SES: A Soils Expert System.

A Thesis presented
in fulfilment of the requirements
for the degree of
Master of Science
in
Computer Science

by

E. G. Todd

July 1988

Massey University

Abstract

An Expert System Development Methodology is proposed, based on experimentation in developing a soils expert system (SES) which identifies a soil from incomplete field data. Tools for conceptual modeling of the soils domain are examined. The tools developed provide a means of recording the conceptual model of the knowledge from three different view points: the inference structures, the domain objects and the functional aspects. A review of the structures used in the knowledge bases of existing classification problems identifies eleven categories for grouping these structures. Using this information with the conceptual model, a detailed design of the knowledge base for SES is created. This design closely models those structures identified as important in the soils domain ensuring that important knowledge is represented explicitly.

Acknowledgements

This thesis is the result of a research program in which the author is indebted to the labours of many people and organizations. I would like to express my appreciation to Mr Ray. Kemp, my Supervisor, for his guidance and motivation throughout the time of this research. I wish to especially thank my husband, Arthur, for continually supporting and encouraging me to complete this piece of research.

I would also like to thank :

Dr Alan Palmer (Pedologist, Department of Soil Science, Massey University) who acted as the main domain expert and also Dr Bob Stewart (Geologist, Department of Soil Science, Massey University) and Mr Hugh Wilde (Pedologist, DSIR Soil Bureau)who both made contributions to the knowledge base of SES.

Members of the Staff of the Computer Science Department, who have given advice, assistance.and encouragement.

Table of Contents

List of Figures

Chapter 1: Introduction

During the nineteen sixties and early nineteen seventies Artificial Intelligence research concentrated on the development of general purpose problem solving strategies. The best known example is the General Problem Solver (GPS) built by Newell, Shaw and Simon (1963). It attempted to replicate the kind of problem solving humans use every day. From the nineteen seventies the direction of this research has altered. This change was initiated by the publication of research that concentrated on the development of programs for solving particular problems in specialized areas of domain knowledge. Such programs are currently referred to as expert systems.

Specifically, the research related to the MYCIN expert system (Shortliffe, 1976), triggered interest in the application of the results of Artificial Intelligence research to the wider software development arena. In the early 1980's the products of the Artificial Intelligence research centres have been taken up by innovative development units in both universities and industry. The main direction of this research has been towards the development of effective expert systems.

Medical domains are the most widely represented in the expert system literature. MYCIN is the best documented of these systems, and contains knowledge about the family of meningitis diseases. It has been extensively tested and modified. Research following on from MYCIN has led to a whole family of expert systems and expert system shells (Clancey,1986). EMYCIN (Empty MYCIN), is one of the earliest expert system shells. EMYCIN was built by removing the knowledge from the knowledge base of MYCIN. PUFF (Kunz et al, 1978) was built using EMYCIN by adding a knowledge base about pulmonary physiology. CENTAUR (Aikins,1983) was built by taking the knowledge in PUFF and redesigning the knowledge structures to represent the same knowledge using a mixture of frames and rules

Expert system technology has also been used successfully by the computing industry itself. The most notable example is R1 (McDermott, 1980), which helps to configure Digital Equipment Corporation's VAX series of computers. DIGITAL has claimed that this program has allowed it to gain a significant advantage over competitors. Other computer companies have followed this lead. PRIME has developed an expert system DOC (Littleford,1985) which analyses a memory image from a crashed Prime computer system. The expert system deduces the cause of the crash and recommends a course of action for repair. DOC only loads that part of the knowledge base that is relevant to the problem after ascertaining what model of CPU is involved and the significant peripheral devices.

The application of expert system technology is only just beginning to be widely reported. A literature search reveals that expert system technology is being applied in the space industry, in business management, the oil industry and by the military. An important application of expert system technology is in interpreting and applying complex codes of practice and specific sets of regulations. In New Zealand, BRANZ (Whitney, 1987) has successfully developed a system to help check that a building design complies with the fire regulations. Government Computing Services (Barton, 1987) has developed a system to help determine a client's unemployment benefit eligibility.

Much expert systems based research has been into knowledge representation. This has been developed in parallel with natural language processing. Both these lines of research are important for the development of large computerized data stores. Current data bases are limited to factual knowledge and lack the semantic and heuristic knowledge of an expert system knowledge base. Although special purpose query languages have been developed for accessing the data, often potential users are either unsure of exactly what they are searching for, or alternately, how to phrase the questions so the required information can be obtained. Natural language research has enabled the development of natural language interfaces to a number of database products. These help users access the information they require. Examples of a number of systems are outlined in Bonnet (1985).

The retrieval of computer based information and the use of application packages by non-computing personnel provides a diverse area for the application of expert systems technology in the commercial environment. Expert systems can be built for existing computerized knowledge sources and application packages. For these expert systems the domain of expertise is a combination of knowledge about the application and about how people typically wish to use the application. SACON (Bennett and Englmore, 1979) is a front end to an application which determines the resistance of different materials. SACON helps the user by giving advice on how to use the system for analysing structures.

Expert system technology is extending the limits to the type of knowledge that can be stored on the computer and the way in which this knowledge can be used. Expert system technology has expanded users expectations concerning the type of information they can request from a computerized system. In particular expert systems have knowledge of how they work and can therefore give some explanation of their actions and lines of reasoning.

1.1 Purpose of the Project

The purpose of this project is to investigate the suitability of using expert system technology in the field of soil science. The main aim is to evaluate relevant methodologies for analysis and design of expert systems. This study emphasizes the role of prototyping in expert systems development and the design of the knowledge base. Development of diagrammatic tools for developing and representing models of different aspects of the knowledge base form a significant part of this research.

The target application area is the identification of New Zealand Soils from incomplete field data. Not only should the system be able to identify a soil accurately from sufficient data but also it should be able to report when it is not possible to determine an unambiguous identification. At this point the system should offer help to the user by identifying the additional data that is required. It should highlight the important features that would either confirm or negate the most likely candidate soil types provisionally identified from the data so far.

1.2 The Soil Science Domain

The areas to which expert system technology have been applied are expanding rapidly. Systems that are used in specialized aspects of the earth sciences were one of the earliest application areas. PROSPECTOR, (Duda, Gaschnig and Hart, 1979) evaluates geological structures for the purpose of identifying and assessing the commercial viability of mineral deposits. The inputs to the system are the geological field data collected by geologists and the output is a site evaluation and maps of the deposit.

Other earth science associated expert systems include: DRILLING ADVISOR (Harmon and King, 1986), DIPMETER ADVISOR (Baker, 1984) and LITHO (Ganascia, 1984). DRILLING ADVISOR provides advice on solving problems encounted with drill bits when drilling exploration and production drill holes. The other two systems are used to interpret the data from down hole wire-line logging of drill holes, particularly in petroleum prospecting.

Massey University has specific expertise in pedology[1], one of the branches of the earth sciences. With the diversification of agriculture and horticulture in New Zealand the application of computer technology to the dissemination of information about soils is a timely project to tackle.

---

1 Pedology is the study of Soil Science.

The identification of a soil from field data has a number of parallels with the evaluation of geological field data. Both disciplines are basically concerned with describing three-dimensional layers including, their characteristics, boundary conditions and the processes involved in their formation. Both disciplines generalize the descriptions by producing classification systems; systems which permit specific instances of a layer or group of layers to be sorted, compared, correlated and contrasted.

Soils, like rock formations, are the product of the intersection of a number of closely interrelated processes. These processes are not discrete but progressively change over time and in space. Creating an hierarchical classification of soils is therefore inherently subjective. Classifying a specific soil also involves a degree of incidental association, as indicated by the following quote from Taylor and Pohlen (1970)

> " a full definition of a kind of soil includes a statement of both differentiating and accessory characteristics, of the permissible ranges in each, and of any likely accidental characteristics that may serve as phase distinctions. "

Before a soil can be classified the pedologist has to describe the soil. A soil description records both the soil forming factors and the soil morphology. Soil forming factors have been recorded in the site descriptor since Dokuchaiev (Neall, Palmer and Pollock, 1987) observed that

> " soils are products of extremely complex interactions of local climates, plants and animals, parent rocks, topography and the ages of landscapes"

The modern pedologist views the soil he can describe in the vertical section profile as a complete integrated, natural body that reflects the combined effects of the soil forming factors. From these direct observations pedologists have noted the associations between the site descriptors and the soil morphology. The association between the soil forming factors can be shown in a simplified soil-function equation

$$s = f(cl, o, r, p, t)$$

where:

- s = soil
- cl = climate
- o = organisms
- r = relief
- p = parent material
- t = time

This equation is a surface reflection of the processes that form a soil. Many of the soil forming processes are as yet poorly understood and the soil forming factors associated with a process are often uncertain.

For example a New Zealand podzol (Neall, Palmer and Pollock, 1987) forms under the following conditions:

cl - adequate rainfall in humid and superhumid regions

o - under kauri forest in Auckland and to a lesser extent under rimu or beech trees

r - flat or rolling relief, not on steep sites

p - coarse or medium textures parent material

t - and enough time for the expression of the process on the soil morphology

The site description holds information about these soil forming factors and the profile description holds the information about the soil morphology. A profile description is a detailed inventory of the changes in the major soil characteristics, beginning at the ground surface and extending vertically down to the underlying rock material. Each soil is made up of layers that are termed horizons. For a specific soil at the site where the profile hole is dug characteristics are recorded and indexed by depth via the horizon designation. For each horizon characteristics such as colour, consistency, porosity, size and shape of aggregates and degree of compaction are recorded to build up a detailed description of the form and structure of the profile.

The classification system used for soil identification purposes is derived from the soil mapping units used on the soil maps accompanying the DSIR Soil Bureau Bulletins. This classification system is a simple hierarchy that forms a pyramid of units, figure 1.1. The soil type forms the smallest unit while the largest unit is the soil class.

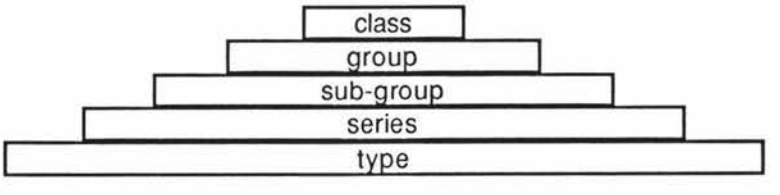

Figure 1.01 Hierarchy of Soil Classification Units.

The Soil Bureau bulletins are published on a County basis thereby forming spatially discrete units of knowledge. The information in these Soil Bureau bulletins and the associated maps is not easy to interpret correctly by either students of soil science or agriculture specialists. An expert system could complement this information by helping the user to interpret the field data they have collected. By identifying the most probable classification unit, the system could enabling the user to make more effective use of detailed soil properties and characteristics in the bulletins.

1.3 Organization of the Study.

Chapter two reviews expert system technology. The review summarizes how expert systems can be classified. The different components of an expert system, of which the knowledge base is the central component, are described. The chapter concludes with a discussion of current expert system development methodologies.

The five stage methodology proposed by Hayes-Roth et al (1983) was selected to guide the development of the Soils Expert System, SES. Chapter three discusses Identification, the first of these stages. An assessment of the feasibility of the proposed expert system. Guide-lines for determining whether a problem domain is suitable for an expert system solution are reviewed. The use of prototyping to determine the feasibility of an expert system is explored.

The conceptualization stage defines both the requirements for the system and the specification of the conceptual design of the knowledge base. The requirements of the users and the modules comprising the expert system are outlined in chapter four. Prototyping is used to clarify specific problem areas, for instance interface design.

Aspects of the specification of a conceptual model for an expert system are reviewed in chapter five. Three views of the domain knowledge are identified as important for describing the conceptual model of a knowledge base. These are used to develop the conceptual model of the knowledge base of SES.

The structural or low level design of an expert system corresponds to Hayes-Roth's formalization stage. A review of the knowledge base structures used in three existing expert systems forms the basis of chapter six. The common features of these knowledge base structures are identified.

Information gained from identifying the common structures in existing knowledge bases was used to redesign the prototype for SES. The evaluation of this prototyping exercise is described in chapter seven. The description of the types of knowledge base structures considered necessary for SES are based on the results of this evaluation.

The final chapter contains a summary of the main points discussed in the thesis with reference to the stated aims of the research. Additional applications for expert system technology in the soil science field are suggested. Extensions and improvements to the current project and associated topics are outlined as suggestions for further research.