

Design and development of SPONGIA, an expert system for sponges identification*

M. DOMINGO and M.J. URIZ¹

Institut d'Investigació en Intel·ligència Artificial (IIIA), Consejo Superior de Investigaciones Científicas (CSIC), Campus Universitat Autònoma de Barcelona, 08193 Cerdanyola, Barcelona, Spain. e-mail: domingo@iiia.csic.es.

¹Centre d'Estudis Avançats de Blanes (CEAB), Consejo Superior de Investigaciones Científicas (CSIC), Camí de Santa Barbara, 17300 Blanes, Girona, Spain. e-mail: IOSUNE@CEAB.CSIC.ES.

ABSTRACT: This paper reports on the design and development of SPONGIA, an expert system to help in the identification of marine sponges. In this field, taxonomic difficulty is noticeable and systematists are scarce. SPONGIA is directed towards identification at any single taxonomic level starting from the class level. With this feature, the system is not restricted to identify only the species rank and the identification stops at a given taxonomic rank (order, family...) depending on the available data about the sample. The design of SPONGIA takes advantage of an architecture for the construction of expert systems in taxonomic domains. This architecture is based on the MILORD II language for expert systems construction. In this paper, we mainly focus on the conceptualisation and implementation of the sponges domain model which constitutes the main effort in the expert system development. We also refer to the particular classification task and problem-solving method for identification that we have adopted from those provided by the architecture to implement SPONGIA.

Key words: Porifera, Systematics, computer-assisted identification, expert systems, knowledge-based systems.

RESUMEN: DISEÑO Y DESARROLLO DE SPONGIA, UN SISTEMA EXPERTO PARA IDENTIFICACIÓN DE ESPONJAS. – En este artículo se describe el diseño y desarrollo de SPONGIA, un sistema experto para la ayuda a la identificación de esponjas marinas. En este campo, la dificultad taxonómica es notable y los especialistas son escasos. SPONGIA se dirige a la identificación de cualquier nivel taxonómico empezando por el nivel clase. Con esta característica, el sistema experto no se restringe únicamente a la identificación del nivel taxonómico de especie. Además, la identificación progresa hasta un nivel taxonómico concreto (e.g. orden, familia...) según la información a nuestro alcance sobre la muestra. El diseño de SPONGIA se ha beneficiado de una arquitectura para la construcción de sistemas expertos en dominios taxonómicos. Esta arquitectura se basa en el lenguaje MILORD II para la construcción de sistemas expertos. Este artículo se centra especialmente en el trabajo de conceptualización y implementación del modelo del dominio de las esponjas, que constituye el principal esfuerzo en el desarrollo de un sistema experto. También se da cuenta de cómo se describe la tarea de clasificación en el sistema y que método de resolución se ha adoptado de los disponibles en la arquitectura para implementar SPONGIA.

Palabras clave: Porifera, sistemática, identificación asistida por ordenador, sistemas expertos.

INTRODUCTION

Expert systems (ES) are computer programs designed to simulate the decision-making processes that are typically performed by human experts. To

achieve this goal, knowledge from experts has to be captured in order to model the domain theories and the heuristics of an expert in the field. As pointed out by Edwards and Cooley (1993), only the combination of domain theories and expert heuristics can lead to a computational system displaying expert reasoning in a field. This domain dependent knowledge constitutes the so-called knowledge base within the

*Received October 9, 1996. Accepted July 7, 1997.

expert system. Since experts already have the heuristic knowledge (i.e. rules of thumb) they are not the ultimate users of an ES but they can also use it to contrast their hypotheses.

Recent literature reports on expert systems in the field of resource management (e.g. Davis *et al.*, 1989), environmental-advisers (e.g. Reinhardt *et al.*, 1989), molecular biology (e.g. Clark *et al.*, 1993), among others. Although computers have been used in taxonomic identification since the 1970s, few expert systems have been developed in the field of Systematics as reviewed in Edwards and Morse (1995). Other potential uses of artificial intelligence techniques in systematic biology are discussed in Fortuner (1993).

Sponges are among the most important animal taxocoenose in hard substrata of benthic ecosystems. However, they are barely taken into account in the marine ecological or management surveys because they are difficult to identify although they are good indicators of the health state of the marine environment (Muricy, 1991). The discovery of sponges being active producers of chemical bioactive products (cytotoxic, antifungal, antimicrobial, & antiviral) (e.g. Uriz *et al.*, 1992) has drawn the attention of pharmaceutical industry and has made evident the scarcity of experts in sponge taxonomy.

The use of ES techniques for identification is particularly suitable in the Porifera domain due to the highly heuristic knowledge of sponge systematists. An ES dealing with the genus *Hyalonema* Ijima (Porifera:Hexactinellida) is reported in Conruyt *et al.*, (1993). Its identification scope is the level of subgenus (12 taxa). Conversely, SPONGIA was conceived to deal with an entire phylum, Porifera, in the Atlanto-Mediterranean biogeographical province. Specifically, it was directed towards classification at any single taxonomic level starting from the class level. Even if incompleteness or uncertainty around the characters do not allow the identification of a specimen to the species or genus level, SPONGIA is likely to identify it to the order or family level. Regarding the taxonomic tree, SPONGIA can identify the orders and families of the class Demospongiae. Specimens of the classes Hexactinellida and Calcarea are detected, but the system is not programmed to continue their identification in lower ranks. The family Geodiidae of the order Astrophorida has been completed to the level of species. All this amounts to around 100 taxa.

There is no technical limitation for the extension

of SPONGIA to identify the remaining taxa. Actually, the present scope has been limited by time and resource constraints. Anyway, although SPONGIA does not reach the species level in the entire phylum yet, it has the advantage of being functional for any sample of sponge because the identification starts at the top (class) level of the taxonomic tree.

The goal of this paper is to report on the design and implementation of a taxonomic ES on the domain of the phylum Porifera, using an architecture specifically oriented to taxonomic domains. In this paper, we mainly focus on the conceptualisation and implementation of the sponges domain model which constitutes the main effort in the expert system development whereas the taxonomic reasoning methods are fully reported in Domingo (1996). Taking advantage of this methodological approach to ES construction, the development of a taxonomic ES on any other taxonomic domain is reduced to the construction of a domain model of the taxon under study. Thus, we show how the difficult challenge of acquiring and formalising taxonomic expert knowledge in an ES may be achieved by non expert programmers.

SYSTEM AND METHODS

SPONGIA requires a Macintosh computer and at least 6MB RAM. The identification of a sample takes a mean of 8 minutes on a Mac Quadra 950 with 16MB RAM. Seventeen questions are queried on average.

The design and development of SPONGIA takes advantage of an architecture for the construction of ES in taxonomic domains (Domingo, 1995; Domingo and Sierra, 1997). This architecture is based on the MILORD II language for ES construction.

MILORD II

MILORD II is a rule-based language specially suitable to cope with large and complex applications that need to be tackled incrementally. The MILORD II programming environment (Sierra and Godo, 1992; Puyol, 1994) was developed using COMMON LISP and runs on a Macintosh environment. MILORD II is not a commercial shell, but it is being developed in the context of a current research project (IIIA-CSIC). It has been successfully applied to several medical ES (Verdarguer, 1989; Belmonte, 1990).

Several characteristics of MILORD II are useful for developing applications for taxonomic purposes. It is possible to define facts (e.g. parts of an organism or the groups of organisms involved) and relations between facts (e.g. dependence relations or hierarchical relations). Rules of MILORD II are able to handle uncertainty through the mechanism of qualifying the likelihood of the statement expressed in each rule. We have rules of the form “if premise then conclusion is certainty value”, where the certainty follows the Many-valued logic as approximate reasoning model (see below). Semantically, premises stand for events that may happen in the real world and conclusions represent the hypotheses or actions that may be enunciated when a given combination of events takes place.

These characteristics and the following ones overcome the problems of earlier rule-based languages for taxonomic ES construction suggested by some authors (Edwards, *et al.*, 1987).

Modularity

The knowledge contained in a knowledge base of MILORD II is structured in compartments called modules. Modules can be combined through several operations and be structured in hierarchies. Since a module is concerned with a local problem, a hierarchy of modules and submodules (i.e. a knowledge base) represents a structured view over a complex problem. For example, the knowledge base of SPONGIA deals with the phylum Porifera and each module of the hierarchy is concerned with a taxon.

While designing the ES, modules may be considered as empty boxes that are to be filled with contents. In this way, modularity allows an incremental approach to the ES development. Thanks to this incremental approach, any partial knowledge base dealing with a taxon is operational and new knowledge can be progressively added when available.

Approximate Reasoning

Uncertain statements may be modelled in ES by means of an approximate reasoning mechanism. MILORD II handles uncertainty and imprecision in the particular approach of Many-valued logic (Lukasiewicz, 1920). Taking this approximate reasoning model, a rule contains a piece of knowledge which could apply regardless of any other rule. Thus, when building the knowledge base, a rule may be added or

deleted without prejudice for the remaining rules and extending or refining a knowledge base is easy by means of this heuristic mechanism. Uncertainty is expressed with linguistic labels (truth-values) ranging from true (or sure) to false (or impossible) (e.g. in SPONGIA they are: impossible, hardly possible, slightly possible, moderately possible, possible, quite possible, very possible, sure) that correspond to the verbal scale that a human expert uses to weight his statements. The granularity and the terms of this list are decided after considering the opinion of the experts in the field. Linguistic terms are treated as labels and no numerical representation of them is needed. The combination and propagation of uncertainty is performed according to some logic connectives among truth-values, namely conjunction, disjunction and negation.

Control Knowledge

This is the knowledge concerning the management of facts and rules within a module. In MILORD II, the basic control element is the meta-rule. The main goal of the control knowledge is to organise the sequence of actions to be performed by the ES in the course of its execution, taking into account the possible conclusions the system could have arrived at each step. A particular control strategy may be implemented within each module. A powerful control knowledge as the one in MILORD II, with meta-rules with a first order logic (i.e. with variables) makes it possible to model a large domain by using a relatively low number of rules (less than a thousand).

Several essential aspects of a taxonomic ES can be handled by means of the control knowledge: order of the questions, dependence of characters, exceptions, taxonomic procedure.

Architecture for Taxonomic Domains

This architecture assigns a formal meaning, in the taxonomic context, to the knowledge representation constructs of a MILORD II knowledge base in such a way that it can be easily understood why a construct has to be used or what a construct stands for in any taxonomic application. Taking advantage of this “skeleton”, a systematist may develop a taxonomic ES application by filling in the “holes” of the taxonomic knowledge base with the knowledge of the particular domain he/she is interested in.

Three kinds of module with a given functionality are devised in the architecture: the classification task, the classificatory domain model (i.e. the taxonomic model and the structural model) and the classification method/s. Methods are designed to infer the taxon membership at all the different levels of the taxonomy. This original contribution, is not currently found in taxonomic devices, that are usually constrained to the identification of species. Methods are completely reusable modules. The classification task module has to be filled with a list of taxa names. The domain model has a reusable structure which is based in two modular hierarchies standing for the taxonomic model and the structural model, respectively. In this basic structure, rules and meta-rules have to be filled in with the knowledge of the particular taxon. The combination of a domain model, a classification task and a method constitutes a taxonomic knowledge base.

Knowledge Acquisition and Conceptualisation of the domain model of SPONGIA

Knowledge acquisition is the process of knowledge storage and design of a knowledge base. In SPONGIA, only one expert (M.J.U.) was directly involved in the acquisition of the knowledge needed to program the domain model. Thus, the scope of SPONGIA is reduced to the taxa with which this expert was specially familiar. As the architecture is oriented to the incremental construction of knowledge bases, SPONGIA is prepared to easily incorporate the knowledge of other experts in the future.

The first step was the analysis of the classification tasks that the expert usually performed (Domingo, 1993). Moreover, we interviewed a number of experts in sponge systematics and we noticed that they did not show strict agreement in terms of specific tactics and procedures.

In the second step, we began a cycle of interviews to the expert involved in the development of SPONGIA. Each one began with an unstructured part, in which the ES developer asked the expert "spontaneous" questions dealing with a given taxon. The developer was already aware of the basic characteristics of that group (thanks to a previous literature revision) in order to be able to conduct the interview efficiently. The goal of these interviews was to arrive at an accurate description of each taxon. Afterwards, the developer and the expert went over local unclear aspects of the knowledge acquired in the previous session.

Although the main body of knowledge was acquired from a human expert, specialised literature was specially important to determine the intervals of character variability in the definition of taxa. In the literature we found several sources of uncertainty which are discussed in detail in Domingo (1996): uncertainty of the author in the judgement of the observed characters (e.g. Sollas, 1888, p. 90; Boury-Esnault and Lopes, 1985, p. 162); contradictory descriptions (e.g. Boury-Esnault *et al.*, 1994 and Sollas, 1888; with reference to the presence of spherasters in *Geodia barretti*, Bowerbank); low understanding of convergence processes (e.g. Fromont and Bergquist, 1990); low understanding of the characters' variability; and subjective taxa definition.

Conceptualisation of the Taxonomic Model

The conceptualisation of the taxonomic model has basically followed the classification of Lévi (1973) with some modifications. According to Lévi, the class Demospongiae include the following orders: Homosclerophorida (subclass Homoscleromorpha), Astrophorida, Spirophorida, Desmophorida, Hadromerida, Axinellida (subclass Tetractinomorpha), Poecilosclerida, Haplosclerida, Halichondrida, Dictyoceratida, Dendroceratida (subclass Ceractinomorpha) and the subclass Sclerospongia. The families Chondrosidae, Epipolasidae and Samidae are considered *incertae sedis*. The order Desmophorida (ss. Lévi, 1973) contains the families of lithistid sponges with triaenes. Several families of Lithistida without triaenes are mentioned along with the Desmophorida without a definite order assignment.

We have introduced some modifications in this classification according to the recent trends of research in Porifera systematics. The genus *Petrosia* Vosmaer, 1887 of the order Haplosclerida is considered to belong to the order Petrosida (Van Soest, 1980). The family Chondrosidae constitutes the order Chondrosida (Boury-Esnault and Lopes, 1985). The genus *Merlia* Kirkpatrick, 1908 is considered *incertae sedis* since some controversial does exist on whether it must be included in the order Merliida within the subclass Tetractinomorpha (Vacelet and Uriz, 1991) or in the order Poecilosclerida (Hadju, 1995). The family Verongiidae is separated from the order Dictyoceratida and placed in the order Verongida (Bergquist, 1978) and the family Halisarcidae is

separated from the order Dendroceratida as suggested by Vacelet *et al.* (1989) and, according to Bergquist (1996) it constitutes the order Halisarcida. The family Dysideidae is moved from the order Dictyoceratida to the order Dendroceratida on the basis of its aquiferous system (Vacelet *et al.*, 1989). All the sponges considered as lithistid sponges (with or without triaenes) are included for practical reasons in the Lithistida (artificial grouping without taxonomic value).

Conceptualisation of the Structural Model

The structural model covers the characters needed to describe at least the taxa included in the present scope of SPONGIA and the relations among these characters. We listed the characters according to the following groups: external features, skeleton, anatomy, ecology, metrics. Within the skeleton group, we separated the characters according to three types: fibrous, spicular and tracts skeleton. Measurable characters were included in the metrics group. In the future other kinds of data could also be included in the domain model thanks to the modular approach.

The specific terminology of sponges used in the knowledge base is in accordance with a Thesaurus of terms for sponges (Boury-Esnault and Rützler, in press).

A character dependence tree was built among each group of characters. This kind of dependence matched the component/subcomponent arrangement of the characters with the correct order of the questions that are to be put to the user. For example, foreign material (sand or other debris) may be incorporated with the fibres, with the skeleton's matrix or simply with the sponge's surface. Foreign material may core primary or/and secondary fibres to a varying degree. Obviously, in the absence of foreign material the character "location of the foreign material" would be inapplicable, and in the absence of foreign material located in the fibres the character "coring degree of fibres" would be inapplicable.

Indirect dependencies between characters, i.e. causal or incompatibility relations, were also drawn up. These were not as straightforward as the component/subcomponent dependencies. Because of this, some of these relations were incorporated after testing a first prototype of SPONGIA. For instance, the typical choanosomal skeleton of the order Haplosclerida is a reticulated or dendritic

arrangement of spicular fibres or tracts which are always composed of monaxon spicules (character tracts). Thus, we could say that this character depends on the presence of monaxon spicules. This was not a component/subcomponent dependence because the presence of a monaxon spicule complement did not imply the presence of the previously described skeletal arrangement. Nevertheless, if we knew that a specimen had spicules of more than one axis we could infer that the character tracts was false. This is an example of incompatibility between characters.

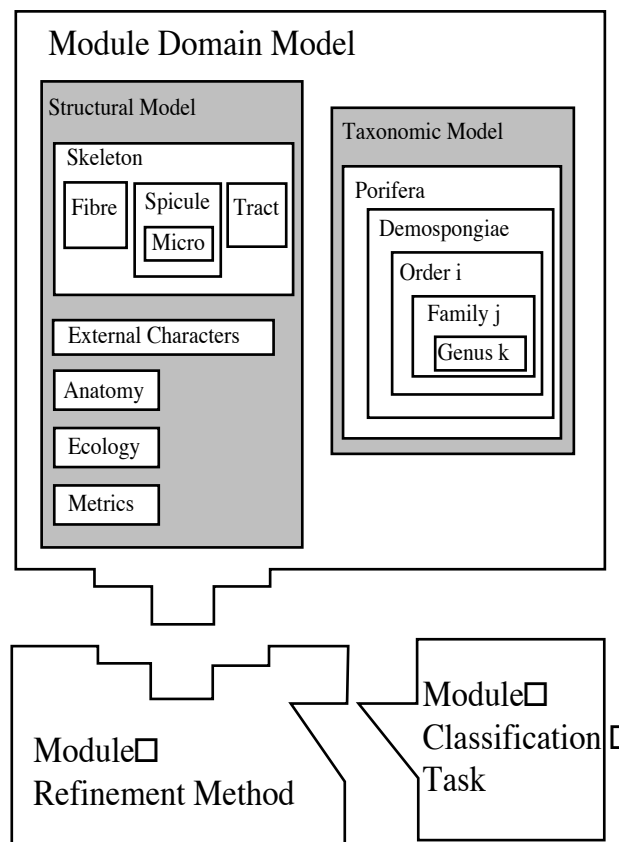


FIG. 1. – Schema of SPONGIA.

Knowledge Representation and Implementation

The knowledge base of SPONGIA consists of: a domain model (a taxonomic model and a structural model), a classification task and a refinement method (Figure 1). The knowledge base

was incrementally constructed, implemented and refined. See Domingo (1995) for the complete knowledge base of SPONGIA.

The rule-based paradigm provides a natural mechanism to represent descriptive and heuristic knowledge about sponges classification. Much of the heuristic knowledge is about combinations of characters (subsets of the prototypical description of a taxon) with particular values or intervals of values which are significant enough to indicate possible

membership of a specimen to a taxon. In this way, rules are used to represent the intra-taxon character's variability.

Implementation of the taxonomic model

Each taxon was given an identifier to avoid excessive typing, e.g. family Polymastiidae was written poly. In SPONGIA, the phylum Porifera was the root taxon and the first taxonomic rank to be

```

Module Demos: Sponges_classification_task=
Begin
  Inherit Hadro
  Inherit Axi
  Inherit Poe
  ...
  Inherit Structures=Structural_model
  Open Structural_model
  Export Taxon
  Deductive knowledge
  Dictionary:
    Predicates: Taxon =Name: "Taxon"
                Type: taxon_type
  Rules:
  ...
  R038 If no(ext/ped) and
        skel/S/size int(megascleres_and_microscleres) and
        skel/S/monax and
        skel/S/monac and
        skel/S/MI/micros int(sigma, rhapsid, toxa)
        then conclude taxon=(poe) is p "F. Biemnidae"
  ...
  R116 If skell/pres and
        ext/consis int (firm, stony) and
        skel/S/diac and
        metrics/max_length > 200 and
        metrics/max_length < 500 and
        eco/asso_sp = (peltodoris_atromaculata)
        then conclude taxon=(petro) is s
        "F. Petrosidae. Petrosia ficiformis"
  ...
  End deductive
  Control knowledge
  Truth threshold: moderately possible
  Deductive control:
  M001 If K(not(Skell/pres),int(s,s)) then
        conclude K(not=(taxon, hadro),int(s,s))
        conclude K(not=(taxon, axi),int(s,s))
        conclude K(not=(taxon, poe),int(s,s))
        conclude K(not=(taxon, astro),int(s,s))
        conclude K(not=(taxon, spiro),int(s,s))
        conclude K(not=(taxon, merli),int(s,s))
        conclude K(not=(taxon, petro),int(s,s))
  ...
  M025 If set_of_instances($x,K(=(skel/S/MI/micros,$x),int(s,s)),$values)
        and intersection($values,(clavidisc)) then
        conclude K(=(taxon, merli),int(s,s))
  ...
  End control
End

```

FIG. 2. – Example of module from the taxonomic model of SPONGIA. Normal style corresponds to the skeleton provided by the architecture and italic style corresponds to the specific implementation of SPONGIA.

identified was the class level. The taxonomic tree was represented by means of a module/submodules hierarchy, in which each module was responsible for the identification of the immediately lower taxonomic level. Each module was given the name of the taxon that it represented. For example, the root module was Module Classes that represented the knowledge concerning the class level. Module Demos represented the class Demospongiae and output the identification of the orders of this class (Figure 2). A submodule of the module Demos was Module Hadro that output the families of the order Hadromerida, and so on.

The implementation of Module Demos is here explained step by step as an example of the taxonomic model. In Module Demos we wrote a submodule corresponding to each order i.e. Inherit Hadro, Inherit Axi, and so on. In this way, we linked the modules representing the families of each order to the module representing the orders of the class Demospongiae. Another submodule declaration, Open Structural model, linked Module Demos to the root module of the structural model i.e. Module Structural model. Thus, data concerning any character may be sought from module Demos when it is needed to satisfy the premise of a rule. Finally, the Export interface specified the output of this module, that is, a taxon name as specified in the Sponges Classification Task module.

In the Deductive knowledge, the dictionary contained a single fact Taxon which is instantiated following the specification of Module Sponges Classification Task, as shown below. The truth threshold was fixed in the certainty value moderately possible. Thus, when an order was concluded in module Demos with a certainty value of possible or higher, the result was considered meaningful enough to investigate lower taxa. Contrarily, if one order was labelled with a certainty lower than or equal to moderately possible, the families of that order were directly pruned.

Rules and meta-rules make a deduction about the order identification within the module. In Module Demos, 138 rules and 26 meta-rules were written to represent the membership to every order of the class Demospongiae.

We represented the incompleteness and uncertainty, involved in the sponges' taxonomic theory, by means of the construction of a wide set of rules and meta-rules, carefully labelled with certainty values. Some of them modelled complex reasoning mechanisms exhibited by systematists such as

TABLE 1. – Examples of rule syntax and semantics using MILORD II. We write Y (yes) when the answer satisfies the premise and N (no) when the premise is not satisfied.

	POSSIBLE USER ANSWER			
	(oxea)	(oxea, style)	(style)	(oxea, tylostyle)
RULE PREMISES				
megas int (oxea, style)	Y	Y	Y	Y
megas int (oxea)	Y	Y	N	Y
no (megas int (oxea))	N	N	Y	N
megas int (oxea) and megas int (style)	N	Y	N	N
megas int (oxea) and no (megas int (style))	Y	N	N	Y
megas=(oxea)	Y	N	N	N
megas=(oxea,style)	N	Y	N	N

default reasoning or non-monotonic reasoning (Domingo, 1996). At the same time, such a rule base can cope with uncertain and incomplete answers from a user consulting the system.

Table 1 illustrates the use of the MILORD II operations disjunction (int), conjunction (and), equality (=) and negation (no). Examples of rule's premises are presented using the character Type of megascleres, megas for short.

In SPONGIA the user may label his answer with a fuzzy certainty value when he is not sure of the character observation. For example, a valid answer to the fact megas could be "oxea is possible and style is very possible". A fuzzy answer with a certainty value higher than the truth threshold would satisfy the premise anyway. Thus, Table I would also apply if the possible answers were qualified with a certainty value higher than the truth threshold.

We illustrate the knowledge representation in SPONGIA by briefly explaining some examples of rules and meta-rules of Figure 2:

R038. When the spicular complement consists of megascleres of monaxon and monactine types, and microscleres of sigma, raphid or toxa type, and the sponge is not pedunculated, the specimen belongs to the order Poecilosclerida. This applies to the members of the family Biemnidae.

The first premise is an example of negation applied to a two-state character. This premise is introduced to prevent this rule from fitting in with a species of the genus *Rhizaxinella* (order Hadromerida), that also exhibits monaxon and monactine megascleres, raphids as microscleres and is typically pedunculated.

Note that the path to arrive at a character may be of different length. In the last premise the fact micros belongs to module MI. This rule was written in module Demos, and from module Demos to module MI, the path goes over three modules: Skel (a submodule of Demos), S=Spicule (a submodule of Skel), and MI=Micros (a submodule of Spicule).

In the last premise, we represent a variety of combinations of states in an multi-state character through a disjunction. This operation (int) must be understood in the context of the Set Theory. Thus, this premise will be satisfied by a specimen displaying any possible subset of the set (sigma, raphid, toxa).

R116. When the spicular complement consists of diactinal megascleres with a maximum length between 200 and 500 microns, the sponge's consistency is firm or stony, and a *Peltodoris atromaculata* has been found over the specimen, then it belongs definitely to the order Petrosida. In this rule the inequality operators “<” and “>” are used to model a numeric interval for the character max_length, i.e. Maximal length of the spicular complement. If desired, we could also use the operators “=”, “≤”, or “≥”, dealing with numeric characters.

Moreover, this rule provides examples of the syntax used to represent two-state characters (e.g. Presence of spicules, namely pres), and fuzzy (logic) characters (e.g. Diactinal spicule, namely diac) which may be answered with a fuzzy certainty value.

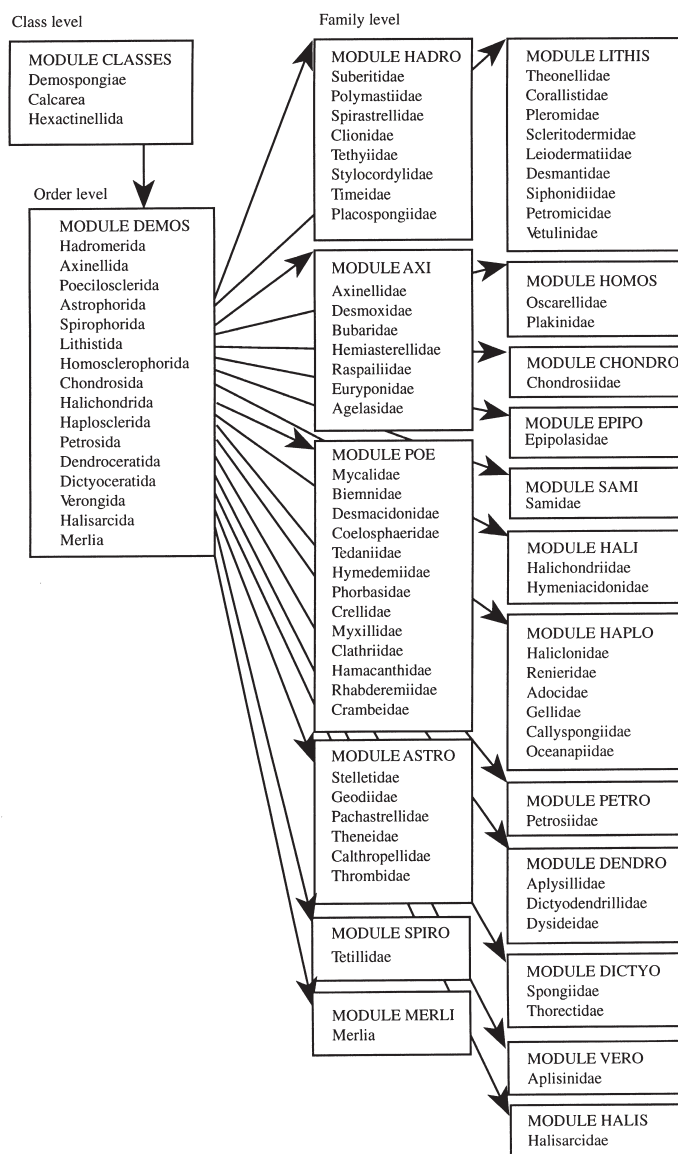


FIG. 3. – Modules of the taxonomic model of SPONGIA down to the level of family. Boxes correspond to modules. Arrows between boxes indicate module/submodule relations.

M001. When no spicules are present within the specimen membership to the following taxa must be discarded: Hadromerida, Axinellida, Poecilosclerida, Astrophorida, Spirophorida, Lithistida, Merliida, Haplosclerida, Halichondrida, Petrosida. Having applied this meta-rule, the system will focus on the Keratosa and the orders that contain species without a skeleton.

M025. When the spicular complement includes microscleres type clavidisc the specimen belongs to the order Merliida.

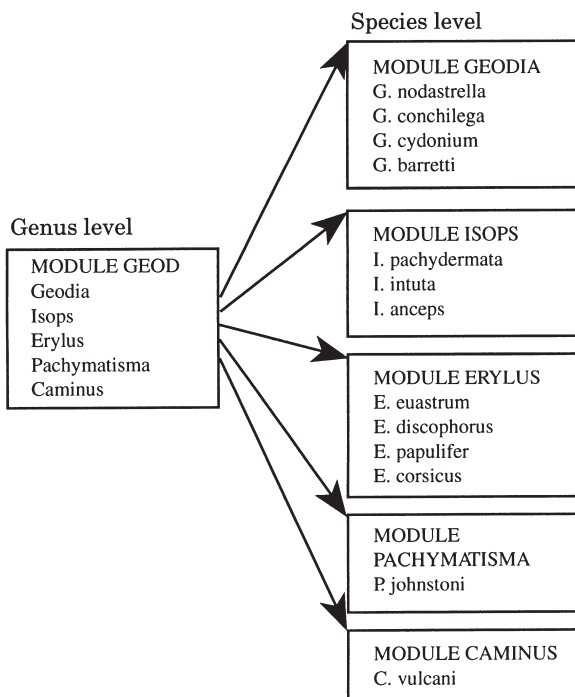


FIG. 4. – Modules of the taxonomic model of the family Geodiidae down to the level of species. Boxes correspond to modules. Arrows between boxes indicate module/submodule relations.

This meta-rule is not used to discard taxa, unlike the previous meta-rule, but to ensure membership to a taxon. This kind of statement is not usually available in every taxon of a taxonomic domain.

In Figure 3 and Figure 4, a scheme of the modules that implement the taxonomy of Porifera has been drawn up. In these figures, each box corresponds to a piece of code similar to the example of taxonomic module (Figure 2). The taxa names written within each box are the taxa concluded by the module's rules and meta-rules. Arrows between boxes indicate module/submodule relations. Each arrow has its origin in a module and points to the submodules of this module.

Implementation of the Structural Model

As in the case of the taxonomic model, each character was given a short identifier. The hierarchy of modules that represented the structural model began with Module Structural Model as root module. This module is linked to the taxonomic model and is accessible from each one of the taxonomic modules via the submodule declaration Open Structural Model (Figure 2). The module Structural Model contains a number of submodules that represent each group of characters, namely Module Ext, Module Skel, Module Ana, Module Eco and Module Metrics and some of them have other submodules, and so on.

The implementation of Module Skel, is detailed here as an example of implementation of a module of the structural model. In Module Skel (Figure 5), we have declared three submodules dealing with different skeletal arrangements. The submodule declaration (e.g. Module F=FIBRE) is a link between two modules that allow the father module (in this case Module Skel) to use the information of its submodules i.e. Fibre, Spicule, Tracts. The Import interface specified those characters that the system's user would have to be asked. Similarly, the Export interface mentioned those characters that the module could give as an answer to the outside (either to the user or to other modules). Note that the character skel was not importable (i.e. not available by querying the user), instead it was inferred by a rule (R001) from this module.

In the Deductive knowledge, each predicate stood for characters. A question is formulated in the definition of importable predicates. In the predicate quim, the relation needs skel indicated that before inquiring the character quim, the character skel should always be found out. Similarly, in the character tracts, the relation needs S/axis indicated that before inquiring this character, the fact axis should always be found out. In this case, axis did not belong to Module Skel but to Module S=Spicule as indicated in the relation by the path S/axis.

Meta-rules were the components that linked each dependent character to the values of the previous characters on which it depended. M003 implemented the relationship explained above between the facts tracts and axis. That is, if we knew that a specimen had spicules of more than one axis we could infer that the fact tracts was false without asking the user about this character.

```

Module Skel =
  Begin
    Module F = FIBRE
    Module S = SPICULE
    Module T = TRACTS
    Import pres, fibre, tracts, quim
    Export skel, fibre, pres, tracts, quim
    Deductive knowledge
    Dictionary:
      Predicates:
        skel = Name: "Presence of skeleton"
              Type: Logic
        pres = Name: "Presence of spicules"
              Question: "Can you observe the presence of a
              complement of spicules in this sponge?
              (foreign occasional spicules must be ignored)"
              Type: Boolean
        fibre = Name: "Spongin fibres present in the skeleton"
              Question: "Can you observe a spongin fibre
              architecture in the skeleton? (with or without
              indigenous spicules or foreign material)"
              Type: Boolean
        tracts = Name: "Spicular fibres or tracts"
              Question: "Is the choanosomal skeleton organised as
              a reticulated or dendritic arrangement of tracts (uni
              or multispicular columns of aligned megascleres with
              or without surrounding spongin)?"
              Type: Boolean
              Relation: needs S/axis
        quim = Name: "Chemical composition"
              Question: "Of which material is the skeleton made?"
              Type: (calcium_carbonate, silica, organic)
              Relation: needs skel

      Rules:
        R001 If no(pres) and no(fibre) then conclude no(skel) is s
        End deductive
        Control knowledge
        Deductive control:
        M001 If K(not(skel),int(s,s)) then conclude K(=(quim, none ),int(s,s))
        M002 If K(not(pres),int(s,s)) then conclude K(not(tracts),int(s,s))
        M003 If K(=(S/axis, $x),int(s,s)) and member($x,(three or four))
              then conclude K(not(tracts),int(s,s))
        Structural control:
        M001 If K(not(pres),int(s,s)) and K(not(fibre),int(s,s)) then filter T
        M002 If K(not(pres),int(s,s)) then filter S
        M003 If K(not(fibre),int(s,s)) then filter F
        M004 If K(not(tracts),int(s,s)) then filter T
        End control
  End

```

FIG. 5. – Example of module from the structural model of SPONGIA. Normal style corresponds to the skeleton provided by the architecture and italic style corresponds to the specific implementation of SPONGIA.

In the structural model, rules concluded characters that could be deduced from other observed characters. Thus, the character *skel* could be inferred from the characters *pres* and *fibre* in R001. On the other hand, the meta-rules of the

structural control (M001 to M004) prune the hierarchy of modules. For example, Module *S=Spicule* is inhibited by the meta-rule M002, by the instruction filter *S*, when the character *Presence of spicules* is false.

Implementation of the Classification Task for SPONGIA

The classification task provided by the architecture corresponds to what has been called simple classification (Clancey, 1985), i.e. identifying an unknown object as a member of a known class of objects by matching observations of the unknown individual against features of known classes. Furthermore, we want this goal to be achieved at every taxonomic rank and even when the collection of characters involves incompleteness or uncertainty.

This task is instantiated here for the application SPONGIA by writing the names of the taxa that were included in SPONGIA within the Classification Task module.

A Classification method for SPONGIA

Among the problem-solving methods provided by the architecture, we have chosen the Refinement method with an additional constraint of consistency preservation to build the application SPONGIA. Following this method, the identification procedure begins with the identification of the class level and goes down to the level of family or of species (depending on the part of the taxonomy). The queries the user is asked range progressively from very general to more specific characters. In this way, the user is guided in the observation of the characters. Moreover, the system is able to detect inconsistency between the answers given to different questions because the diagnostic characters are verified by means of different questions if the answers are not given with absolute certainty.

The consistency preservation constraint constitutes a restriction over results of the inference process. In each taxonomic level, when a taxon is concluded with the maximum certainty value (i.e. sure in SPONGIA) any other taxon which is concluded with a lower certainty value than sure is eliminated from the answer presented to the user.

DISCUSSION

The design of the domain model is very intuitive for systematists because it follows two well-known hierarchies: the taxonomic tree in the taxonomic model and a component/subcomponent tree in the structural model. A classification task module and

several classification method modules are available in the architecture. The implementation of the classification task is reduced to fill the gaps in the classification task module with the particular taxonomic names of our domain. Since problem-solving methods are generic knowledge, the pieces of code implementing methods are directly reused in any taxonomic ES. The problem-solving method in SPONGIA provided the identification of every taxonomic rank. Thus, even when the user is uncertain in his answers, SPONGIA can give an identification at least at high taxonomic ranks. Moreover, since the system begins the identification from the class level the user only has to identify previously whether the sample is a sponge or not.

The major stumbling block is the development of the domain model. Variability, uncertainty and incompleteness of the taxonomic knowledge in the field of sponges have been modelled in rules and meta-rules of MILORD II using the operators: conjunction, disjunction, equality, negation and inequality. Complex reasoning processes exhibited by systematists such as default reasoning and non-monotonic reasoning are modelled through the combination of rules and meta-rules.

Taking advantage of AI techniques, we have faced up to difficulties in knowledge modelling that would have to have been extensively simplified if we had used any other technological approach to build a computerised identification tool. In our opinion, the only way to build a consistent model on the taxonomy of many taxonomic domains, in particular the Porifera, is to acquire the knowledge directly from a human expert, who can balance and unify the meaning, the importance and the accuracy of data from existing sources (journals, monographs, text books...) according to his heuristics. It may be argued that this is not a model of expertise but a model of the expertise of an expert. In fact, we agree with this view and this is not a problem if the model is a recognised expert. As is the case in ecological modelling, if we manage to represent the complexity of an entity in the world that we wish to model on (e.g. ecosystem, expert knowledge) we are likely to succeed in simulating its behaviour.

When we began the construction of SPONGIA we were aware of the taxonomic problems associated with the phylum Porifera. In the course of its development, we discovered the extent to which Porifera was a problematic domain concerning

classification. The simple work of character definition in the domain model turned out to be a non-trivial task. A Thesaurus of terms for sponges that was to contain an important consensus among the European experts in sponge systematics was in the pipeline and its preliminary versions were available to us. However, the current vocabulary of “our” expert was not always standardised according to the previous consensus.

A second problem we faced was the continuous evolution of the established taxonomic theory. In Porifera, this theory is intensively revised when new findings do not fit current observations. Most of the current changes in the sponge taxonomy are at the genus or species level but even changes in higher levels have also been proposed (e.g. Vacelet *et al.*, 1989). In the domain model construction, we were obliged to compromise the current views.

Indirectly, this situation was a suitable challenge to test the feasibility of introducing local changes in the domain model. We experienced that, thanks to the modularity of the knowledge base, this entailed little work and no prejudice for the rest of the knowledge. So, with this experience we guaranteed that a maintenance survey (i.e. a periodical up-dating of the domain model) would be feasible.

After the development of SPONGIA, we consider that the MILORD II language was rich enough to cope with the requirements for taxonomic ES construction. Moreover, we benefit from an uncommon situation; since MILORD II was an experimental language (instead of a commercial shell) we could ask for particular facilities of the language that were programmed ad hoc. This feedback between the language developers and the application developer led to research results in both fields (Puyol, 1994). On the other hand, the validation of SPONGIA (Domingo, 1995) demonstrates that it obtains similar quality results to the experts in Porifera systematics.

ACKNOWLEDGEMENTS

This research has been funded by the projects CICYT TESEU-TIC91-0430 and MARS95-1764, and the Government of Catalonia: fellowship (FI/91-193) to M. Domingo and grant 1995SGR 00443.

REFERENCES

- Belmonte, M. A. – 1990. *RENOIR: Un sistema experto para la Ayuda en el diagnóstico de Colagenosis y Artropatías Inflammatorias*. Ph.D. Thesis, Universitat Autònoma de Barcelona, Barcelona, 297 pp.
- Bergquist, P. R. – 1978. *Sponges*. Hutchinson, London, 268 pp.
- Bergquist, P. R. – 1996. The marine fauna of New Zealand: Porifera: Demospongiae. Part 5 Dendroceratida and Halisarcida. *New Zealand Oceanographic Institute Memoir*, 107: 1-53.
- Boury-Esnault, N. and M. T. Lopes. – 1985. Les Démosponges littorales de l'archipel des Açores. *Ann. Inst. Océanogr., Paris*, 61: 149-225.
- Boury-Esnault, N., and K. Rützler (eds). – 1997. *Thesaurus of Terms for Sponges*. Smithsonian Institution Press, Washington, D.C. (In press).
- Boury-Esnault, N., Pansini, M. and M. J. Uriz. – 1994. Spongiaires bathyaux de la mer d'Alboran et du golfe ibéro-marocain. *Mém. Mus. nat. Hist. Nat.*, 160: 1-174.
- Clancey, W. J. – 1985. Heuristic classification. *Artif. Intell.*, 27: 289-350.
- Clark, D. A., Rawlings, C. J. and S. Parsons. – 1993. Current Research in Artificial Intelligence and Molecular Biology. *The Knowledge Engineering Review*, 8: 349-357.
- Conruyt, N., Manago, M., Le Renard, J. and C. Lévi. 1993. – Une méthode d'acquisition de connaissances pour la classification et l'identification d'objets biologiques. Application au domaine des éponges marines. *Proc. 13th International Symposium on expert systems and Applications, Avignon, France*, 1: 485-495.
- Davis, J. R., Nanninga, P. M., Hoare, J. R. L. and A. J. Press. – 1989. Transferring scientific knowledge to natural resource managers using artificial intelligence concepts. *Ecol. Model.*, 46: 73-90.
- Domingo, M. – 1993. *Towards a knowledge level analysis of classification in biological domains*. In: Piera-Carreté, N. and Singh, M. (eds). *Qualitative reasoning and decision technologies*, pp.535-544, CIMNE, Barcelona.
- Domingo M. – 1995. *An expert system architecture for taxonomic domains. An application in Porifera: the development of SPONGIA*. PhD Thesis, Universitat de Barcelona, Barcelona, 290 pp.
- Domingo, M. – 1996. *Models of practical taxonomic reasoning in knowledge-based systems: an application to Porifera*. In: Willenz, P.(ed). *Recent Advances in sponge biodiversity inventory and documentation*, Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Biologie Vol. 66 suppl: 27-35.
- Domingo, M. and Sierra, C. – 1997. A Knowledge Level Analysis of Taxonomic Domains. *International Journal of Intelligent Systems*, 12:105-135.
- Edwards, M., Morse, D. R. and A. H. Fielding. – 1987. Expert systems: frames, rules or logic for species identification? *CABIOS*, 3: 1-7.
- Edwards, M. and R. E. Cooley. – 1993. Expertise in expert systems: knowledge acquisition for biological expert systems. *CABIOS*, 9: 657-665.
- Edwards, M. and D. R. Morse. – 1995. The potential for computer-aided identification in biodiversity research. *TREE*, 10: 153-158.
- Fortuner, R., (ed). – 1993. *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. The Johns Hopkins University Press. Baltimore. London, 560 pp.
- Fromont, J. P. and P. R. Bergquist. – 1990. *Structural Characters and Their Use in Sponge Taxonomy: When is a sigma not a sigma?* In: Rützler, K.(ed). *New perspectives in sponge biology*, pp.273-278, Smithsonian Institution Press, Washington D. C.
- Hadju, E. – 1995. *Macroevolutionary patterns within the Demosponge Order Poecilosclerida. Phylogeny of the marine genus Mycale and an integrated approach to biogeography of the seas*. PhD, University of Amsterdam, 173 pp.
- Lévi, C. – 1973. *Systématique de la classe Démospongiaria (Demosponges)*. In: Grassé P.P. (ed). *Traité de Zoologie. III Spongiaires* pp.577-631, Masson et Cie, Paris.
- Lukasiewicz, J. – 1920. On 3-valued logic. *Przegląd Filozoficzny*, 23: 189-205.
- Muricy, G. – 1991. Structure des peuplements de spongiaires autour de l'égout de Cortiou (Marseille, France). *Vie Milieu*, 41: 205-221.

- Puyol, J. – 1994. *Modularization, Uncertainty, Reflective Control and Deduction by Specialization in MILORD II, a Language for Knowledge-Based Systems*. PhD Thesis, Universitat Autònoma Barcelona, Barcelona, 230 pp.
- Reinhardt, E., Wright A. H. and D. H. Jackson. – 1989. An advisory expert system for designing fire prescriptions. *Ecol. Model.*, 46: 121-133.
- Sierra, C. and LL. Godo. – 1992. – *Modularity, uncertainty and reflection in Milord II*. pp.255-260. Proc. IEEE International Conference on Systems, Man and Cybernetics, Chicago.
- Sollas, W. J. – 1888. Report on the Tetractinellida collected by H.M.S. "Challenger" during the years 1873-1876. *The Voyage of H.M.S. "Challenger"*, 25: 1-458.
- Uriz M.J., Martín D. and D. Rosell. – 1992. Relationships of biological and taxonomic characteristics to chemically mediated bioactivity in Mediterranean littoral sponges. *Mar. Biol.*, 113: 287-297.
- Vacelet, J., Boury-Esnault, N., de Vos, L. and C. Donadey. – 1989. Comparative Study of the Choanosome of Porifera: II. The Keratose Sponges. *J. Morphol.*, 201: 119-129.
- Vacelet, J. and M. J. Uriz. – 1991. – *Deficient Spiculation in a New Species of Merlia (Merliida, Demospongiae) from the Balearic Islands*. In: Reitner, J. and Keupp, H.(eds). *Fossil and Recent Sponges*, pp.170-178, Springer-Verlag, Berlin, Heidelberg.
- Van Soest, R. W. M. – 1980. Marine sponges from Curaçao and other Caribbean localities. Part II. Haplosclerida. *Stud. Fauna Curaçao Caribb. Islands*, 62: 1-174.
- Verdaguer, A. – 1989. *PNEUMON-IA: desenvolupament d'un sistema expert d'ajuda al diagnòstic mèdic*. Ph.D. Thesis. Universitat Autònoma de Barcelona, Barcelona, 400 pp.

Scient. ed: J. Lleont

