domain of the cataloger, who provides the missing pebbles for the mosaic of cataloging expertise.

WORKS CITED

A.L.A. *cataloging rules for author and title entries*. 1949. Prepared by the Division of Cataloging and Classification of the American Library Association, 2d ed., ed. Clara Beetle. Chicago: American Library Association.

Anglo-American cataloging rules. 1967. Prepared by the American Library Association, the Library of Congress, and the Library Association, North-American text. Chicago: American Library Association.

Anglo-American cataloguing rules. 1978. Prepared by the American Library Association, the British Library, the Canadian Committee on Cataloguing, the Library Association, and the Library of Congress, 2d ed., 1978, ed. Michael Gorman and Paul. W. Winkler. Chicago: American Library Association.

Anglo-American cataloguing rules. 1988. Prepared under the direction of the Joint Steering Committee for Revision of AACR, a committee of the American Library Association, the British Library, the Canadian Committee on Cataloguing, the Library Association, and the Library of Congress, 2d ed., 1988 revision, ed. Michael Gorman and Paul W. Winkler. Chicago: American Library Association.

Anglo-American cataloguing rules. 1998. Prepared under the direction of the Joint Steering Committee for Revision of AACR, a committee of the American Library Association, the British Library, the Canadian Committee on Cataloguing, the Library Association, and the Library of Congress, 2d ed., 1998 revision, Chicago: American Library Association.

Brisson, Roger. 1995a. The cataloger's workstation and the continuing transformation of cataloging. Part I. *Cataloging & classification quarterly* 20: 3–23.

———. 1995b. The cataloger's workstation and the continuing transformation of cataloging. Part II. *Cataloging & classification quarterly* 20: 89–104.

Davies, Roy. 1987. Outlines of the emerging paradigm in cataloging. *Information processing & management* 23: 89–98.

———. 1992. Expert systems and cataloguing. In *The application of expert systems in libraries and information centres*, edited by Anne Morris. London: Bowker-Saur.

Davies, Roy, and Brian James. 1984. Towards and expert system for cataloging: Some experiments based on AACR2. *Program* 18: 283–97.

Gibb, F. and C. Sharif. 1988. CATALYST: An expert assistant for cataloging. *Program* 22: 62–71.

Ercegovac, Zorana. 1990. Research on knowledge-based descriptive cataloging of cartographic publications: An experimental advice-giving system—Mapper. Ph. D. diss., University of California, Los Angeles.

———. 1992. A multiple-observation approach in knowledge acquisition for expert systems: A case study. *Journal of the American Society for Information Science* 43: 506–17.

Ercegovac, Zorana, and Harold Borko. 1992.

Design and implementation of an experimental cataloging advisor—Mapper. *Information processing & management* 28: 241–57.

Fidel, Raya and Michael Crandall. 1988. The AACR2 as a design schema for bibliographic databases. *Library quarterly* 58: 123–42.

Hjerppe, Roland, and Birgitta Olander. 1989. Cataloging and expert systems: AACR2 as knowledge base. *Journal of the American Society for Information Science* 40: 27–44.

Jeng, Ling Hwey. 1986. An expert system for determining title proper in descriptive cataloging: a conceptual model. *Cataloging & classification quarterly* 7: 55–70.

———. 1987. The title page as the source of information for bibliographic description: An analysis of its visual and linguistic characteristics. Ph.D. diss., University of Texas, Austin.

———. 1988. The language of a title page. In *Proceedings of the 51st ASIS annual meeting, Atlanta, Georgia, Oct. 1988*, 31–35. Medford, NJ: Learned Information.

———. 1991a. The structure of a knowledge base for cataloging rules. *Information processing & management* 27: 97–110.

———. 1991b. Knowledge representation of the visual image of a title page. *Journal of the American Society for Information Science* 42: 99–109.

———. 1992a. The conceptual obstacles to building the knowledge base for cataloging expert systems. In *Information technology: IT's for everyone! Proceedings of the LITA 3rd national conference, Denver*, edited by T. W. Leonhardt, 67–69. Chicago: ALA.

———. 1992b. Networking and transformation of cataloging expertise. In *Networking, telecommunications, and the networked information revolution: ASIS 1992 mid-year meeting, Albuquerque*, 182–92. Silver Spring, Maryland: American Society for Information Science.

———. 1997. Using verbal reports to understand cataloging expertise: Two cases. *Library resources & technical services* 40: 343–58.

Kilgour, Frederick G. 1969. Computerization: The advent of humanization in the college library. *Library trends* 18: 29–36.

Library of Congress rule interpretations. 1990–. Washington, D.C., Cataloging Distribution Service, Library of Congress.

Meador, Roy III, and Glenn R. Wittig. 1991. AACR2 rules used in assigning access points for books in two subjects: Implications for automatic cataloging expert systems. *Library resources & technical services* 35: 135–40.

Molto, Mavis, and Elaine Svenonius. 1991. Automatic recognition of title page names. *Information processing & management* 27: 83–95.

———. 1998. An electronic interface to AACR2. *Cataloging & classification quarterly* 26: 3–24.

Morris, Anne. 1992. Overview of expert systems. In *The application of expert systems in libraries and information centres*, edited by A. Morris, 1–33. London: Bowker-Saur.

Sandberg-Fox, Ann Martha. 1972. The amenability of a cataloging process to simulation by automatic techniques. Ph. D. diss., University of Illinois, Urbana–Champaign.

Svenonius, Elaine, Betty Baughman, and Mavis Molto. 1986. Title page sanctity? The distribution of access points in a sample of English language monographs. *Cataloging & classification quarterly* 6: 3–21.

Svenonius, Elaine, and Mavis Molto. 1990. Automatic derivation of name access points in cataloging. *Journal of the American Society for Information Science* 41: 254–63.

Taniguchi, Shoichi. 1996. A system for analyzing cataloging rules: A feasibility study. *Journal of the American Society for Information Science* 47: 338–56.

Weibel, Stuart. 1992. Automated cataloging: Implications for libraries and patrons. In *Artificial intelligence and expert systems: Will they change the library? Proceedings of the 1990 Clinic on Library Applications of Data Processing*, edited by F.W. Lancaster and L.C. Smith, 67–80. Urbana, Ill.: Graduate School of Library and Information Science, University of Illinois at Urbana–Champaign.

Weibel, Stuart, Michael Oskins, and Diane Vizine-Goetz. 1989. Automated title page cataloging: A feasibility study. *Information processing & management* 25: 187–203.

Weiss, Paul J. 1994. The expert cataloging assistant project at the National Library of Medicine. *Information technology and libraries* 13: 267–71.