Application of Expert Agents/Assistants in Library and Information Systems

James G Williams and Ken Sochats*

Abstract

This article presents a brief overview of expert agents and assistants, and then discusses activities in several areas of library and information centres where expert systems have been developed and where intelligent agents/assistants could be of practical use in such activities. A discussion of what an intelligent agent or assistant might perform in these areas is outlined.

1. INTRODUCTION

Although expert systems can and have been applied to many areas within library and information science, the most popular areas have been online retrieval, reference and referral services. Other areas which have received a relatively significant amount of attention are cataloguing and classification, indexing and abstracting, collection development, acquisitions, and management information.

McDonald & Weckert[1] provide a set of articles from a conference and workshop held at Charles Sturt University in Riverina, Australia, which discuss expert systems in many of these areas. Cavanagh[2] and Durkin[3] also provide excellent overviews of the application of expert systems in library and information centers. Cavanaugh's article covers selection/acquisitions, cataloguing, classification, indexing, reference, user models and interfaces for searching, post-search

* Department of Information Science and Telecommunications, University of Pittsburgh, USA.

processing and intelligent tutoring. He provides a brief description of prototypical systems that exist for these areas. Durkin has compiled an extensive catalogue of expert system applications in wide ranging areas.

Over the last several years, there has been considerable effort in developing intelligent agents and assistants which are experts in a particular domain such as office automation. These agents and assistants have been developed with varying levels of intelligence using expert system principles and techniques although they are not necessarily expert systems in the classical sense.

2. EXPERT SYSTEMS

Generally, expert systems are classified as a subset of the field of *Artificial Intelligence* (AI). Expert systems are systems that employ a codification of human expertise to solve problems. The goal of the expert system is to emulate the problem solving process of the expert(s) whose knowledge was used in the development of the system. Thus, the expert system's performance is highly dependent on the quality of the expert. Typically, expert

systems are narrowly limited in their domain of operation.

The structure of an expert system is shown in Fig. 1. An expert system is comprised of three
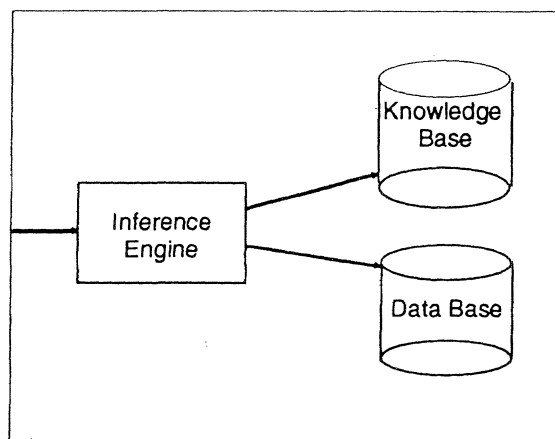

Figure 1. Expert System Structure

component areas. First is the knowledge base. The knowledge base contains the representation of the expert's knowledge in the problem domain. Next is the database that contains specific facts about the problem situation under study. Finally, the inference engine is a programme that applies the expertise of the knowledge base to the facts in the database to solve the problem at hand.

Knowledge base is the 'nerve centre' of the expert system. The knowledge base represents both the expertise and the reasoning strategies of the expert in the problem domain. Many different representational schemes have been used by expert systems researchers to structure the knowledge base. These range from network structures that interrelate the expert's domain knowledge to frame representations that represent knowledge as schematic data structures to rule based systems that codify the expert's knowledge as a set of rules or heuristics.

Two major issues arise during the development of the knowledge base. The first is the source of the knowledge base content. Knowledge bases can be derived from the expertise of single experts in single domains as in the Caduceus system for medical diagnosis (Pople, Myers, and Miller[4]). Many common

systems rely on the use of knowledge from several experts in the same domain that are merged to form a single knowledge base.

The Prospector system for mineral exploration (Duda, Gashnig, and Hart[5]) has several sub knowledge bases derived from experts with differing areas of expertise. The Prospector system applies each of these knowledge bases to a problem and evaluates the outcomes of these to arrive at a single final recommendation.

The second issue in knowledge base development is the extraction of the expert's knowledge. Many expert system developers rely on interview techniques to get their expert(s) to explain the techniques and information that they bring to bear in solving the domain problems. Ericsson and Simon[6] have introduced the formal methodology of protocol analysis to capture an expert's reasoning. In protocol analysis, the actions of an expert are completely recorded as the expert solves a series of problems. The expert is also encouraged to 'think out loud' or to talk through the steps he or she is taking during the solution of the problem. These recordings become a protocol which, when analysed, provide insight into the expert's problem solving skills.

Rulecons[7] is an expert system to aid in the formulation of knowledge bases. Rulecons assists the expert system developer or knowledge engineer during interviews of the expert(s). It is a system for creating domain independent, rule based knowledge bases. Rulecons helps the designer to maintain consistency and achieve completeness in the knowledge base.

3. INTELLIGENT AGENTS

An agent is a software module that monitors a specified set of events, messages and information flow for a user or group of users and performs useful functions that assist the user(s) in reducing his/her need to be notified, involved, interrupted or otherwise bothered to deal with all aspects of the activity, task or related sub-tasks. Agents that have been developed in various application areas range from simple extensions to existing software

such as a word processor or spreadsheet to complex work-flow procedures. Such agents assist in tasks such as scheduling meetings, filtering e-mail messages, searching for specific types of task related information or making decisions about routing forms for authorisation or performing anticipated additional work functions.

Two main features of intelligent agent software for assistance are:

⊕ it must be able to learn and

⊕ it must be able to negotiate.

Negotiation is required because software must be able to interact with other agents (which may include human being) in order to both learn and complete its tasks. It must be able to learn since it must adapt to the user for whom it is an assistant and the changing environment in which it exists. Such agents have been developed for calendar management, optimised use of the UNIX operating system and 'surfing' the Internet. The most popular application of intelligent agents has been in the office domain for handling information flow such as mail filters, news filters, and library browsers.

Intelligent agents must work autonomously in the background so that they can perform their work with as little user intervention as possible. This requires the agents to negotiate with other agents and reason about available resources and alternative solutions to task related problems in order to minimise user interaction.

Intelligent agents must be able to learn a user's (expert's) approaches to performing tasks. A number of approaches have been taken to provide learning mechanisms such as 'learning-by-being-told' or 'dialogue learning' and 'programming-by-demonstration.' Maes[8,9] discusses four methods for acquiring data for learning that are incorporated in intelligent agents:

1. Observing the actions of the user and finding patterns of regularity.

2. Using user feedback such as when a user discards an agent action.

3. User provided explicit examples as training exercises.

4. Inter-agent exchange of task specific knowledge.

In all cases, expert agents need some knowledge acquisition mechanism which may include the selection of various machine learning algorithms. Machine learning consists of tools and algorithms for learning classifiers, decision trees, rules for control knowledge, learning protocols, finite state machines and other operational structures for learning group support structures. Learning protocols permit an agent to acquire or improve strategies for negotiation with other agents and humans. It is possible to distribute the learning task to many agents thereby enabling reasoning types of systems. The most recent agent/assistance software is constructed so that they improve over time by learning dynamically during actual operations of their designated tasks. This type of software is called a learning apprentice or learning assistant.

Toolkits for intelligent agent development are beginning to appear. *Lotus Notes* has under development customisable intelligent agents for filter and search assistance for various databases. The SHADE system[10] implemented using the Knowledge Query and Manipulation Language (KQML) supports concurrent engineering which acts as a mediator for information consumers and providers. The *Knowledgeable Agent-oriented System (KAoS)* is a distributed artificial intelligence environment based upon the Common Object Request Broker (CORBA) standard that enables expert agent applications to be developed in distributed *Smalltalk*.

Agents which have been developed and are currently quite popular fall into two major classes, namely, filtering agents and search agents. The development of agent systems is being actively pursued. These are briefly discussed below.

3.1 Filtering Agents

Information filtering agents have been the focus of attention in the e-mail domain by pre-sorting or pre-selecting e-mail messages for

certain types of actions such as forwarding to colleagues. *Lotus Notes, Inbox Assistant, Mailbox assistant* and *Microsoft Exchange Server* have such filters. Mailbox assistant augments its user supplied rules with learned rules based on e-mail transmissions. Similar types of agents have been developed for filtering Internet news, WWW pages, library documents, video database browsing and other sources of information.

3.2 Search Agents

Filtering agents assist in interactive browsing while search agents automatically search for information. This can be of enormous help with millions of WWW pages and large library collections as well as corporate databases. The computer programs referred to as Web Crawlers or Spiders index the WWW pages of interest for their user(s). *Web Watcher* is an example of a learning apprentice which acquires knowledge about useful HTML links that correspond to a user's interest. The learning mechanism is relevance feedback from the user. Frequently Asked Questions (FAQ) files at the various Internet sites are the target of Hammond's *FAQ-Finder* which applies parsing and classification techniques to find relevant keywords and the type of question in a request (how to, verify, when, etc.). The Commerce Business Agent reasons and plans for requests that may be asked of the Commerce Business Daily's online database. Many other such agents have been developed for use with business information as well as technical reports on the Internet.

3.3 Systems of Agents

While the agents discussed above have been developed for a specific type of task, agent systems will be developed for multiple tasks that are interrelated or at least performed by the user on a frequent basis. A system of agents which can share knowledge and assist users for complex information processes is the desired goal of expert agents. At present there are no fully operational, generally applicable agent systems but rather various categories of specialist agents each designed to perform a specific task. But experiments with such systems

are underway in several academic and commercial environments. Several critical issues are being investigated related to intelligent agents. These are:

1. Natural language support in terms of natural language understanding relative to cost and technical requirements.

2. Consistent user behaviour in terms of how consistent must user behaviour be for an agent to learn.

3. Personal privacy in terms of unwanted surveillance and control.

4. Authority transfer in terms of how willing a user is to have an agent acting on the user's behalf.

These and other issues provide a fertile ground for research and development over the next several years.

4. INFORMATION RETRIEVAL AND CATALOGUE SEARCHING

Studies of OPAC use have revealed that effective use of online retrieval is difficult to achieve for novices as well as experts. A major problem area is the match between a user's search terms and correct subject terms. The correspondence between the terms a user's supplies and the terms actually used for indexing items, is typically less than 50 percent. Additional studies have shown that understanding how to use an online database or OPAC accounts for its ineffective and inefficient use.

Borgman[11] found that simple typing errors and logical errors accounted for a significant amount of user dissatisfaction with online searching. An expert agent or assistant would need to understand the user's needs, intentions and goals and develop a plan to meet the goals. An expert agent to aid in searching would also need dealing with the problem of term 'selection'. This might be done in a variety of ways such as spell checking, thesaurus look up of antonyms and synonyms, rearranging words in a multi-word search term, dealing with prefixes, suffixes as well as infixes, accounting

for varying spaces and utilising the results of previous searches to find additional terms or eliminating terms. An intelligent agent would perform the following tasks:

⊕ Learn typing, spacing and spelling errors of each user and make appropriate corrections during operations.

⊕ Learn the vocabulary used by each user and perform the appropriate thesaurus or authority file look up for legitimate search terms.

⊕ Perform word and phrase substitution based on previous retrieval results and user feedback.

⊕ Monitor the sources of information used by each user and perform a check for new additions of interest in those sources and new sources that cover the areas of interest.

⊕ Maintain an updated profile of information needs, preferences and use for each user by ageing interest areas based on date of last use.

IR-NLI is an expert system designed to enhance information retrieval performance[12]. It acts as an intermediary to assist a non-technical user in selecting the optimal data base for searching and in formulating an appropriate query.

EP-X assists users in defining and focusing their areas of interest during information retrieval. EP-X uses an iterative technique to present to the user a topic area refinement based on user input and a knowledge base of search strategies and domain knowledge. The domain of EP-X is limited to the field of environmental pollution literature.

5. REFERENCE SERVICES

Reference questions are answered using a range of sources of different types such as atlases, dictionaries and encyclopaedias. 'When to use a particular type of source' and the 'specific tool' to utilise within the source class is the domain of the expert reference librarian.

This requires a knowledge of what reference tools are available, general strategies of when to use which tools for a specific type of question and finally, a deeper knowledge of interviewing

users and the structure of knowledge within a particular domain. This can be viewed as a decision-making process. An essential step in answering a reference question of the ready reference type is to map the question to a category of tools. For example, IF the question requires information about basic concepts THEN an encyclopaedia is a good source. IF the question requires information about a word, THEN a dictionary is a reasonable choice.

Thus, the question must first be categorised into a type of reference tool category before a specific tool in the category can be selected. This can be thought of as a simple forward chaining technique which, after analysis of the question which was asked, is the initial expert system task to be performed. The next step or task can be considered a backward chaining task that attempts to map the elements of the question to the most appropriate reference source within the category determined in the first step. This is done by matching characteristics of the question to characteristics of the source. IF the question is initially mapped to the category of dictionaries because it dealt with the meaning of a word or phrase and it was a medical term for which the user wanted an illustration, THEN a specific illustrated medical dictionary such as *Dorland's Illustrated Medical Dictionary* would be an appropriate selection.

The fact that it is possible to codify the rules for answering 'ready reference' questions based on a relatively simple model does not mean that it can replace a reference librarian for all ready reference questions but it does mean that there is a level at which an expert agent can support answering ready reference types of questions. The difficulty is in getting questions into a form that can be analysed by an automated system so as to first categorise the question by appropriate tool and secondly by extracting features from the question so as to be able to choose a specific tool within a class.

As Forrester and Garner[13] point out, the world of answering reference questions is not as simple as that portrayed above. An expert system for reference must perform the role of a

human intermediary which they state comprises the following stages:

1. Translate initial user needs as stated into prepositional form

2. Confirm, and if necessary, redefine the users needs

3. Recognise the appropriate level of information required

4. Obtain the correct tools to find this information using an appropriate strategy

5. Refine the information obtained and select exception criteria

6. Respond to the initial propositions in an appropriate format, and

7. If required by the user, repeat steps 1 through 6.

The generation of a reference map based upon declarative semantics of a user's question in a prepositional form are categorised by Forrester & Garner into seven proposition types:

1. Goal statements.

2. Single assertions.

3. Information—sources relationships.

4. Technique—result relationships.

5. Explanation types.

6. Activity—methods modelling.

7. How-to (activity-select).

This scheme requires an intelligent agent to have both lexical and semantic information of all referenced knowledge sources.

The capability to respond to a user's natural language request is a goal of some expert systems in the reference arena. The ability to respond to a request such as *I need a bibliography on distributed databases* or *I need the name of a distributed database system* represents two quite different types of questions. The first can be answered by bibliographic data but the second requires the content of a source such as a textbook or a reference source such as Datapro. What they both have in common is that to understand what the question is about, some level of natural language processing must be performed and mapped to a question database which contains the answer to the question.

Several relatively successful attempts have been made to answer the first type of question (bibliography) but successful attempts at answering the second question at a general level remain elusive. The National Agricultural Library of the United States Department of Agriculture (USDA) developed a system called Answerman in 1986 for assisting in answering questions related to a wide variety of reference books and other reference materials in the area of agriculture. This was referred to as an 'Expert Advisory System.'

This was followed by a system called AquaRef which acted as an expert advisory system in the narrow domain of aqua culture. A major concept in this system was to permit chaining by users to build 'chains' of knowledge bases representing their own collections and expertise which could be chained to other knowledge bases thereby extending the overall knowledge base using a form of distributed expertise.

Ready Reference Advisor is an expert system designed to assist novice users in finding answers to common library reference questions. The system was developed for the National Agriculture Library (NAL). The system was written using an expert system development tool called 1st class[14]. A second system implemented with 1st class is the Finders system. Finders was developed for the Economic Research Service of the Department of Agriculture. Finders uses a menu based system to guide users to the information they are seeking.

RAS (Reference Advisory Systems) is an expert system developed at San Diego State University to assist in providing reference services[15]. The RAS system provides a natural language interface to reference service in the areas of materials science, computer science, public health and nursing.

Intelligent agents[16] to be experts in specific domains for answering reference questions would need:

- A database of domain sources, a database of frequently asked questions (FAQ) with answers

- A thesaurus for vocabulary control, a concept or topical hierarchy or network of domain knowledge

- A parser plus lexical and semantic analyser for analysing questions being asked

- A powerful pattern matcher

- A classifier for determining question and answer categories

- A search engine for retrieving relevant answers, and

- A relevance feedback mechanism to learn how to better perform the reference task.

The intelligent agent would continually update its databases, its concept hierarchy and its classification scheme based on relevance feedback. If the intelligent agent maintained a profile of users who asked questions and if relevance feedback information related to previous questions asked, it could perform its task in a very personalised manner.

6. REFERRAL SERVICE

McDonald[17] describes a referral/reference prototype called Distref used by students involved in distance education at Charles Sturt University in Australia. This prototype system captures knowledge of academic disciplines as well as knowledge provided by lecturers and librarians who deal with inquiries from students. The system is composed of the following major components:

- Student profile which contains the courses and subjects the student has or is taking which indicates the level of complexity for a question or a concept.

- Local reference sources to which a student can be referred which is a knowledge base for the system.

- CD-ROM catalogue of bibliographic references available.

- Search function for searching the catalogue.

- Discipline map which contains concepts and associated relationships found in textbooks, lecture notes, thesauri and knowledge classification schemes.

- Browse function for navigating the discipline map using hypertext type links.

- Item map, which is a structure that a domain expert believes contributes to the organisation of a subject field and these are used to construct the discipline map.

- The system has a knowledge editor module for creating and maintaining concept hierarchies and a knowledge integrator module for building structural relationships among concepts.

- Distref attempts to stimulate concept formation and exploration by users who do not have a reference librarian or instructor readily available for asking questions and receiving answers.

An intelligent agent for the student in a distance education programme might observe the information (in the form of a course syllabus, lecture notes, exercises, problem sets, etc.) related to the current task (specific course related activities) a student is working on and could launch a search for potential relevant referral items, experts, locations, organisations and other artefacts. The agent might have learned a 'pattern of interaction' and when the student needs to be referred to is capable of detecting sources of help and offer suggestions without being explicitly asked to provide such help. This is analogous to a teacher observing students in a class during a lecture or question-answer session and determining whether they need help in understanding the concepts being presented. It is also analogous to a teacher observing a student attempting to solve a problem or perform an experiment and detecting that they need assistance. In some cases, the amount of time spent on a particular activity indicates a need for assistance whereas in other cases there is a pattern of interaction with the material, problem or experiment that indicates a need for assistance.

The difficult task for such an intelligent agent is not monitoring what the student is doing at

the workstation since it is relatively simple to capture all keyboard, mouse, window and data related activities. The difficult task is determining how to classify and interpret the monitor data such that interaction patterns can be detected and utilised to provide assistance. Such an intelligent agent must be able to learn from past experience with the user or by having an expert such as a teacher to enter patterns known to indicate when there is a need for assistance.

7. INDEXING AND ABSTRACTING

Traditional information storage and retrieval (ISR) systems use an inverted file approach to indexing. This causes problems with the post-coordination of uncontrolled vocabulary resulting in poor recall/precision and the storage of unused terms. The use of positional, truncation and substitution operators alleviates some of this problem for searching but syntactic and contextual information is lost in an inverted index.

The Structured Information Management Processing and Retrieval (SIMPR) project, funded by ESPRIT, attempts to process source documents and make them retrievable using expert system techniques rather than full inversion of the documents. SIMPR maintains a model of the information base into which documents are entered and from which retrieval is performed. The system uses lexical and syntactic processing modules for storage and search rather than simply character matching. The text processing system validates text words, performs a structural mark up of a document, performs a morphological analysis of each word via a lexicon lookup using 400 disambiguation rules and then performs a syntactic analysis of words relative to the role they play within a sentence and appends an appropriate role tag.

The task of automatic indexing or subject cataloguing of documents has been approached from many angles but all approaches require the selection of terms from the text of the document or document surrogate. This involves selecting those terms most indicative of the content or 'aboutness' of

the document and mapping it to existing terms via a thesaurus or creating a new category of terms when no thesaurus match is found. Expert systems attempt to select anchor terms or phrases using a set of rules or heuristics such as:

- If a word or phrase occurs both in the title and other document structures, it is preferred.

- For conferences and workshops, title terms are preferred.

- If an abstract or summary exists, terms or phrases in these are preferred.

- If a classification number exists, select terms that match the class number's textual definition.

The SIMPR indexing software works by extracting analytics (index terms) that represent the information content of the text and classifies the index terms for subject searching. Each sentence in the text is passed through a series of augmented transition networks (ATN) for extraction of candidate analytics using a set of processing rules. The candidate analytics are then passed through a series of normalisation and transformational procedures to ensure standard structure and word forms. The words are also compared to a set of stop word lists for possible rejection as an index term. The results are presented to the user for rejection, modification or acceptance.

The SIMPR classification system uses an expert system to learn how a user classifies documents and suggests likely classification categories. This learning approach is seen as the most practical method for determining classification attributes. This function falls into the category of an intelligent agent since it must observe an expert user and learn how the user performs a task, e.g., classification. The suggestion of classification categories is an assistant function. Therefore, the SIMPR classification expert system falls into the fuzzy area between a classical expert system and an intelligent agent/assistant.

Index Expert is an expert system[18] that assists librarians in selecting the appropriate index terms for a document Bailey. It also assists in abstract development. Index Expert is designed as a generic indexing tool.

Martinez, Lucey and Linder[19] describe an indexing system developed for the abstracting & indexing service of the American Petroleum Institute. This system is domain specific to the area of petroleum technical literature. The system applies its rules to the American Petroleum Institute's Thesaurus to suggest terms for indexing.

The Artificially Intelligent Document Analyser (AIDA) project developed a range of document analysis techniques to indicate the content of a document. AIDA divides document into three stages:

⊕ Document structure analysis.

⊕ Content analysis.

⊕ Output generation.

The document structure analysis breaks the document into words, sentences, paragraphs, sections, etc. and identifies structural entities such as headings, numbering conventions, and other document structural and countable facts.

Content analysis is performed using sentence position, cue identification and section headings to identify key sentences which are then selected as representing the content of the document in much the same manner as an expert abstracter. The knowledge acquisition portion of AIDA used a professional librarian in the Australian Parliamentary Library in order to develop the knowledge base. The output is an abstract indicating the content or 'aboutness' of the document using the 'key' sentences found during the content analysis processing phase.

The AIDA system can be thought of as having many characteristics needed for an intelligent agent or assistant. If an abstracter, indexer or subject cataloguer had such an agent available, it would present its suggestion on the content of the document which could then be utilised as a draft abstract, a surrogate for indexing purposes, or a statement useful for selecting subject terms and classification numbers. This agent might interact with an indexing agent, a classification agent and or a subject cataloguing agent to provide a basis for their specific tasks.

An expert agent might also observe professional abstracters and indexers as they perform their work tasks and learn the patterns used by them to create abstracts and select or generate index terms. The agent could then determine the sources within documents, and thesaurus or subject classification schedules of 'key' sentences for generating abstracts and the index terms selected for documents. The agent could then determine what patterns exist in these selections for different document types as well as similarity measures among documents so as to use this knowledge to make suggestions to an abstracter or indexer.

A simple similarity measure between a new document to be abstracted or indexed and previously abstracted and indexed documents could provide prototypical abstracts and index terms for a new document. This will require maintaining information about the source of key sentences and index terms for documents already abstracted and indexed in such a way that the intelligent agent could utilise this information in making suggestions for new documents to be processed. Structural information about documents will be necessary to perform this task. This would suggest that the structural markup of documents such as that provided by the Standard Generalised Markup Language (SGML), will be required for such an intelligent task. Thus, it will be possible to easily and quickly identify structural components of documents such as title, author, abstract, summary, conclusion, headings, tables, lists, formulas, diagrams, etc.

It is hypothesised that expert abstracters and indexers utilise the structural components as cues for selecting key sentences, index terms, classification numbers, subject headings, etc. The use of thesaurus, subject heading list or classification schedule as a verification mechanism for selecting terms, phrases and key sentences might be more useful if the entries in these devices had some indication or explanation of how they were derived, what the sources of derivation were and under what circumstances they were most useful. This is analogous to the roles and links concept associated with index terms and subject headings.

8. COLLECTION DEVELOPMENT AND ACQUISITION

The economic crisis related to the acquisition of library materials has increased the importance of selecting and acquiring the best and most relevant materials within budgetary constraints. Selection of materials has always been the domain of experts in collection development. Expert systems have been developed to select materials in very specific domains such as the Monograph Selection Advisor which supports the selection of materials in classical Latin literature and others such as that described by Rada[20] which makes decisions on journal selection. An expert in any area including collection development must have three types of knowledge, namely:

* Domain theory

* Problem solving, and

* Heuristic rules

where domain theory + problem solving = heuristic rules.

In collection development, the following kinds of knowledge is required:

⊕ General knowledge of libraries, e.g., a library model.

⊕ Knowledge of selection rules such as selection criteria, relevance of criteria to a particular situation, relative importance of rules, etc.

⊕ Knowledge of domain attributes such as authors, publishers, existing collection, development objectives, user requirements, etc.

⊕ Knowledge of ratings for authors and publishers in specific fields and why they are rated at a certain level.

⊕ Knowledge of library goals as related to collection development and the relative importance of these goals.

⊕ Problem solving knowledge in terms of general strategies, sequence of using criteria and relative weighting of criteria.

The critical element is the selection criteria used in the expert system. These can be acquired via user defined criteria as an enumerated list with sub issues, range of values for certainty and how different criteria can be combined for establishing priorities. Another approach for establishing selection criteria is to use machine learning based on a decision theoretic model by observing how an expert collection developer classifies potential selections in terms of acceptance and rejection and the explanation for such decisions. The statistical model resulting from such observations and the features of the items selected and rejected become the criteria for future recommendations to users for acceptance or rejection.

It may also be possible to train a neural network to build patterns for determining the features or criteria for acceptance or rejection of specific items. Of course, the selection criteria will change over time and therefore it becomes necessary to constantly monitor the selection patterns and determine when such changes are occurring. These changes must then be used to update the knowledge base by adding new rules, deleting old rules or modifying old rules.

The Librarian Assistant expert system has been built using the expertise of a professional technical librarian Debrower and Uones[21]. The domain of this system is the collection of the Applied Physics Laboratory at Johns Hopkins University. The subject areas covered include: engineering, computer science, physics, applied mathematics, telecommunications and other technical areas relating to the Laboratory.

Johnston and Weckert[22] describe an expert system called Selection Advisor. This system assists in collection development at academic libraries. The system has rules for the application of the following six selection criteria:

⊕ Subject.

⊕ Intellectual content.

⊕ Potential use.

⊕ Relation to existing collection.

⊕ Bibliographic considerations.

⊕ Language.

If we had an intelligent agent that could observe sources of requests for materials (such as e-mail requests), changes in the environment served by the collection (such as new courses, research projects, services or products), the content of the existing collection (such as an OPAC or document database), and sources of new materials (such as publisher announcements and lists), then it would be possible for the agent to be given a set of initial rules matching the collection development policy of the library or information centre and let it observe how an expert bibliographer makes decisions concerning requests and newly available materials.

The intelligent agent would then build a flexible decision model and a set of heuristics for assisting the collection development personnel. For example, when an e-mail request for materials arrives, the agent could parse the request, determine whether it is already in the collection, determine whether it is in an other collection from which it could be borrowed, verify the bibliographic entry, process it against the collection development and heuristics provided or learned and prepare a recommendation with explanation as to whether the item should be acquired and from whom. Budgetary constraints could also be a part of the rule base and budgetary data could be used as part of the decision making process.

The collection development personnel's task would change from largely data collection and analysis that currently characterise the job to evaluating the explanation and data used by the intelligent agent for its recommendations. An agent that observed multiple collection development personnel might impose a level of consistency that is difficult to achieve using manual systems.

9. CLASSIFICATION OF LIBRARY MATERIALS

The classification of library materials using the Universal Decimal Classification (UDC) system, Library of Congress Classification (LC) system, Dewey Decimal Classification (DDC) system or others is a fundamental function for libraries which has traditionally been performed by experts in classification. The N-Cube expert system reported by Cosgrove[23] is an example of using expert system technology to assist in the classification of library materials using the UDC. This system combines shallow knowledge usually represented as production rules and deep knowledge usually represented in a frame structure (representing classification knowledge as an integration of object oriented structures with associated rules and hypotheses). The N-Cube architecture consists of a parent class, sub-classes, a rule set, class inheritance and a set of hypotheses. This is represented as a tree structure which is searched based upon information in the bibliographic elements of a title to be classified. When a rule (one or more facts) matches a node in the tree, it is verified via a user or a knowledge base of facts and its classification number is extracted. The node's sub-tree can then be searched to determine more specific extensions to the base number as desired.

Classification schemes may be knowledge based such as UDC and DDC or they may be materials based such as LC. Knowledge based schemes would appear to be easier to apply in expert system technology than a materials based scheme. The basic concept of classification is to keep material about the same subset of knowledge together. Therefore, a set of rules could be formulated that depict the logic used by a lassifier to select a classification category and then build the appropriate extensions to the number. Extensions to a classification number are formulated by rules that dictate how to generate those digits beyond the general category number to provide specificity for the location of the material. Rules for handling form of material (dictionaries, periodicals, associations, historical, collections, bibliographies, etc.) are provided in the schedules of the classification schemes.

A particularly sticky problem is the classification of non-traditional materials. Schultz[17] describes the design of an expert system for classifying music scores using the Dewey Decimal System. The MUSic CATaloguing (MUSCAT) system is also designed

to assist in cataloguing musical pieces (Fulton[18]).

An expert agent could assist a classifier by utilising bibliographic elements in the catalogued material to suggest one or more general categories. For example, the author entry may be used to lookup other materials written by the same author and use its classification number as a potential starting point. This is based on the assumption that authors tend to write on the same general knowledge area for extended periods of time. Therefore, the most recent materials would be the most useful.

A match between a previously written material's subject headings or index terms might be used to verify whether the current material is in the same knowledge area. The intelligent agent might then examine other bibliographic elements such as subject headings, title words and phrases, etc. to generate a more specific classification number within the general category and to determine whether the material requires form sub-divisions. If the material requires form sub-divisions, the classification schedule form sub-division rules would be used to generate further extensions to the classification number. This would require that the form-subdivision rules be expressed as a set of IF - THEN - ELSE production rules. The expert agent could then collect data on the classifier's acceptance, rejection and modifications to its recommended classification number. This is to refine its rule base for future processing of bibliographic entries to determine recommended classification numbers. Obviously, such an agent could have a highly interactive component to assist a classifier in navigating a classification schedule and formulating form sub-divisions of classification numbers, e.g., country classes for national bibliographies.

10. ONLINE DATABASE SELECTION

The number of online databases has grown from several hundred to several thousand over the last ten years which poses a problem to information specialists or end-user who must

select a database for answering a specific query. Even though database selection aids exist in printed and electronic form they are not efficient or effective to use because of their volume. An expert system developed by Trautman and Flitter[26] attempts to provide database selection assistance. This system consists of database attributes with rankings, a user modeller, a question clarifier, a search module, an evaluator module and a ranker module to produce a recommended set of database files to use for a query.

A browser module also permits a free text search of text files and indexes of the database entries. Databases are classified using nine attributes such as kind, period of coverage, language, geography, vocabulary, viewpoint, sources, audience, and updating. Each database attribute has category descriptions and ranks used by the expert system to determine which databases are appropriate for a query. The user modeller is used to determine where the end-user fits on a novice-expert scale as well as the type of results required for its ultimate utilisation. Ambiguities and uncertainties in the query are resolved by the question clarifier module using expert knowledge analogous to a professional intermediary. This expert system is an aid to selecting the correct database from the thousands that exist.

An intelligent agent might be built that utilised the same data and module structure as Trautman and Flitter but with the added capability to store queries, user modelling information and feedback on databases that successfully responded to the query. When a new query is entered, the expert agent could match it with similar queries and user modelling data to determine whether it already had a database selection list from a previous query. The agent might still do a search and match the results with the stored selection list to determine whether its previous results were still valid. This is similar to the human expert who remembers previous searches that are similar to a current query and utilises the same databases that provided success. An extension to this process might be to have a small but representative sub-set of each database's content such that a local search could be

performed to show the user what the results are likely to be and get user feedback from this subset search.

11. LIBRARY EDUCATION

Another area of intelligent system application is in the training of professional librarians and in topics related to library science. Expert systems can provide a specialised, concentrated, and low cost method for obtaining practical library-related experience. Obensen *et al.* describe the Yes Sir expert system which is designed to assist in retrieval tasks. Associated with Yes Sir is an intelligent tutoring system that teaches the user improved information retrieval skills. As expert and intelligent agent systems become more abundant in library science, modules for enhancing the knowledge and skills of the users will be added.

12. SUMMARY

In this article we have tried to provide an introduction to expert system technology and intelligent agents/assistants as well as how this technology has potential uses in work activities performed in libraries and information centers. The problem area we concentrated on was the use of such agents to reduce information overload and reduction of clerical and repetitive tasks associated with many activities. In addition, we have tried to provide some indication of expert systems developed in these areas of activities which have proved useful and whose techniques could be incorporated into intelligent agents/assistants.

It should be clear that there is no generally applicable agent or agent system that combines all the ideas and results of the many different developments that have been explored. But there are various types of specialist expert systems and agents that have been used to explore specific problem domains with library and information centers. There are very few intelligent agents commercially available but this should change over the next few years. There does not appear to be any standardised approach to developing intelligent agents. Intelligent agents can be classified as belonging to categories such as:

1. Planning agents which plan a solution for complex tasks by acquiring case-based knowledge and using some inductive learning techniques.

2. Broker agents which act as mediators or facilitators by matching user requests against information or known solutions in databases or provided by database agents. These agents rarely use learning methods.

3. Group support agents which act as special 'broker' agents for performing tasks that support group activities such as scheduling meetings. These are based on speech-act theory and use personal preferences of group members to negotiate a solution to a problem.

4. Personal assistant agents[27] or apprentices which contain formalised learning algorithms and heuristics developed for a single, specialised application or task. Information filters are an example of this class of agent applied to browsing types of tasks.

Future agents will explore information spaces based on user activities and develop pathways to information sources which they will cache on the assumption that these sources will become valuable in the future. Future agents will be combined into systems of agents which will exchange information, execute tasks for each other and negotiate solutions to information related problems. Of course, the acceptance and usefulness of such agents will depend on user acceptance of a software product acting on his/her behalf and observing and collecting data on what he/she is doing.

REFERENCES

1. McDonald, Craig and Weckert, John. Libraries and Expert Systems. 1991, Taylor Graham, London.

2. Cavanagh, Joseph. Library applications of knowledge-based systems. *In* Expert Systems in Reference Services, edited by Roysdon, Christine and White, Howard. Haworth Press, New York, 1989, pp.1-19.

3. Durkin, John. Expert systems catalogue of applications. Intelligent Computer Systems, Akron, OH, 1993.

4. Pople, H.E. Jr, Myers, J.D, and Miller, R.A. DIALOG: A model of diagnostic logic for internal medicine. *IJCAI*, 1975, 4, 848-855.

5. Duda, R.O, Gashnig, J.G and Hart, P.E. Model design in the PROSPECTOR consultant system for mineral exploration. *In* Expert systems in the micro-electronic age, edited by Michie, D. Edinburgh University Press, Edinburgh, 1979, pp. 153-167.

6. Boy, G.A. Intelligent Assistant Systems. Academic Press, London, 1991.

7. Ericsson, K.A. and Simon, HA. Protocol analysis: Verbal reports as data. 1984, MIT Press, Cambridge, Massachusetts.

8. RULECONS. Georgia Tech Research Institute, Georgia Institute of Technology, Atlanta, Georgia, 1992, 30332.

9. Maes, P. (ed.) Designing autonomous agents. MIT/Elsevier Press, London, 1994

10. Maes, P. Agents that Reduce Work and Information Overload. *Communications of the ACM*, 1994, 37(7), 30-40.

11. McGuire, J., et al. SHADE: Technology for knowledge-based collaborative engineering. *Concurrent Engineering: Research and Applications*, 1993, 1(3).

12. Borgman, Christine. 'Psychological research in human-computer interaction.' *In* Annual Review of Information Science and Technology, edited by Williams, Martha. White Plains, NY. Knowledge Industry Publications, 1984, 19, 33-64.

13. Guida, Giovanni and Tasso, C. An expert intermediary system for interactive document retrieval. *Automatica*, 1983, 19(6), 759-66.

14. Forrester, C.L and Garner, B.J. Knowledge-based support for library users in *Libraries and Expert Systems*, edited by McDonald, Craig and Weckert, John. 1991, Taylor Graham, London, pp. 17-27.

15. 1STCLASS expert system applications guide. AI Corporation, Waltham, Massachusetts.

16. Carande, Robert J. Reference advisory systems board. *Information Technology and Libraries*, 1990, 9(2),180-84.

17. Riecken, D. Special Issue on intelligent agents. *Communications of the ACM*, 1994.

18. Bailey, C. The index expert system: A knowledge based system to assist users in index selection. *Reference Services Review*, 1989, 17(4), 19-28.

19. Martinez, C., Lucey, J., and Linder, E. An expert system for machine-aided indexing. *Journal of Chemical Information Computing Science*, 1987, 27(4), 158-62.

20. Rada, R., Bacus, J., Giampa, T., Gidds, C., and Goel, S. Computerised guides to journal selection. *Information Technology and Libraries*. 1987, 6, 173-84.

21. DeBrower, A.M. and Uones, D. Application of an expert system to collection development: Donation processing in a special library. *Library Software Review*, 1991, 10(6), 384-89.

22. Johnston, Mark and Weckert, John. Selection advisor: An expert system for collection development. *Information Technology and Libraries*, 1990, 9(3), 219-25.

23. Cosgrove, S.J. Item classification using the N-Cube's hierarchical knowledge representation scheme. *In* Libraries and expert systems, edited by McDonald, Craig and Weckert, John. Taylor Graham, London, 1991, pp.88-98.

24. Schultz, Lois. Designing an expert system to assign Dewey Classification numbers to scores. National Online Meeting, Proceedings, New York, May 9-11, 1989. 9-11, pp.193-97.

25. Fulton, C. MUSCAT: A music cataloguing inquiry system from California. *Library Software Review*, 1988, 7(3), 192.

26. Trautman, R. and Flitter, S. An Expert system for microcomputers to aid selection of online databases in expert systems. *In* Reference Service edited by Christine Roysdon and Howard White, Haworth Press, New York, NY. 1989.

27. Bocionek, S. Software secretary kernel: An extendible architecture for learning and negotiating personal assistants. *AI Communications*, 1994, 7(3/4), 147-60.