A prototype expert system involved in classificatory tasks for sponge identification*

MARTA DOMINGO i GOU
Institut d'Investigació en Intel·ligència Artificial, Centre d'Estudis Avançats de Blanes, CSIC, Camí de Santa Barbara s/n, 17300 Blanes, Girona, Spain.

SUMMARY: A first evaluation is made of the advantages of using artificial intelligence techniques in producing automatic tools for biological classification and identification. The general principles of knowledge-based expert systems are set forth and their application tested in the domain of the Porifera by means of SPONGIA, a rule-based expert system devised to aid in the classification of North Atlantic-Mediterranean sponges. Special attention is given to problems involved in constructing the knowledge base.

Key words: Artificial intelligence, expert systems, knowledge engineering, classification, identification, Porifera, Demospongiae.

INTRODUCTION

The basic goal of taxonomic studies of living organisms is to establish their phylogenetic relationships on the basis of some observable features and to build up a store of information from which deductions can be made about biological processes. This goal is difficult to achieve in an actual classification because of the uncertainty surrounding the knowledge used to arrive at biological classifications, the lack of strong theory in this domain, the inherent imprecision of natural language, and the subjective interpretation of features. Thus a certain degree of controversy tends to arise around any given classification.

Nevertheless, no ecological investigation can move forth without some taxonomic analysis of the organisms present. An error in identification at this stage could invalidate any subsequent work in a given area (KNOWLTON et al., 1992). At the same time, authoritative identifications require years and years of study of more or less extensive taxa. These are among the reasons that efforts have been made to introduce automation to the difficult task of classification.

Classification, it should be mentioned at the outset, refers to the recognition of groups among a number of species, while identification presumes the existence of groups to which an unknown could be assigned. Although different methods have been developed to facilitate these two activities, they are in essence closely related, and it is difficult to keep them separate (see GOWER 1973). Thus, although taxonomic literature deals mainly with classification problems, the underlying issue is one of identification. Conversely, a fundamental purpose in identifying an unknown specimen is to obtain an idea of its classification.

The central concern of this study was to devise an artificial intelligence (AI) tool to aid in the identification of North Atlantic-Mediterranean sponges. The MILORD II programming environment was used for

* Received December 17, 1992. Accepted June 9, 1993.
this purpose. Several kinds of automatic identification devices that have been developed in the last decade. Expert systems are particularly well suited to capturing biological (taxonomic) expertise and making it available to scientists worldwide (WOOLLEY and STONE, 1987). From the cognitive viewpoint, classification can be defined as a broad activity of human intelligence concerned with the ability of humans to identify things (e.g., a broken chair can be recognised and differentiate from a pile of wood). Consequently, classification is seen as a main (or generic) task of a system that is to be able to identify or diagnose.

The Porifera have been chosen to test the applicability of AI tools to biological classification for two reasons. First, this is a field in which experts are scarce and thus there is a great need for efficient automatic identification tools. Second, the problems associated with constructing a knowledge base for the Porifera can help researchers improve the MI LORD II tool from the knowledge engineering point of view.

It should be emphasised that the current system has only been tested on a small scale. That is to say, an exhaustive protocol of validation is still needed to test its actual capabilities and refine the expertise. A further goal of this research is to determine what is known about the biological classificatory task itself (hierarchy structures, problem solving strategies, domain models) in order to formalise the underlying methodology involved in biological classification in a broad sense.

In the next section a general explanation of expert systems is in order, along with definition of other concepts of AI as knowledge acquisition or knowledge representation (for further details on AI contents, see RICH, 1983; WINSTON, 1992).

MATERIAL AND METHODS

The material on which this research is based is the knowledge associated with the problem domain, in this case the Porifera. This knowledge has been obtained from two chief sources: direct interaction with an expert and several discussions about particular topics with other field experts, and data supplied by books and journals. Where controversy between different authors existed, the criteria followed have been explained. The vocabulary and the structural concepts in the knowledge base have been formulated along the lines suggested by BOURY-ESNAULT and RÜTZLER (in press).

The process of extracting and formalising expert knowledge is called knowledge acquisition, and it is considered a major bottleneck in the development of expert systems (Hoffman, 1987). Some proposed methodologies for knowledge acquisition (Breuker and Wielinga, 1984; Wielinga and Breuker, 1986a, 1986b; Steels, 1990) have been explored in this work.

Expert Systems and Expertise

An expert system (ES) is an artificial intelligence device that is able to solve problems typically requiring human expertise. In other words, an ES is a program capable of making inferences on the basis of some knowledge provided by an expert, and, from this viewpoint it can be also referred to as knowledge-based system. According to this definition, an ES has two components: the inference engine, which makes the inferences, and the knowledge base, which encapsulates the knowledge of the given domain. The inference process develops some general mechanisms for interpreting data over the particular domain knowledge.

The design of the knowledge base must follow syntax and semantics appropriate to the inference engine to be used. In other words, knowledge needs to be expressed in some formal way to be used by a knowledge-based system. Whatever paradigm is chosen to represent knowledge—whether frames, semantic nets, or production rules (for further explanation see REICHGEIT, 1991)—the knowledge engineer must design and build a data structure representing the knowledge that has been acquired in a particular domain from one or several experts in the so-called process of knowledge acquisition. In practice, this is referred to as implementing the knowledge base.

Several preliminary steps must be taken before any effort is made to build such a system (Breuker and Wielinga, 1984):

a) Assess ES feasibility: An ES stands no chance of success if there is no actual interest in the task it performs. The time of human experts is scarce and expensive and an expert can easily lose his interest.

b) Identify the User: The interface characteristics and scope of a system will depend on the kind of user who will apply it.

c) Plan the domain structure: Aspects concerned with the knowledge acquisition must be explored, namely, conceptual structure, domain theory, availability of data.

Some attention must also be given to the concept of expert and expertise. The ability of an expert to perform a specific task is the product of a great deal
of practice. Some aspects of expert performance can be delineated (C努t et al., 1988):

a) The expert obviously has a great deal of domain knowledge.

b) Experts easily perceive meaningful patterns in their domain, which reflect a detailed and clear organisation of their knowledge.

c) Experts can solve problems rapidly and with little error, in part because of the two characteristics just mentioned plus their ability to automate some routine tasks and thus make more efficient use of memory in induction tasks.

d) Experts can comprehend a problem at a deep level, that is, at the level of its underlying principles, although they may be unable to make this comprehension explicit.

These factors must all be taken into consideration in designing a knowledge-based system that is to perform an expert task. Since these characteristics are domain independent, they apply to any expert system.

By taking these and other characteristics of expert performance into account, it is possible to arrive at a cognitive approach to knowledge engineering (Slatter, 1987). The resulting approximation will capture not only the knowledge of a given kind of expert, but also the way in which the expert represents, organises, and uses that knowledge.

Thus specific attention must be given to expertise in order to capture the typical strategies of expert performance in a concrete domain. In other words, to discover how an expert achieves the solution to a problem. This is the subject of problem solving and can be seen as the search strategy for achieving a path to a solution.

Another important factor to consider in the approach to problem solving is uncertainty. Even for experts, it is not always possible to arrive at a decision with absolute certainty. Actually, this highlights another of their abilities, which is to obtain more accurate diagnoses than do novices in real world environments that are characterized by uncertain information. The uncertainty may arise from several sources: the loss of information, instrumental error in scientific measurement, observational subjectivity, uncertainty in judgement, incompleteness of the domain theory, random events, and so on.

Assumptions may have to be made about information that is not available and judgements made in real world situations. In other words, the expert is often forced to reason with uncertain factors and sometimes in limited situations.

**MILORD II Programming Environment**

The MILORD II programming environment (Sierra and Godo, 1992b) was developed using COMMON LISP and it runs on a Macintosh environment. This expert systems shell (currently in development) is an extension of MILORD (Sierra, 1989), which has been successfully applied to develop several medical ES (Verdarguer, 1989; Belmonte, 1990). Several characteristics of MILORD II such as local logic definition, the partial deduction mechanism or the declarative backtracking, are beyond the scope of this paper (see, e.g., Agustí et al. 1992; Puyol et al., 1992; Sierra and Godo, 1992a).

Some features of Milord II which are particularly interesting to model our problem domain must first be explained. In addition, the programming constructs of MILORD II are illustrated with some examples from SPONGIA.

The first feature to mention is modularity. The basic structural unit of MILORD II expert systems is the module (Table 1). Each module contains (a) an import interface that enumerates the information to be provided by the user or by other modules (e.g. features are directly queried to the user while the process is running) and (b) an export interface that fixes the output contents of this module. Other components are a set of knowledge units (facts, rules, metarules) concerning the domain knowledge. Modules can be given a structure by several operations. This structure is essential to model the strategies of the expertise task. There are three basic techniques for combining modules in MILORD II (Sierra and Godo, 1992b): (a) modules can be composed through the declaration of submodules, in order to establish hierarchical relations between modules; (b) modules can be refined, that is used to perform inheritance between modules and capture the notion of incremental programming; and (c) modules can be composed through the definition of generic modules, that allow the user to define specific operations of composition.

The knowledge units mentioned above can be briefly described as follows.

a) **Facts**

Facts fix the domain concepts and are stored in a dictionary. Each fact is given a fact identifier, an associate question if the fact is going to be directly asked to the user, and a type, in order to fix its possible values. It is also possible to define relations between facts to give them a logical structure. Four
types of fact can be represented: three of them—enumerate, boolean, and numeric—can be understood as multistate, binary (or two-state), and numeric characters, respectively; the fourth one is the logic type, associated with facts that are usually expressed by the experts with a linguistic certainty term (e.g., that a specimen belongs to the family Polymastiidae is possible, or probable, or sure, or ...). Table 2 presents some examples of different types of fact.

b) Rules

Rules establish associations between facts in the following form: “if premise, then conclusion is certainty value.” A premise is a conjunction of conditions, and a wide set of expression possibilities is provided by MILORD II syntax. After each rule, information about the sources of data or other observations can be stored explicitly. For instance:
TABLE 1. — Example of module definition.

Module Classes =
  Begin
    Inherit Skel
    ...
    Import geo, conti
    Export geo, demos, calc, hexa
    Deductive knowledge
      Dictionary:
        Predicates:
        demos = Name: “Class Demospongiae” Type: Logic
calca = Name: “Class Calcarea” Type: Logic
        ...
      Rules:
        R001 if not(skel/skel) then conclude demos is sure
        R002 if skel/quim = (silica) then conclude demos is possible
        R003 if skel/saxis = (one or four) then conclude demos is sure
        ...
    end deductive
    Control knowledge
      Deductive control:
      M001 if K(not(skel/pres,s)) then inhibit rules Ext/brush
      ...
      M002 if K(demos, int ($x,S,y)) and ge ($x, s) then conclude K(not(hexa),s)
      ...
      M003 if K = (geo, $x,s) and member ($x, (mediatianoan or atlantic)) then conclude K(not(contl),s)

      Structural control:
      M001 if K(hexa, int ($x,S,y)) and ge ($x, p) and K(demos, int ($r,S)) and lt ($r, s) then definitive solution
      end-1
      ...
      M002 if K = (geo, (others) ),s and K(not(contl),s) then definitive solution end-2
      ...
    end control
  end

R004 If numb > 1 and numb < 6 then conclude pauci-tract is sure
  “Paucispicular tract (BOURY-ESNAULT and RUTZLER, in press): a column of aligned megascleres with
  2–5 spicules arranged alongside one another.”

R020 If ext/colour=(yellow or orange-to-red or blue) and ext/cribe
  then conclude Order Poecilosclerida is possible
  “The facts colour and crible belong to module External Characters.”

R025 If skel/S/MI/micros=(chela)
  then conclude Order Poecilosclerida is sure
  “The fact micros belongs to module Microscelares.
  This is a submodule of Spicule, and Spicule is a submodule of Skeleton”.

3. Metarules

Metarules are components that manage the rules and the modules to achieve an efficient search in the
knowledge base and to model the strategies of the expertise performance. In short, they can be consid-
ered structurally similar to rules but their expression capabilities are stronger. Two different goals can be
distinguished in the control. The deductive control can affect rules by different kinds of inhibition. The
structural control can affect the modular architecture by filtering or ordering submodules (Table 3).

TABLE 2. — Examples of several facts of SPONGIA.

<table>
<thead>
<tr>
<th>Name</th>
<th>Identifier</th>
<th>Type</th>
<th>Values</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth upon the substrate</td>
<td>grow</td>
<td>Enumerate</td>
<td>encrusting massy, erect, repent</td>
<td>How is the sponge growing upon the substrate?</td>
</tr>
<tr>
<td>Presence of cribes</td>
<td>cribe</td>
<td>Boolean</td>
<td>yes, no</td>
<td>Is the surface of the sponge covered by cribes (circular perforated areas of some mm of diameter)?</td>
</tr>
<tr>
<td>Uni-, pauci-or multispicular tracts</td>
<td>numb</td>
<td>Numeric</td>
<td>integer</td>
<td>How many spicules are there in the vertical tracts, alongside one another?</td>
</tr>
<tr>
<td>Family Suberitidae</td>
<td>sube</td>
<td>Logic</td>
<td>sure, very possible, quite possible, moderately possible, slightly possible, impossible</td>
<td>NO question as this fact is not consult- eed to the user</td>
</tr>
<tr>
<td>Hierarchical reticulate spongin skeleton</td>
<td>hierar</td>
<td>Logic</td>
<td>sure, very possible, quite possible, possible, moderately possible, slightly possible, impossible</td>
<td>Do you observe a visible difference between two kinds of fibre? (primary = vertical and secondary = horizontal, either in terms of orientation or diameter)</td>
</tr>
</tbody>
</table>

EXPERT SYSTEM FOR SPONGE IDENTIFICATION 277
TABLE 3. — Examples of several meta-rules of SPONGIA.

<table>
<thead>
<tr>
<th>Type of control</th>
<th>Example</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDUCTIVE rules</td>
<td>If no ( \text{pres} ) then inhibit rules ( \text{pres} )</td>
<td>Inhibition of each rule containing the fact \text{presence of spicules} in the premises</td>
</tr>
<tr>
<td>prune rules</td>
<td>If no ( \text{quim} = \text{silica} ) then prune rules ( \text{S}^\text{megas} )</td>
<td>Inhibition of each rule that uses directly or indirectly the fact \text{megascleles} of the module \text{Spicule}</td>
</tr>
<tr>
<td>STRUCTURAL rules</td>
<td>If ( \text{lab} ) is impossible then filter ( \text{suse} )</td>
<td>Local inhibition of the module \text{Suberitidae}</td>
</tr>
<tr>
<td>filter modules</td>
<td>If no ( \text{ske} ) then order ( \text{homs, halis, chondro, dendro} )</td>
<td>Determination of an order or modules consultation: Mod. \text{Homosclerophorida} Mod. \text{Hulsarida} Mod. \text{Chondroidea} Mod. \text{Dendroceratida}</td>
</tr>
<tr>
<td>order modules</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thanks to modularity, MILORD II is especially useful for the incremental construction of knowledge bases (Puyol et al., 1991). In this way, a complex problem can be decomposed into a number of simpler problems, and the initial problem can be solved by combining the solutions to those subproblems. Modularity in MILORD II allows for an incremental construction of an application and makes it possible to cover the knowledge of large domains such as the Porifera in progressive steps. This feature must be highlighted since it is a key point in the knowledge base construction, as well as a functional advantage of MILORD II.

Two particular knowledge representation problems of SPONGIA provide examples of modularity advantages. The first one deals with spicular dimensions. At certain levels of taxonomic identification, spicular dimensions have to be taken into account. This requires a global treatment in the knowledge base construction to avoid unnecessary questions about the particular features of each explored taxon. In this case, a generic module has been implemented that allows for a particular instantiation of different significant intervals of spicular dimensions, without defining them explicitly as separate facts.

The second one is related to ectsosomal differentiation. A knowledge representation problem has arisen concerning Poecilosclerida spiculation. In this case, it was necessary to differentiate the choanosomal and ectsosomal spiculation to discriminate some families. If a kind of spicule existed in both areas, questions about its particular features should be posed (monactine/dialectite; smooth/spined). A generic treatment has been employed to rationally cover the number of possibilities that exist and, while doing so, reduce the number of questions posed to the user.

Another important feature of MILORD II is that it allows for uncertainty in order to handle any imprecision in the experts’ statements curing the construction of the knowledge base. Logic facts and rules are labelled with a linguistic term related to certainty (such as “sure,” “possible,” “moderate possible,” and so on). The certainty values obtained from different rules are combined by a particular logic (which can be locally defined inside each module, if desired) and they are propagated during the inference process following the principles of Many-valued Logic (Rescher, 1969). Moreover, the system is able to manage incomplete information on a given specimen. That is to say, the user is allowed to answer “unknown” to the system questions but, obviously, the less information, the weaker the final decision.

Implementing SPONGIA

Thus far, researchers have failed to agree on who is better prepared to develop an ES. While some authors claim the expert should be the knowledge engineer himself (Taylor, 1985), others argue that the expert knowledge can be biased as a side effect of knowing about techniques or methods of AI (McIntosh, 1986). Since it can take years to develop an ES, the expert will only be able to build an ES himself under special conditions.

From the experience of the author, who is also the knowledge engineer of SPONGIA, it can be said that the time needed to obtain the knowledge from an expert can be substantially shortened when the developer is someone with a solid background (not necessarily an expert) in the same field. Thus, this option is probably the right way to optimise the development of an expert system.

Several types of information have been represented in the knowledge base of SPONGIA by means of modular structures to meet three important requirements.
Taxonomic knowledge: A module hierarchy matches the taxonomic hierarchy. That is to say, the knowledge (deductive and control knowledge) concerning each taxon was formalised in a particular module that took the name of this taxon.

Data gathering: A group of data modules was connected at different levels of the previous structure to collect the observational information by asking questions of the user.

Strategies of classification: Another correlated hierarchy of modules involved only in control tasks was built to model the search strategy.

RESULTS

SPONGIA is a prototype expert system, constructed using the MILORD II programming environment. The purpose of this ES is to help in the identification of marine sponges by asking questions about structural and ecological features of a given sample. Although any kind of information could be represented, some features such as chemical or genetic have not been considered to be sufficiently established in the Porifera to be included at present.

The prototype begins the taxonomic identification at a Class level and follows to more specific taxa. In other words, given a specimen to be identified, the search starts at the highest taxonomic level and progressively follows down looking for the more promising taxa by obtaining the requested information about the sample.

There are two striking features emerging from the preliminary results of this work. The first has to do with the knowledge base as the storage of contents and the second with the interaction of the user with SPONGIA, that is to say, the system interface.

SPONGIA Knowledge Base

The current prototype of SPONGIA consists of 46 modules, 210 facts, 325 rules, and 100 metarules.

A taxonomic restriction has been made since Class Demospongiae is the only one that the expert system deals with. The current program is able to identify the orders and families of Class Demospongiae. In this way, specimens of classes Hexactinellida and Calcarea are detected, but the system is not programmed to continue its identification in lower ranks.

Another constraint was imposed by the geographic scope. The knowledge base has been built on North Atlantic-Mediterranean sponges because this area is well known by the experts involved in this work. If the user wants to identify a specimen collected in another area, the system recommends against doing so unless variability due to geographic situation is not important in the case under study (these variations have not been considered in the knowledge base).

SPONGIA Interface

This is how SPONGIA interacts with the user: Suppose the user ignores any taxonomic rank for the specimen at hand. Then he enters a first module, which initiates the inference process from the top of the taxonomic tree (Fig. 1). The first goal is to obtain a certainty value for each one of the possible classes, by asking suitable questions about the sponge to satisfy any of the rules about each class. Different levels of trace facilities are optional (Fig. 2) to follow the search process. If the system finds a possibility for Class Hexactinellida or Class Calcarea higher than for Class Demospongiae, the inference process will be stopped (because the knowledge base is programmed only to identify demosponges). Only if Class Demospongiae is identified with a high degree of certainty is the next hypothesis to be proved considered: the taxonomic rank order. Usually, several orders can be directly rejected with the information at hand. After this, particular questions will be addressed to the user to determine the level of certainty for the other orders. Afterward, questions will be directed to the family diagnose within the selected order (or orders). The degree of severity in the cluster assignment is fixed in the knowledge base by a certainty threshold. Finally, a list of the more likely taxa is given, each one associated with a certainty value to express its likelihood (Fig. 3). SPONGIA may find one or more possible solutions for a problem case; in other words, the system will not always be sure of its final decision. Obviously, the more accurate the knowledge base is, the more precise the final answer will be.

The accessibility of modules is fixed by the implementation of the module structure, and the user is guided by the system in the right way to move among them, depending on the case at hand. The order of the questions is a consequence of the answers given by the user, and so, it is different in each interaction.

Finally, regarding the scope of utilisation, this expert system is addressed to a user who is able to understand the basic vocabulary of sponge features. A systematisct expert in this field may employ it as a consultant, industry taxonomists can be interested in a specific taxon, and novice biologists sufficiently aware of the knowledge domain can employ it in a more pedagogical interaction.
DISCUSSION

Expert systems have proved to be an interesting approach to identification in that they fulfil some requirements not usually met by other identification tools. For example, they can deal with missing data, uncertainty, or dynamic efficiency.

The primary objective of this study was to construct an ES system and to develop the necessary knowledge base. The problems that had to be resol-
ined in building this system fall into four main categories: coherent conceptualisation, the handling of variability, the incompleteness of the domain theory, and the observational capabilities of the user.

Coherent conceptualisation. An initial problem has been the vocabulary in the Porifera field, both the number of terms available for the same concept and the several meanings attributed to a single term. Although a thesaurus of terms for sponges (BOURVENSAULT and RÜTZLER, in press) was available to us, it did not entirely resolve the problem since the literature—that is to say, the experts—use terms in various ways and thus it was difficult to establish their correct interpretation, especially since the experts could not be interviewed.

Another problem is the controversy surrounding several taxa, sometimes at high taxonomic levels of the classification. This made it necessary to assign taxonomic relationships while designing the knowledge base. Furthermore, there is no clear association between a taxonomic level and the characters employed in its classification. This complicated the control system, since a certain type of knowledge can be required at any level of the classificatory tree and this can affect stylistic issues of interacting with the user.

Variability. Populations of biological organisms are highly variable, both in their external and structural characters. In sponges, structural variability usually relates to differences in latitude but variability of external characters can also be a function of local environmental conditions. So, external characters are desirable as identification characters but they are not reliable. Since they are observed with ease, external characters would be very useful when they could discriminate a taxon. Unfortunately, this seldom happens, and such information must play a supporting role in the system, after other characters (i.e., skeletal) have been explored by the search strategy.

If the user would rely only on external characters, the system can arrive at a weak approximation, but this would not be a reliable basis for identification. Certainly, after obtaining low-confidence information, the ES answer would be cautious and the range of possible answers wide, but the human expert answer would probably be uncertain as well.

Domain theory’s incompleteness. Some problems of knowledge representation are due in part to the lack of information about many biological and cytological features of sponges, and to the continuous discovery of new taxa (e.g., BERGQUIST and FROMONT, 1988; LEVI, 1988). Another difficulty in structuring the knowledge base is that information in the literature derives from different levels of observation (light or electron microscopy). Under light microscopy, for example, a spicule might be considered smooth, whereas examination under electron microscopy might reveal that spines are present. This may lead to confusion in nomenclature in the literature.

User observational capabilities. Every observational process is open to a certain degree of error, as a result of subjective observations, imprecise measurements, and the like. Depending on the user’s skills, the observation will be more or less accurate. Moreover, because of other constraints (e.g., kind of laboratory equipment used, damaged samples), some information may be lost. One way to avoid this problem is to choose a system such as MILORD II, which is able to manage incomplete information. Nevertheless, a graphic aid would be desirable to avoid misunderstandings, but at the moment this is not one of our immediate goals.

CONCLUSIONS

This research has indicated that expert systems technology is a suitable and efficient method of representing the particularly subtle problem solving of the expert systematist.

The practical outcome of this work is the actual construction of an expert system for identifying sponges, which represents a considerable advance since expertise in this field scant and a great complexity is involved in this task. A noteworthy feature of SPONGIA is that it has been designed to deal with identification of anyone of the taxonomic levels. In other words, the identification of a sample go through the taxonomic tree from phylum to species giving an identification at each taxonomic rank (e.g., family or genus). Thus, the system is not restricted to the identification of the lowest taxonomic rank (i.e. species). This way of working requires to base the classificatory knowledge on a given classification theory but, on the other hand, this helps to structure the domain knowledge and to make evident some research problems in the sponge systematics field. Although its present capability is at the family level, the interesting challenge is obviously to expand the tool for the identification of species. Nevertheless, it is important to carefully evaluate SPONGIA’S performance before taking steps in this direction. In any case, once the data structure has been established and the strategies (methods) implemented, domain extension is only a question of time. The ultimate
goal of SPONGIA is to cover the entire phylum which is without doubt a difficult one to achieve for the very extensive knowledge that needs to be acquired to operate the system and to discriminate taxa.

**Future Work**

The work that lies ahead in refining the expert system can be divided into three main functions.

a) Elaborate and initiate a validation project for the prototype of SPONGIA. Since the problem domain has a large extension, it is reasonable to consider the current prototype as a whole and evaluate its prediction capabilities up to the family level.

b) Extend the SPONGIA knowledge base to more specific taxa (genus and species). To date, the domain had to be narrowed in scope in order to attain enough depth. However, both the modularity of MILORD II and the symmetrical distribution of the data structure improve the chances of extending the system easily. At the same time, other techniques that have been explored to tackle this second part (i.e., case-based learning) may be appropriate to cover the species variability.

c) Investigate the kind of knowledge required to perform classification tasks in biology. This will likely lead to an abstraction of domain models, methods, and tasks involved in the identification problems of systematics. The taxonomic classification problems could then be modelled in a MILORD II structure that could accommodate specific knowledge of a concrete domain (Porifera in our case).

**ACKNOWLEDGMENTS**

I am grateful to the Porifera experts who participated in the Seventh International Workshop on Systematics of Atlanto-Mediterranean Sponges (Marseille, France, 1991), whose general comments provided some guidelines for my research in the Porifera domain. Sincere thanks go to Dr. M. J. Uriz who acted as the Porifera expert in the acquisition of the knowledge needed for the application of the program to sponges. The Centre d’Estudis Avançats de Blanes (CSIC) provided technical assistance. This research was supported by a Generalitat de Catalunya fellowship FI91 to the author and by the CICYT TIC91 project TESEU.

**REFERENCES**


