Designing an expert system for classifying office documents

Savic, Dobrica

Expert systems technology, as a category of computer based artificial intelligence, offers a number of possibilities for further automation and improvements in managing office documents. Because of its potential, expert systems represent an interesting subject for study and implementation. To design and develop an expert system application is a challenge. Different expert systems are already being applied in many areas of human activity. Classification of different subject domains is one of the popular targets of expert systems. Yet, there are not many expert systems developed for office document management.

In the book, "Artificial Intelligence--Programming Techniques in Basic," an elementary expert system is described as "one of the easiest computer programs to develop from a programming standpoint."(1) When developing an expert system, it is only necessary to ask a series of questions; input the answers; have a series of IF-THEN statements to eliminate any conclusions that do not fit the data provided; and print the conclusions that were not eliminated. This knowledge, combined with some skill with the QUICKBASIC programming language, coupled with the author's understanding of the scheme for classification of office documents in the International Civil Aviation Organization (ICAO), gave impetus to the project. It was named a Classification of Office Documents Expert System, abbreviated as CLOD-X. The project became more than just a design and development of a single application. It became a means to learn and understand expert systems technology as a category of artificial intelligence. It also became a gateway to the exploration of possibilities for wider use of what expert systems can offer to records and document management within the contemporary office environment.

This is a small-scale prototype project with some limitations. Firstly, from a broader perspective, because of its constraints the chosen approach may reach unique conclusions that may not be duplicated by other similar projects. More research is needed for a sound generalization. However, the project indicates that this may be a viable means to automate office document classification. Secondly, from the micro perspective, the overall analysis of the project framework suggests that limits had to be established. There are over nine thousand different file titles used by ICAO, so that scope had to be constrained to a manageable number. At the minimum, there had to be a limit for the initial design of the prototype version such as CLOD-X. Therefore, only the section of the ICAO Central Registry File Guide(2) dealing with Distribution and Sale of Publications and Documents (Code A10) was selected. Added to this section was part of the registry file section dealing with the sale of Audio Visual Aids (Code AN12). The selected sample contained all the files dealing with purchase and sale of ICAO publications. As a result, the chosen subset included over four hundred file titles or possibilities (classification suggestions), which the expert system had to take into consideration. It was a sufficiently large database (number of files) to use in the design and programming of a simple expert system. It is believed that based on this experience, a comprehensive expert system could be developed covering the whole range of ICAO file subjects.

THE FEASIBILITY STUDY

The starting point of any well planned project is a feasibility study. The design of an expert system for document classification should be no exception. Before beginning any sort of knowledge acquisition, its representation, programming and expert system building, a crucial question has to be answered: Is the project feasible? In other words, is the development of an expert system for classification of office documents feasible?

In feasibility studies basic comprehension of project planning techniques and system design suggests that the following portions may be considered:

*technical feasibility,

*economic feasibility,

*environment related feasibility.

All three factors in the feasibility assessment are essential. Because moreover, they should be studied as three interdependent parts, it is difficult to separate one factor from another. The feasibility factors in this paper are examined individually because of analytical and presentational reasons.

TECHNICAL FEASIBILITY

Technical feasibility is considered the crucial element in building a system or in designing a project. Determining if an idea can be translated using existing technology and mastered methodology into a working application, is important. It is necessary to know, before actually beginning the project design, if existing knowledge and available tools support solving of the problem. A negative answer to the above automatically disqualifies the project as feasible. Bearing in mind, however, that the study of technical feasibility cannot always reach a definite and absolute answer, and using terminology borrowed from expert systems probabilistic answer best defines the answer. In other words, if the question is whether something is technically feasible or not, the answer should be taken only as a guideline. The answer usually indicates the probability of reaching a desired solution to the problem, should proposed methodology and circumstances remain unchanged.

On the other hand, the challenge in answering the question of technical feasibility lies with the need to leave a margin of "breakthroughs." These breakthroughs occur when, against all the odds, a solution to a problem believed unsolvable is found.

In order to decide the technical feasibility of designing an expert system for the classification of office documents, two questions must be considered:

*Is it a suitable problem for application of an expert system?

*Is the required knowledge easily available?

Deciding the suitability of a particular problem or domain for the design or application of an expert system is a difficult task. Expert systems is a powerful technology that

can be of great help in many areas, but it is not suitable for every problem. "A knowledge based system is not a panacea; it cannot solve every problem. 'Horror stories' abound of disasters resulting from attempts to address problems using inappropriate technology."(3)

First, it is necessary to decide whether a particular task in a domain is suitable for coding an expert system. This is done by examining if the problem is regularly and satisfactorily being solved in its everyday "non-expert-system" environment by human experts. In other words, it should be possible, given adequate knowledge and expertise to solve the problem. Diagnosing for engine malfunction and creating an optimal computer configuration are tasks which are regularly and satisfactorily performed by many experts in their respective areas. Thus, the above areas can be regarded as suitable candidates for application of expert systems technology. In fact, expert systems designed to solve problems in these domains presently exist.

Contrary to this are problems where predicting an outcome is uncertain due to the nature of the problem. In these unsuitable areas, however, proficient experts' background knowledge and experience might be, their expert skills can bring nothing more than an educated guess. "For instance, a knowledge-based application that could accurately predict stock prices or winners in a horse race would be extremely valuable; however, since no collection of humans has that type of expertise, a knowledge-based application built to solve these problems is not feasible."(4)

Classification of office documents is a job regularly and satisfactorily performed in almost all organizations and institutions. It can be defined as "the act of identifying documents or records in accordance with a predesigned filing system."(5) Classification of office documents (official records--received correspondence) in ICAO is similarly defined. The definition includes document contents analysis, matching of its subject with the subject of some existing registry file, file number identification (physical marking of the document), and a decision on its circulation. Based on the rules and ICAO records management procedures, classification is carried out daily. During the course of a normal work day, usually over 500 documents are scanned. Out of that number, some 150 documents are assigned file numbers and placed in appropriate registry files and circulated to corresponding offices for further action. Even though a small number are misclassified, the great majority of documents are classified correctly and placed in proper files. This indicates that satisfactory classification using existing rules is possible and that it is carried out on a regular basis. It requires some special knowledge and experience, but according to the initial premise, it is a suitable task for an expert system program. Although oversimplified, the above summarizes the reasoning utilized in the decision to develop an expert system for classification of ICAO office documents.

The second element in deciding the suitability of applying expert system technology to a particular problem is the existence of rules which allow experts to develop heuristics. Heuristic rules are usually the outcome of long working experience and sound judgement, and are used as intellectual short-cuts for making more efficient and effective decisions in the domain area. It is necessary to have sustainable rules and heuristics that will govern problem-solving activities and allow knowledge engineers to design and construct required expert systems. Common sense, so readily available and taken for granted, is almost impossible to translate into computer programs. Computers operate more efficiently with a previously developed set of rules, such as, classification, indexing, or circulation rules-the rules in the area of records management. As suggested by Waterman,(6) development of expert systems is possible only if the tasks are cognitive, easily understood, and do not require common sense.

Classification of office documents meets all of the above requirements for the use of expert systems. Expert system technology is regarded as an excellent tool for helping with classification tasks. Many expert systems in the marketplace support this assumption. It has been argued that the classification done by expert systems is a relatively straightforward task.(7) That is, it is a process of deciding on a single choice among a set of predetermined, specifiable, and enumerable solutions. This being so, there is a high chance that an expert system project will succeed in classifying office documents.

Determining the availability of expert knowledge required for building an expert system, as a part of a technical feasibility study, seems to be a straightforward task. Unfortunately, as it often happens with some "obvious cases," this task is sometimes more complicated than it may first appear.

Expert knowledge is usually available in two different forms:

*written knowledge,

*knowledge possessed by a certain individual (domain expert).

Both forms bring their own peculiarities and inherent difficulties to the process of knowledge extraction --acquisition. Written material regarding some expert knowledge can have a number of deficiencies. It can be dated, cumbersome, partial, complex, etc. Knowledge acquisition becomes a difficult exercise rarely offering a certainty of success.

Acquiring knowledge from an expert can also be a frustrating experience. The problem is multidimensional. Experts in a specific domain may be difficult to reach or not readily available. Some may demonstrate such confusing practical knowledge that it can be of hardly any use to knowledge engineers. Another important aspect to consider is the willingness of experts to reveal and share knowledge. Monopoly over information is a fact of life.

Usually there exists few written rules in the area of classification of office documents, and that is the case in ICAO. Still, there are governing rules and procedures which are understood and known to the person in charge of classification. The ICAO Central Registry File Guide and ICAO General Secretariat Instructions are two important sources of written instructions. There was not a major problem with the availability of expert knowledge since the author is himself a classifier knowing many of the "expert secrets" involved in the classification of received office documents. The person who does classification on a full-time basis was also available and ready to give assistance with some of the questions which required special insights.

Keeping in mind the questions discussed above and their direct application to the design of a specific expert system for classification of office documents, it can be concluded that all criteria required as part of the technical feasibility study were met. Since there were no obstacles pertaining to technical feasibility, design proceeded as planned.

ECONOMIC FEASIBILITY

In today's world of shrinking financial resources and competitive environments, a proposal for the development of an expert system must be financially sound. An answer

to the question of financial feasibility or, in other words, whether a project is cost effective, is important. Appropriate financial calculations of cost versus benefits must be performed prior to the launching of the project.

General cost factors are more or less standard ones. They include costs for:

*hardware components,

*software tools,

*time spent on knowledge acquisition, design, testing and implementation,

*consultancy or expert fees,

*computer down-time or work-process interruption if required,

*training and support,

*system maintenance.

Benefits of implementing a knowledge-based system are not always easy to quantify in monetary terms. However, all necessary efforts should be made to translate benefits, such as faster and easier processing, increased consistency and dependability (an expert system is unlikely to apply for a job somewhere else and to leave the organization), availability 24 hours a day, transportability, decreased training cost, etc. into financial figures.

Development of the CLOD-X system was primarily self-initiated, something that might be called a spare time intellectual exercise. (Why not have besides aerobics something like "intelobics?") The temptation of exploring the possibilities for the use of expert system technology in the area of records management, so as to make some of its tedious functions easier and more interesting, was enough motivation to begin and conclude the project. A number of hours were set aside from the author's spare time, and were devoted to the challenges of extracting the required knowledge, making the initial design, and writing the program. Financial costs, except for the purchase of some literature and photocopying, were relatively non-existent.

However, this does not mean that an expert system design is cost-free; usually it is a long and expensive exercise.

ENVIRONMENT RELATED FEASIBILITY

With knowledge base systems, environment related feasibility must answer two questions:

*Are the users ready to use an expert system?

*Are all other related issues, i.e., technical, psychological, organizational and managerial resolved?

It has been suggested that collaboration with the prospective users of the expert system is vital since the success of a project almost invariably depends on it.(8) Important issues to consider include:

*User characteristics--age, sex, level of intelligence, level of computer literacy, etc.

*Job specification--the user's current job and the effect an expert system might have

*User requirements--what is really required to help users with their work

*User needs--what they want and expect from the system (level of knowledge and advice, error information and explanations, operating modes)

*User modeling--the ability of the system to adapt to different circumstances and users, and to different levels of experience and competence at various stages throughout the consultation/use of the system

*User collaboration--users' involvement in the development and testing of the system.(9)

Issues such as organizational changes, management readiness, and technical support are also of importance in this part of the feasibility study. If the technical support for the maintenance of equipment and the system as a whole (including knowledge base maintenance) is insufficient, and if management is unprepared, inflexible, and unable to accommodate necessary changes, chances of success are slim.

KNOWLEDGE ACQUISITION

Knowledge acquisition is probably the area most studied in expert systems technology. It is defined as a process of "eliciting, analyzing and interpreting the knowledge which a human expert uses when solving a particular problem, and then transforming this knowledge into a suitable machine representation."(10)

Some authors argue that "the process of acquiring heuristic knowledge from a small number of experts is unique to the development of such applications."(11) Others argue that "knowledge acquisition is not an entirely new endeavor. Those involved in the development of heuristic programs and decision-support systems have, for quite some time, faced very much the same problem as is now faced in expert systems."(12) Whatever the case might be, knowledge acquisition is of utmost importance for the development of expert systems, so particular attention has to be devoted to this project phase. It should be also mentioned that knowledge acquisition and knowledge representation are two sides of the same process. Knowledge acquisition and knowledge representation are phases of expert systems development that proceed virtually hand in hand. Both phases are vital to the integrity of the rule base for the expert system.(13)

The crucial moment in the knowledge acquisition phase is the decision concerning the methodological approach to be taken. As mentioned, knowledge can be acquired either from written sources or from domain experts. Dealing with experts and eliciting desired knowledge requires long and thorough preparation. All suitable existing techniques as well as newly designed ones have to be carefully examined before application. The scope of available techniques is wide and generally covers the following:

INTERVIEWS. This is the most common method to acquire knowledge used by many disciplines. It has a number of advantages including a long history of usage and the perfection of the technique as a result of its exploitation by other fields. In other words, this methodology is readily available.

OBSERVATION. As a technique, this can be applied with or without active participation. This choice depends on the circumstances such as the domain expert's willingness to cooperate and the necessity to observe the actual work without interruption and outside interference.

MULTIDIMENSIONAL TECHNIQUE. This technique concentrates mainly on gathering structural criteria used for organizing domain knowledge. This includes card sorting, multidimensional scaling, repertory grids, proximity analysis and matrix techniques.(14)

THINK ALOUD PROTOCOL. A very popular technique, used mostly by cognitive scientists, where the knowledge engineer records, on paper or using some other means, exact steps which an expert takes in order to solve a problem. The problem solver, in this case the domain expert, explains the activities he/she undertakes while working on the problem. This is a very time consuming technique because it requires a very detailed list of steps taken and actions which later on have to be analyzed in such a way that enables subsequent extraction of knowledge.

The technique used for acquisition of knowledge required for the development of CLOD-X system differs from the above. While the interview technique was used to clarify some of the problems encountered, the main technique applied was that called "the domain expert as the knowledge engineer technique." Since the author acted as the domain expert, extra work was necessary also to act as the knowledge engineer and achieve a working expert system. It was demonstrated in this project, as well as in some other projects, that "it takes less time to train a domain expert in expert systems than it does to train a knowledge engineer in the specific domain."(15)

KNOWLEDGE REPRESENTATION

After collecting the information needed for designing the expert system, that is, acquiring the necessary knowledge, the next phase was proper knowledge representation. The representation must take a form where its structure becomes meaningful and easy to manipulate by a computer. A number of ways were designed to meet this objective. A comprehensive classification of various knowledge representation forms is offered as follows:

*rules,

*frames,

*multiple contexts,

*models,

*blackboards.(16)

The first two ways, which are the most common ways of representing knowledge in expert systems, will be discussed in greater detail. These are also the representation techniques used in the development of the CLOD-X expert system.

RULE-BASED REPRESENTATION

The most popular, most often used, and probably most suitable form of knowledge representation in expert systems, is rule-based representation. This is the method used for representing knowledge in the CLOD-X. Rules are also called production or IF-THEN rules. From a programming aspect these are IF-THEN-ELSE rules. (Most of the expert systems shells, readily available for purchase and relatively inexpensive, use the rule-based representation.) However, the popularity of rule-based type of representation should be viewed with some caution since this type of knowledge representation is at times unsuitable for every unique situation.

The main advantage of rule-based knowledge representation is that "rules are relatively easy to construct. They enable rapid prototyping; tests can begin with just a few rules. They often seem a natural way to summarize much of what we know."(17)

Expert systems consist mainly of three parts:

*knowledge base,

*inference engine,

*user interface.

The knowledge base is a collection of IF-THEN rules. The inference engine is the software equivalent of experts' heuristics for deciding how to go about solving a problem, which rules to use, and in which order. The user interface is a "communication window" which allows the user to define the problem to the expert system and, at the same time, enables the expert system to ask questions, or give answers, and display results.

A topic often mentioned in rule-based systems is the order in which the inference engine executes the rules. It could be either a forward or backward chaining technique. If we define the IF part of a rule as a predicate and the THEN part as a consequent, then in "backward.-chaining, the inference engine works backward from hypothesized consequents to locate known predicates that would provide support. In forward-chaining, the inference engine works forward from known predicates to derive as many consequents as possible."(18)

Practically speaking, if a problem to be solved in the area where the goal is stable, or the possible solutions could be listed, then backward-chaining is the best approach.

Such applications are diagnosing, selection and/or similar problems. Forward-chaining is the more suitable approach if it is necessary to begin with facts that do not indicate the direction for a possible solution. Whatever the case, it should be noted that most applications are some type of so-called mixed chaining. This refers to the strategy of using both forward and backward chaining within a single knowledge base. There are good reasons to do this. The line that distinguishes forward and backward chaining is thin and difficult to draw. From the point of view of how an expert may solve a real-world problem, these differences may even be considered as artificial. Dividing problems into "types" is somewhat arbitrary since many problems display properties of both types. Accordingly, some problems may benefit from a combination of backward and forward approaches. Some problems may need only backward chaining or only forward chaining. There are "coincidences occurring in a world rich with problems in which both techniques can be useful for addressing any one problem."(19)

FRAME-BASED REPRESENTATION

Frame-based representation is another way of structuring knowledge for use by an expert system. It is often combined with a rule-based method, thus making a powerful system for solving complex problems. A frame is a structural representation of an object or an idea.(20) Therefore it can represent physical as well as an abstract entity. Its main characteristic is its internal structure. Each frame consists of a number of slots which are given distinct names. The slot contents are the attributes, values which describe the object or the idea. These values could be numbers, words, symbols, pictures, pointers and so on. This type of representation is very useful when the expert system is to handle a large amount of facts or data.

As mentioned earlier, this type of representation is often combined with a rule-based system. CLOD-X used this type of representation when a list of countries and corresponding country numbers were created and later used by the system for constructing the file number. A number of registry files is subdivided by countries so the file extension is given the appropriate country or regional extension number. (For example A10/100 is a general file number for the sale of publications to countries, while A10/100.10 is the same subject file for Canada.)

BUILDING THE EXPERT SYSTEM

Initially, an attempt was made to build the system using an expert shell. Although the shell allowed straightforward and easy design of the rules, it required extensive preprocessing of data. As the number of the programs for preprocessing grew, it became obvious that it would not take much more effort to write a whole expert system using a computer language. Therefore, the decision was made to write the CLOD-X expert system in QUICKBASIC. Writing the program was relatively easy. As the CLOD-X project demonstrated, QUICKBASIC is a user friendly language capable of producing satisfactory results even in expert systems research. CLOD-X is a simple prototype program designed with the main goal of doing the job, rather than impressing with some attractive programming features, such as windows, fancy help screens, etc. However, this is one of the areas in which some further development and programming is desired.

CLOD-X was developed for a standard MS-DOS based personal computer equipped with a color VGA monitor. There are no additional RAM or CPU speed requirements. Since CLOD-X comes in a compiled form, as an EXE file, accompanied with some extra ASCII files, it does not require special installation.

KNOWLEDGE BASE

During the knowledge acquisition phase it was concluded that it was necessary to find answers to only three main questions and to have all the information necessary for proper classification. It was assumed that all the documents (received correspondence) were requests for the purchase or for the distribution of printed ICAO publications. Logically, this should have been the starting question, but since it was the basic assumption, that question was omitted. Therefore, the three principal questions were:

*What type of material is requested?

*Who is the originator of that particular request?

*What is the originator's geographic location?

A modular approach was adopted as the one which is the most easy to develop, follow, and change or modify if necessary. Based on these three questions, three modules were developed with the total of 43 IF-THEN production rules. The two first questions were put to users in a menu type form. It was left to users to decide the appropriate category and to enter the corresponding number. Features are built into the program to warn the user if an incorrect answer has been entered.

In some cases it is sufficient to know only the document type so as to determine the file number. An example of a simple rule that would provide the appropriate file number and the office where the order should be sent to is:

IF -- It is a request for an Air Transport Reporting Form

THEN -- The file number is A10/16. It should be sent to C/STA.

Another example is a rule that would be fired only after learning the answers to the first and second question. It is a slightly more complex rule:

IF -- It is a request for the printed document

AND -- The originator is an international organization other than United Nations

THEN -- The file number is A10/14. It should be sent to S/DSU (Supervisor Document Sale Unit).

The third question, the originator's geographic location, was at first considered to pose little problem. It turned out to be the most complicated to incorporate into the program and to write the actual coding. There was the possibility to just make a list of countries and let the user choose the corresponding number, but due to the number of countries (165) this approach turned out to be impractical. As a result, it was decided to let the user type in the answer, i.e., the name of the country from where the document (correspondence) originated. This was the exception to all other questions which were presented in menu forms. Therefore, a short subprogram was created to search a country data file and to find the corresponding country number which had to be added at the end of some files which were subdivided by countries. The country

names and numbers were stored in an ASCII file which was easy to build and to edit if required. A provision was also made for the program to accept various name inputs, such as UK, United Kingdom; or US, USA, United States.

The problems with the geographic location did not end with the above approach. Since the ICAO File Guide has further subdivisions for Canada and US files, two additional subprograms were added to meet these requirements. These were two additional menu type questions which would appear only if the indicated country was Canada or US.

Further on, two extra menu type questions were added to extract the needed information from the user concerning geographical location. The first one was the question of region, applicable only if the originator was a bookseller. The second one was the city where the ICAO Regional Office was located, applicable only if the answer to one of the previous questions was the ICAO Regional Office, and therefore, the file was one of the ICAO Regional Office files.

Here are two samples of production rules after all the questions in CLOD-X were answered:

FIRST EXAMPLE:

IF -- It is a request for a printed document

AND -- The originator is government

AND -- Country is Canada

AND -- It is Department of Transport

THEN -- Your file number is A10/100.10.1. It should be sent to C/RDA.

SECOND EXAMPLE:

IF -- It is a request for a printed document.

AND -- The originator is ICAO Regional Office

AND -- The office is in Nairobi

THEN -- Your file number is A10/100.4. It should be sent to C/RDA.

At the end of this review of the knowledge base implemented in CLOD-X, it can be concluded that the knowledge base included the following elements:

*knowledge of ICAO Registry File Guide

*files circulation procedures in ICAO

*structure of ICAO

*ICAO document classification rules

*structure of file codes.

The expert system also "knows" the numbering system used for identifying ICAO's member states, ICAO's regional offices, and geographical regions.

INFERENCE ENGINE

The manner in which one solves a particular problem is represented by an inference engine. One of the safest means to determine the type of inference engine used in the process of logical reasoning is to determine if induction or deduction was used. Once known, consideration on whether backward or forward chaining should be applied in the inference engine can begin.

Forward-chaining systems are considered data-driven. They work in bottom-up or inductive fashion. This type of system is most useful in problem domains where there are many possible goals and all that is made known to the program are details of current conditions.(21)

The logic used for classification of office documents is inductive in nature. One usually begins from some known facts to higher, more general conclusions. In office document classification, the number of determining factors is small, while the number of possibilities is large. For this type of exercise, an inference engine with forward chaining is more suitable. This approach is the one utilized in CLOD-X.

USER INTERFACE

"Even the most sophisticated expert system is worthless if the intended user cannot communicate with it."(22) CLOD-X is a menu driven expert system designed to be user friendly. Every menu contains a prompt line which informs users about the next step they are expected to make. Menus, where the user must choose a number, have an error-protection module to help with incorrect entries ("WRONG ENTRY -- PLEASE TRY AGAIN").

A sample CLOD-X session is shown in the illustrations on pages 27 and 28.

TESTING AND EVALUATION

United Nations Food and Agricultural Organization (FAO) funded the development of an Expert System for Agricultural Project Evaluation in Developing Countries

(PROJEVAL). As the designers of PROJEVAL noted, testing has to be performed as the system evolves.(23) Rules have to be modified whenever necessary. In order to ensure that the system's performance matches the expert's performance, it has to be constantly refined and debugged. The PROJEVAL's testing was started as soon as the knowledge base contained enough rules to handle a reasonable number of factors.

It is very likely that almost every expert designer does the same thing -- "build a little, test a little."(24) It is not only the anxiousness to see the result of the designing and programming efforts that urges developers to test-run the system as early as possible, it is also the necessity to see if the line of reasoning chosen is the correct one. A wrong result (wrong answer), rarely indicates just a change of one or two program lines in order to correct it. Usually it is the logic which has to be modified and, consequently, this ongoing testing of the system during its creation phase becomes absolutely essential.

So was the case with CLOD-X. Testing and refinement were constant activities. Also, once completely finished, a number of actual documents were tested (classified) using the CLOD-X. Excellent results were achieved. However, the system covers only one segment of the total amount of office documents received by ICAO, namely documents regarding sale and distribution of ICAO publications. In order to use CLOD-X, the classifier has to go through a process of time consuming pre-selection--a type of sorting or rough classification of received documents. That was the sole method to determine the applicability of CLOD-X to certain documents. This pre-selection slowed down the classification process and reduced the benefits achieved.

Experience with CLOD-X emphasized, as did the experience with designing some other expert systems, an interesting fact: expert systems are built by experts, not for experts. It is unlikely that experts in a specified domain will wish to use a developed expert system to solve their problems. Usually domain experts find available expert systems too slow, cumbersome, below their own capabilities, and time consuming. As one ICAO classifier put it, "I can just glance over a document and tell you the file number, so why should I waste my time using the expert system?"

This brings us to the question of the actual practical value of expert systems. Generally speaking they are found to be most suitable in emergency situations when the human expert is not available. They are also great tools for training "future experts," and enabling them to gain proper experience and develop their skills. "AI can be immediately exploited to assist classers in understanding the rules of classification schemes and in applying them consistently."(25) The area of training can certainly benefit the most from the user of expert systems technology. Designers may need to reconsider their priorities and take into consideration the training aspect of expert systems, keeping in mind the requirements that that category of novice may have.

CONCLUSIONS

Developing an expert system prototype is a challenge and a great learning experience. If it is designed by a domain expert, it provides an enriching experience because of the possibility of viewing one's use of knowledge and processing patterns. For many, it is a unique opportunity to see one's own logic analyzed and systematically put on paper, or in computer program form. Being able to translate human reasoning into computer language, and to see it being used by the computer, is also enlightening. If a machine can do something that until recently was thought to be "intelligent," is our understanding of intelligence being pushed to its limits? Even without any attempt to touch upon the subject such as "Could a Machine Think?"(26) it is astonishing to realize that a simple 200-line program can perform some "intelligent" functions, such as office document classification.

The CLOD-X prototype had a goal to examine: whether expert system technology is suitable for record and document management. As has been shown by this project, it is a very promising area which should be explored further and exploited by office and administration managers. A knowledge base approach to classification of documents provides sufficient tools for the construction of effective applications. However wide the span of a classifier's knowledge is, its proper acquisition and appropriate representation enables automation of the inference process. Any user, with very little training, can correctly classify the documents pertaining to CLOD-X. Although developed for a limited domain and a relatively small number of file subjects, CLOD-X proved to be useful in solving problems such as classification. Furthermore, it demonstrated a way for a wider use of this type of artificial intelligence techniques in document automation. Use of expert systems seems to be a way for improving efficiency and effectiveness, not only of classification, but of overall office document management.

The results of the CLOD-X prototype project offer a good base for further research in this area. Once demonstrated that this type of office activity can be automated, doors open to build a system for automatic classification of documents where the computer program "intelligently" scans a document, decides on its subject without any human assistance and provides a correct classification number. The importance of such an application would be tremendous. Automatic classification can let the computer do the tedious work of scanning analyzing and classifying numerous documents received daily, saving money, time and freeing human resources to do other creative tasks. Once this first step becomes automatic, the whole chain of further activities can be computerized and also automated. This can include circulation of documents through local and wide area networks, to appropriate destinations and some required preprocessing. What might be just "some pre-processing" in the beginning, can later become a completely automatic activity. In the case of CLOD-X's domain, the full cycle might cover analysis and classification of received requests for purchase or free distribution of some ICAO documents. It can incorporate proper computer based filing and document management, trigger physical document delivery, and update stock inventory and accounting data. Printing of required reply letters and invoices to accompany the actual delivery, as well as any other type of reporting is also a potential.

The possibilities for using artificial intelligence techniques for solving office related tasks are promising. It is left to creative minds of knowledge engineers, information and documentation managers to find the most beneficial ways for their application.

SCREEN NO. 1 (WELCOME SCREEN)

ICAO DOCUMENT CLASSIFICATION SYSTEM--CLOD-X

Created by Dobrica Savic

Press ENTER to start

SCREEN NO. 2

PLEASE ENTER THE CORRESPONDING NUMBER FOR THE TYPE OF MATERIAL REQUESTED:

1--Printed Documents

2--Audio-visual material

3--Air transport reporting forms

4--ICAO Journal

Material requested is: 1

SCREEN NO. 3

PLEASE ENTER THE CORRESPONDING NUMBER FOR THE ORIGINATOR OF THE RECEIVED CORRESPONDENCE:

1--Government

2--ICAO Regional Office

3--United Nations

44--International Organizations other than United Nations

5--University

6--Training institution

7--Library

8--Bookseller

9--Corporation or person

SCREEN NO. 4

PLEASE ENTER THE NAME OF THE COUNTRY FROM WHERE THE CORRESPONDENCE CAME:

Country name: Canada

SCREEN NO. 5

PLEASE ENTER THE CORRESPONDING NUMBER OF THE ORIGINATOR:

1--Canada--general

2--Department of Transport

3--Department of National Defense

4--Department of Industry, Trade and Commerce

5--National Research Council

6--Canadian Commercial Corporation

Originator is: 5

SCREEN NO. 6

YOUR FILE NUMBER IS: A10/100.10.9 Please send it to C/RDA

SCREEN NO. 7

Do you have another document to classify?

Please answer: Y or N N

SCREEN NO. 8

It was my pleasure working with you.

Bye for now? CLOD-X!!!

REFERENCES

1. Leithauser, David (1987) Artificial Intelligence--Programming Techniques in Basic. USA: Worldware Publishing, p. 196

2. ICAO Central Registry File Guide (1990) Montreal: ICAO

3. Walters, J. & Nielsen, N.R. (1988) Crafting Knowledge-Based Systems: Expert Systems Made Easy Realistic. New York: John Wiley & Sons, p. 60

4. Ibid. p. 61

5. Evans, Frank B., Harrison, Donald F., Thompson, Edwin (compilers) (1974) "A Basic Glossary for Archivists, Manuscript Curators and Records Managers." Reprinted from American Archivist, Vol. 37, No. 3, July 1974, p. 419

6. Waterman, D.A. (1986) A guide to expert systems. Reading. USA: Addison-Wesley

7. Clancey, William J. (1984) Classification Problem Solving. Proceedings of the National Conference on Artificial Intelligence. August 6-10, 1984 at University of Texas at Austin, p. 49-55

8. Candy, L. & Lunn, L. (1988) Design Strategies for Expert Systems: A Case Study. Conference on Human and Organisational Issues of Expert Systems. Stratford-on-Avon, UK, 4-6 May 1988

9. Morris, Anne & O'Neil, Margaret (1990) "Library and Information Science Professionals and Knowledge Engineering." Expert Systems for Information Management. Vol. 3 No. 2, p. 115-128

10. Kidd, A. L. (1987) Editor. Knowledge Acquisition for Expert Systems: A Practical Handbook. New York: Plenum Press

11. Ibid. Walters, J. & Nielsen N.R. (1988), p. 35

12. Ignizio, James P. (1991) Introduction to Expert Systems: the Development and Implementation of Rule-Based Expert Systems. New York: McGraw-Hill, p. 111

13. Ibid. p. 111

14. Neale, I.M. & Morris, A. (1988) Expert Systems for Information Management, Vol. 1 No. 3, p. 178-192

15. Ignizio, James P. (1991), p. 126

16. Walters, J. & Nielsen, N.R. (1988)

17. Pedersen, Ken (1989) Expert Systems Programming: Practical Techniques for Rule-Based Systems. New York: John Wiley & Sons, p. 27

18. Walters, J. & Nielsen, N.R. (1988), p. 196

19. Pedersen, Ken (1989) Expert Systems Programming: Practical Techniques for Rule-Based Systems. New York: John Wiley & Sons, p. 81

20. Minsky, Marvin (1975) "A Framework for Representing Knowledge." In: P. Winston, ed. The Psychology of Computer Vision. London: McGraw-Hill

21. Teft, Lee (1989) Programming in Turbo Prolog with an Introduction to Knowledge-Based Systems. New Jersey: Prentice Hall

22. Mishkoff, Henry C. (1986) Understanding Artificial Intelligence. Texas: Radio Shack, pp. 3-5

23. Khan, Kemal & Doukidis, Georgios I. (1988) "PROJEVAL: An Expert System for Agricultural Project Evaluation in Developing Countries." Expert Systems for Information Management. Vol. 1, No. 1, Spring, p. 22-42

24. Liebowitz, Jay (1989) "Expert Systems in Business and Information Systems Management: Developing CESA (Contracting Officer Technical Representative Expert System Aid)." Expert Systems for Information Management. Vol. 2, No. 1, p. 37

25. Travis, Irene L. (1988) Applications of Artificial Intelligence of Bibliographic Classification. Classification Theory in the Computer Age. Conversations Across the Disciplines. Proceedings from the Conference, November 18-19, 1988. Albany, New York, p. 32

26. Churchland, Paul M. & Churchland, Patricia Smith (1990) "Could a Machine Think?" Scientific American, January, p. 32-37

AUTHOR: Dobrica Savic holds an M.Phil degree in Library and Information Science from Loughborough University of Technology, as well as a B.A. and M.A. in International Relations from Belgrade University. He has over 15 years of working and consulting experience in the area of information and documentation. The last 9 years he spent with various United Nations agencies working in registries, archives, libraries and documentation centres.