THE EXPERT GEOGRAPHIC KNOWLEDGE SYSTEM
APPLYING LOGICAL RULES TO GEOGRAPHICAL INFORMATION

Kenneth D. Gardels
Landscape Architecture Computer Applications Laboratory
University of California
Berkeley, California 94720

ABSTRACT

The Expert Geographic Knowledge System (EGKS) represents the merger of techniques of computerized expert systems with those of geo-processing and database management. It involves the application of heuristic rules developed by experts in land management and related disciplines to the data within a geographic information system. EGKS construction must conform to rigorous design criteria to ensure that the system is capable of addressing the variations in the planning domain, that expert knowledge is accurately codified into rules reflecting its complexity and uncertainty, and that textual and graphic information is meaningfully communicated to the user.

INTRODUCTION

Environmental planning comprises the detailed information, rigorous analysis procedures, creative design and synthesis capability, and communication facilities necessary to understand and manage the relationship between humans and their environment. Planning practice relies on the application of expertise by specialists and generalists to the environmental decision-making process. The application of expertise to geographic information is the underlying concept of the Expert Geographic Knowledge System (EGKS).

The EGKS, using a computer and accessing large geographic and textual data bases, works in much the same way as the human expert. It applies logical rules stored in a knowledge base to the data in one or more geographic information system databases and to a large textual data base (called a domain cyclopeadia) in order to provide textual answers to specific geographic inquiries.

THE EGKS ARCHITECTURE

The goal of the expert geographic knowledge system is to provide expertise to an environmental specialist, a planner, or other decision-maker on the subject of a site, event, or topic specific inquiry. The overall architecture reflects the relationship between typical planning issues and the codification of expert knowledge in a computerized form. A planning problem emphatically does not involve the derivation of a specific statement from a limited set of facts. The converse is true: the planner must synthesize a wide realm of knowledge, and isolate that which is necessary to make the decision at hand. Thus the expert computer system must selectively utilize all the information available to it to deduce the information the decision-maker needs. It must be flexible enough to address a wide range of demands and meaningfully communicate its conclusions.

The architectural model of the EGKS is as follows: The user provides the system the basic parameters of the inquiry such as the location of a proposed development, the circumstances of an environmental event, or concerns about a resource issue. With this data, the expert system performs a complete investigation of relevant automated geographic information. Based on what it finds, the expert system applies rules established by human experts to make conclusions about the inquiry - the nature of the problem and realistic solutions to it. The system then assembles textual information relevant to the inquiry and produces documents detailing its findings. Finally, the planner or decision-maker can interact with the system for more knowledge or explanation.

DESIGN CRITERIA FOR AN EGKS

The construction of an expert geographic knowledge system requires a systematic approach to design. Design criteria must be rigorously applied and rationales for each standard must be explicitly defined for each component of the EGKS.

> The inference engine must be able to address the class of problems represented by environmental planning while retaining a high degree of domain independence for specific applications.
>
> The knowledge base must adequately reflect the complexity of planning and specialist expertise while remaining internally consistent and logical.
>
> Conclusions and explanations derived by the knowledge base rules must be supportable by general domain knowledge contained in an environmental cyclopeadia.
>
> The link to the GIS and other databases must retrieve the exact information necessary to address an inquiry.
>
> The output must be in formats useful to specialists, planners, and decision-makers and sufficiently flexible to accommodate varying needs.
>
> The user environment must be comfortable and encourage productivity while providing adequate power and capability to service advanced expertise requirements.
>
> The system must be amenable to updating and expansion in an open-ended, incremental fashion as new knowledge critical to landscape analysis becomes available.

Standards for ensuring that these basic requirements are met are necessary for any system design, whether domain independent or application specific.

The EGKS Inference Engine

The inference engine is the body of software code that translates the rule statements into actual actions and produces conclusions or other results. It uses the computational ability provided by list processing software to apply high-level rules to variable data.

Domain Independence. The key criterion for the inference engine is that it function independently of the facts of the domain(Hayes-Roth, Waterman, and Lenat, 1983) and remain adaptable to divergent sets of domain facts. For the EGKS, a workable inference engine must consider the types of relationships between correlated environmental data and the analytical and heuristic approaches used to study them, without being tied to specific relationships that exist only in one application or one location. Domain independence ensures technical compatibility between multiple systems, in terms of languages, coding, and so forth. Database adaptability ensures that systems can be implemented for new expertise need situations without rewriting the entire package. For a proposed application, these two factors mean that a system that meaningfully addresses local environmental issues can be developed fairly quickly.

Logical independence of the domain also implies a logic path that is complete and is not biased toward specific classes of solution. This may be described as due process reasoning, since it is based on both advocacy and skepticism. It has the distinct advantage in a planning domain of not being adversely affected by the absence of data that could be critical to a problem (Hewitt, 1985).

Antecedent-Driven Reasoning. Most expert systems are rule-based production systems, meaning that pattern matching, scheduling of rule-firing, and other logical operations are under the explicit control of an executive program. Within this framework, systems may be either forward or backward chaining, or antecedent-driven or consequent-driven, respectively (Infotech, 1981). If rules are defined as having two parts, an antecedent (the "if" part) and a consequent (the "then" part), the distinction between forward and backward chaining becomes a matter of whether any identified true antecedent produces its associated consequent(s) or any desired consequent is evaluated by testing its antecedent(s).

Most of the expert functions of the EGKS require a forward-chaining inference algorithm, so that a wide range of possible scenarios can be explored starting from the basic data. Thus the system is free to draw any reasonable conclusion from the data rather than seek out a particular conclusion (or diagnosis). This deductive process corresponds to the way in which geographic data, both manual and automated, are typically used, ie., data driven rather than goal driven. As intermediate hypotheses are derived, additional conditions may need to be established to verify a conclusion. At this point, the system may initiate backward-chaining logic in an attempt to determine if the relevant conditions are supported by the data.

Blackboard Hypothesis Interaction. A type of forward-chaining system that provides some of the desired data and goal driven functionality is the blackboard model, in which intermediate hypotheses are "posted" for examination by other rules. The blackboard-type procedure is closely analogous to traditional planning, where answers develop slowly and iteratively. The blackboard always represents partial analysis in which islands of truth begin to appear until its contents are complete and resolved (Waterman, 1981). By posting results to the blackboard complex interactions of basic environmental phenomena can be tracked and then reevaluated in terms of new knowledge and data.

One body of rules may be considered specialists and form hypotheses from basic data for posting on the blackboard - for example, the existence of a sensitive waterfowl habitat. Specialist rules are selected ("fired") based on activation rules that specify which rules, based on their content and their certainty, can best address the current best hypothesis on the blackboard - for example, to confirm or deny the existence of a wetland-associated soil type. Finally, strategy rules choose the activators that correspond to the class of knowledge needed to answer an inquiry - for example, to determine if a wetland/habitat area is of relevance to the study issue.

The blackboard can also point to slates containing relevant data extracted from the GIS or other databases, cyclopeadic information about the utilization of particular rules, and user-volunteered information about a site or event. As new information becomes available, or as new parameters are introduced, the blackboard will dynamically adjust to the new environment, and the program will select and fire new rules, reject old hypotheses and propose new ones, and backtrack and eliminate incorrect lines of reasoning.

Explanation. Most expert system packages being used today provide explanation to the user in the form of rule restatements. While explanation is important to users of any type of expert system, including the EGKS, the expert geographic knowledge system architecture expands the explanation capability significantly via the domain cyclopeadia. Material expertly extracted from the cyclopeadia explains the significance of each conclusion, not just the logical path used to reach it. Moreover, obtaining cyclopeadic materials based on content analysis of the text isolates the cyclopeadia from the rule-making and reasoning process. This allows the cyclopeadia to be revised, incrementally expanded, and updated without any effect on the inference engine or the knowledge base rules.

The EGKS Knowledge Base

The performance of an expert system is most closely related to the content of the knowledge base. Shridharan notes that the key to a thorough knowledge base and an expert level of performance is "formalizing, structuring and making explicit the private domain of a group of experts" (Infotech, 1981).

Knowledge Engineering. The acquisition and codification of expertise is the function of knowledge engineering. It involves identifying both the organization of the domain and the strategies of domain problem solving. The expert geographic knowledge engineer understands the nature of the environmental planning domain and the capabilities and constraints of the EGKS architecture. He or she is thus able to translate the heuristics of expert analysis and decision-making into the set of rules comprising a planning knowledge base. The knowledge engineer also defines the logical reasoning framework for the utilization of the rules - priority ordering, finding and hypothesis and conclusion building, etc. The duality of content and structure is especially important to environmental planning, since professional expertise consists largely of reasoned explanation and description rather than logical conclusion.

The knowledge engineer constructs a working EGKS by mapping the formalized knowledge into the representational framework provided by the blackboard/antecedent-based EGKS engine. For example, a general rule such as "Steep slopes with clayey soils tend be be unstable when wet" may become:

> If slope | 12%,
> and if soiltype = fine clay,
> then potential slope instability is high.

The use of a high level rule-writing language allows rules to be stated in a restricted natural language format, and thus rules can be back-checked against the original experts' heuristic problem solving methods. Since codifying of reasoning used in environmental analysis involves restating ad hoc general rules of thumb into much more precise language, these restatements should be carefully reviewed with domain experts to ensure their applicability and internal consistency. The knowledge engineer and domain expert together postulate rule credibility indicators to indicate the relative reliability of each rule.

<u>EGKS Rule Content.</u> Although rules are by definition high-level expressions of domain expertise, each rule in a domain knowledge base should be relatively primitive. Attempting to express too much in a rule reduces the overall certainty of that rule, and as individual rules approach the universality of large-scale models they lose the incremental, hypothesis-building advantages of the expert system. For example, attempting to add to the previously cited slope stability rule considerations of nearness to upslope water sources, drainage nets, or precipitation rates would reduce the overall credibility of the original if-then statement. Limiting the scope of rules does not mean that individual rules cannot have multiple antecedents or consequents or numerical expressions of validity. However, each rule in the knowledge base must be simple enough that a single concept is represented - one that can be empirically confirmed or denied by research and field investigation.

Rules should also address important user questions and information needs. Extraneous or marginal rules reduce the surety of fundamental domain rules because, in the forward-chaining model, each activated rule that proves true from the database is combined with other rules proved true from other data to form intermediate and high-level hypotheses.

All cited rules become part of the final explanation or conclusion, and thus multiple rules incorporate multiple database elements as well. Therefore, the final conclusion reflects reductions in reliability as a function of map overlay, as well as the lowest common denominator of each individual rule's reliability.

Finally, rules should adequately represent both explicit "textbook" planning knowledge and the intuitive, experiential knowledge implicit in expert analysis. It is the latter that produces environmentally sound advice in a complex problem-solving or decision-making context. The knowledge engineer will probably have the greatest difficulty in building an EGKS knowledge base at the level of substantiating these heuristic rules. The point to be made is that they are the rules used in traditional approaches to planning, and their insertion into an automated system does not in itself lessen or enhance their status. However, leaving them out of an EGKS knowledge base severely restricts the degree to which the system can emulate a human expert and provide meaningful information.

Logical Reasoning Paths. Rules must express not only facts of the environmental planning domain, but must also direct the interaction of other rules. Just as the planning process involves relative weighing of multiple pieces of information and assigning of priorities, the expert geographic knowledge system architecture requires rules that direct the logical flow of knowledge from database to user. Although rule consequents point directly to other rule antecedents, a strategy must be imposed to ensure that the data based process is directed toward a class of explanation or conclusion. As described above in relation to the blackboard, this prioritizing is handled via strategy rules that reduce the solution space. Strategy rules point out additional data sources or initiate a query to the user when more information is needed to formulate or confirm a hypothesis.

Basically, a rule is a statement in the form "if a, then b." a may be a primitive comprising geographic data extracted from a GIS map, or it may be in effect the b of another rule. Both the a and b of a rule may have several components as well. Thus the rule-implementing process quickly becomes a network of conditions contemporaneously interacting on the blackboard. At any one time, a set of basic data, findings about those data, and conclusions in the form of hypotheses and explanations may all be represented. The inference engine in combination with high-level strategy rules is responsible for reducing the blackboard knowledge to a series of expert statements about the environment that resolve the original problem and address the user's inquiry.

Ultimately, the reasoning, blackboard updating, antecedent checking, and consequent firing process must reach an end. That is, the system must reach a stable state in which its expertise has produced conclusions about the environment that remain unchanged unless new data are provided. At this point, the expert system can communicate the statements as findings or explanations of the environment relative to the original inquiry.

The Domain Cyclopeadia

The EGKS domain cyclopeadia is the repository of non-rule knowledge in the specific environmental planning application. It provides the user with an organized summary of knowledge related to the findings of the rule evaluation process. Cyclopeadic materials that explain findings and elaborate on their decision-making significance may be expertly retrieved by the EGKS based on descriptors of material content. By being thus indirectly keyed to hypotheses and findings, the system can also extract knowledge explaining the means and results of its advanced deep reasoning processes. Based on application needs, the cyclopeadia may contain a variety of information types and be organized according to a range of structures.

The content of the domain cyclopeadia in the EGKS architecture comprises digests of knowledge relative to specific planning problems. The material included in the cyclopeadia serves various explanation and description functions. It can justify individual rules by describing the environmental relationships between an antecedent and consequent; for example, the relationship between mapped soil type or slope and the potential for slope instability. It can describe strategies and the hierarchies of rule applications, including procedures followed by environmental scientists in determining relevant resource parameters; for example, based on a finding of soil instability, what confirming indicators, such as vegetation or geology, are used. Thus it can amplify the situations defined by hypotheses and conclusions far beyond that expressed by a rule-series consequent. Most importantly, knowledge represented in the cyclopeadia can incorporate recommendations based on laws, statutes, or other land regulatory jurisdiction, not just on the EGKS reasoning process. For example, if the knowledge base rules determine that an environmental hazard exists at a site and that development should minimize adverse effects through use of setback zones and engineering restrictions, the cyclopeadia can explicitly define what those zones or restrictions should be and substantiate them in terms of precedence or jurisdictional authority.

A useful structure for representing knowledge within the cyclopeadia is to define each bit of organized knowledge as a kernel comprising text and/or graphics, each identified by keywords. The process of retrieving cyclopeadic knowledge is a function of identifying the kernels describing a rule or conclusions. Knowledge sources are mapped to keywords, and kernels satisfying the requirement are extracted and ordered. Keyword-based retrieval is in itself an expert process, since each rule, finding, description, and so forth must have some means of identifying additional cyclopeadic information relevant to itself. This function translates into a series of expert rules of the form, "if conclusion a, then find knowledge about subject b". Of course, rules may have complex interactions that build and reject hypotheses concerning what knowledge is relevant, just as the basic environmental rules do.

The Geographic Information System Link

The knowledge-based rule and cyclopeadia architecture rests fundamentally on basic geographic information contained within an automated database. Therefore the design criteria for the GIS data, their format, and their accessibility by query is critical to the operation of the expert geographic knowledge system.

Standards for data accuracy and currency are taken as *a priori* requirements for a GIS. Beyond these standards, though, are more exacting specifications to ensure that data obtainable from the GIS are capable of supporting expert analysis represented in the knowledge base. To be truly useful, GIS data must correspond to the accuracy and scale of user inquiries, and the data classification to the environmental information that is key to analysis and planning. Thus every feature on the automated map is useful, needed delineations are all present, and unnecessary data are minimized.

The format of the GIS data is also important to system performance, though, like the data, it may be out of the hands of the system designer or knowledge engineer. Because of the high-level rule structure of the EGKS, it is preferable that the GIS be accessible via fairly high-level calls at an operating system or query language level. GIS query should comprise a functional description of the mapped data (eg., natural vegetation and elevation province) rather than a structural description of the GIS organization (eg., columns 5-7 and 12). Otherwise, database information must be built into the EGKS and any GIS update, change, or expansion requires a major effort. A relational data model (Blumberg, 1975) and a modular software toolbox of functional data manipulation capabilities is critical to efficient data retrieval (Dangermond, 1983).

EGKS Output

The expert geographic knowledge system is designed to provide documented expertise to the environmental planner (or specialist or other decision-maker) about a site proposal, an environmental event, or other land-related issue. Expertise may be disseminated in a user-specified document containing the findings reached by the system, substantiated by textual and graphic material from the cyclopeadia. Optionally, maps of selected geographic features may be incorporated into the document as well. System expertise may be obtained interactively: the user directs queries to the system following the completion of an expert review of a proposal, event, etc. The system uses its logical record and information available on the blackboard to explain its reasoning, extract additional information, or even modify its conclusions based on new data.

The EGKS User Agent

The expert geographic knowledge system is designed to interact with planning professionals, not computer operators or programmers. By eliminating the data manager link for routine automated geographic queries, users are given much freer access to their data. Such direct contact requires that the system be easy to operate, recover gracefully from errors, and provide significant amounts of assistance to the novice user. Computer systems developed in the last few years, especially in the personal computer realm, have introduced the concept of a user agent that stands between the user and the actual operating system or program. An expert user agent not only makes an expert system easier to use, it understands the types of queries being made of it and can thus interpret application needs more accurately.

The EGKS user agent is the mechanism for translating initial user inquiries into the expert procedures used to reach conclusions and the means by which the system conveys those responses back to the user. The most complex part of the user agent is concerned with the interactive inquiry review - requesting explanation, eliciting information, or resetting parameters. For explanation requests, the agent displays the rules producing a specific finding. For information requests, the user chooses from a list of available topics to obtain more information from the database or cyclopeadia; for example, soil types in an area and their suitability for construction. For the parameter resetting - or "what-if" - function, the user agent conveys the causal relationships between parameters and findings by displaying the current parameter set as defined on the blackboard. A change in any parameter (or addition of a new one) results in a new reasoning cycle and the development of new conclusions. The new findings may be displayed on the screen and a map updated, using shading or color, to show new or changed areas of concern and to facilitate straightforward comparisons of different scenarios.

Continuing Knowledge Acquisition. The process of knowledge acquisition does not end after the initial knowledge engineering phase. On-going interaction with experts and incremental addition to the knowledge base and cyclopeadia are assumed. Moreover the system expertise can be directed toward acquiring new knowledge on its own. Interactive expertise transfer programs guide the expert in explicating and formalizing his or her knowledge.

Continued knowledge acquisition programs are particularly important to the EGKS because of the size and open-endedness of the environmental planning domain. The EGKS must consistently track its own decisions and determine where its reasoning is inconsistent with decisions made by human experts. Where this occurs repeatedly, it must identify the invalid assumptions or missing rule logic and postulate a remedy. The goal is a systematic, incremental implementation of EGKS technology and knowledge that accurately assesses the complex realm of environmental planning.

CONCLUSION

Expert systems represent the evolution of computer-aided decision-making from sequential number crunching to advanced reasoning, and as such represent the cutting edge of computer applications to real-world problems. The expert geographic knowledge system applies expertise-based reasoning capabilities to environmental planning problems - land use and management, environmental protection and monitoring, resource utilization and assessment - to identify impacts of and alternatives to development, to describe environmental effects of various activities, and to explain complicated resource issues.

The sophistication of the EGKS concept requires an equally sophisticated inference engine as well as data, information, and knowledge bases that are complete, accurate, current, and consistent. Where this is done, planning and decision-making in complex natural and institutional environments can benefit enormously.

REFERENCES

Blumberg, M. H. 1975, A Generalized Geographic Information System for Managing a Non-Redundant Demographic and Resource Data Base: Computers, Local Government and Productivity, Volume I, O. M. Anochie, ed., p. 181-193, URISA, Chicago. (Papers from the Thirteenth Annual Conference of the Urban and Regional Information Systems Association, August 24-28, 1975, Seattle.) IBM, Federal Systems Division.

Dangermond, J. 1983, A Classification of Software Components Commonly Used in Geographic Information Systems: Proceedings, United States/Australia Workshop on Design and Implementation of Computer-Based Geographic Information Systems, D. Peuquet and J. OCallaghan, eds., p. 70-91, IGU Commission on Geographical Data Sensing and Processing, Amherst (NY). Environmental Systems Research Institute.

Hayes-Roth, F., Waterman, D. A., and Lenat, D. B. 1983, Building Expert Systems, Addison-Wesley Publishing Company, Inc., Reading (MA). (Teknowledge Series in Knowledge Engineering.) Teknowledge, Inc.; The Rand Corporation; Stanford University.

Hewitt, C. April 1985, The Challenge of Open Systems: Byte, 10(4): 223-244. MIT Artificial Intelligence Laboratory.

Infotech 1981, Machine Intelligence (collected papers), Pergamon Infotech Limited, Maidenhead, Berkshire, England. 407 pages.

Waterman, D. A. 1981, Rule-Based Expert Systems: Machine Intelligence, Infotech, ed., p. 323-338, Pergamon Infotech Limited, Maidenhead, Berkshire, England. 407 pages. The Rand Corporation.