Knowledge engineering in environmental sciences: application to Systematics of corals of the Mascarene Archipelago

N. Conruyt[1] and **D. Grosser**[1] and **Y. Geynet**[1]
and **G. Faure**[2] and **M. Pichon**[3] and **M. Guillaume**[4] and **N. Gravier-Bonnet**[5]

Abstract. Systematics, the scientific discipline that deals with listing, describing, naming, classifying and identifying living organisms is at a turning point. Expertise is becoming extremely rare. For future biodiversity studies relying on species identification, environmental technicians will only be left with monographic descriptions and collections in museums. With the emergence of knowledge management on the Internet, it is possible to enhance the use of systematician's expertise, by providing them with collaborative tools to widely manage, share and transmit their knowledge. Knowledge engineering in Systematics means to revise descriptions of specimens and to bring them alive on the web. We have designed an Iterative Knowledge Base System ($\mathcal{IKBS}$) for achieving these goals. It applies the scientific method in biology (conjecture and test) with a natural process of knowledge management: 1/ acquisition of a descriptive model and related descriptions from collection specimens, 2/ processing of this knowledge for classification and identification, 3/ experimentation and validation. The product of such a tool is a collaborative knowledge base of a domain, that can evolve (by updating the knowledge) and be connected to distributed databases (bibliographic, photographic, geographic, taxonomic, etc.) that will yield information on species after the identification process of a new specimen. The $\mathcal{IKBS}$ system is presented here as a life science application facilitating the identification of coral specimens of the family *Pocilloporidae*.

keywords : Knowledge Management, Knowledge Base System, Case Base Reasoning, Inductive Learning, Corals Systematics,

1 INTRODUCTION

Today around the world, scientific databases are increasingly delivered on CD-ROM or through Internet (e.g. World Biodiversity database from ETI in Netherlands, Reefbase and Fishbase from ICLARM in the Philippines, Hawaii Biological Survey databases, Coral Id at AIMS and Description Language for Taxonomy at CSIRO in Australia, etc.). These applications are taxonomic and bio-geographic information systems with some identification keys for biologists (students, amateurs) and professionals (environment, tourism). In fact, they reproduce mostly electronically what already exists in books (i.e. textual descriptions, identification with diagnostic characters). This approach is interesting when the taxa are well known and stable, but it is not sufficient when the knowledge of groups evolves rapidly, which is particularly the case in the marine environment (corals, hydroids, sponges, etc.).

In such domains, products for knowledge management in Systematics are also needed, with a new methodology of knowledge extraction. This method is based on the re-examination of specimens in various collection in order to get more robust classifications (definition of the taxa) and identifications. In fact, the description of specimens is the key point for engineering Systematics: this descriptive information in the application can always be retrieved in the future and compared again with the museum sample collections. For young systematicians, this specimen-oriented approach brings more robustness to the learning process than working with old monographs based on conceptual species descriptions. Moreover, end-users of such a system (e.g. environmental technicians) can directly compare a newly collected specimen with the description of other specimens in collections.

For engineering Systematics, we have developed a type of knowledge base that supports the above methodology. The tool that generates such applications is called $\mathcal{IKBS}$ (*Iterative Knowledge Base System*)[2]. $\mathcal{IKBS}$ is a knowledge management system written in Java language, which is available on the Internet [17]. This tool was co-designed with specialists and end-users for 15 years in different domains such as plant pathology diagnosis [12] and computer aided Systematics [15]. For making descriptions, classifications and identifications, our knowledge bases rely not only on observed things (the database of specimen descriptions) but also on observable things (the knowledge of a descriptive model of the domain).

Such knowledge bases are built by following the natural learning process of an expert, i.e. the inductive or experimental approach (conjecture and test): experts acquire their experience by observing, interpreting and learning from a huge number of specimens in collections and in the field. Our aim is to reproduce explicitly this methodology of knowledge extraction. Computer scientists can give taxonomists an interactive tool allowing them to: 1) define structured objects of the observable domain, 2) describe examples (or cases) of specimens in collections, 3) process this descriptive knowledge with classifications, decision trees, comparison of cases, and synthesis of descriptions, and 4) make experiments with end-users, i.e. carry out identifications of new specimens. Thus, experts themselves

[1] IREMIA, Université de la Réunion, 97715 Saint Denis, BP 7151, France. e-mail : ikbs@univ-reunion.fr
[2] ISIM, Université de Montpellier II, Pl. E. Bataillon, 34095 Montpellier, France
[3] EPHE, Université de Perpignan, Av. de Villeneuve, 66025 Perpignan Cedex, France
[4] LBIMM, Museum National d'Histoire Naturelle, 55 rue de Buffon, 75005 Paris, France
[5] ECOMAR, Université de la Réunion, France

can validate the robustness of the predefined descriptive model and descriptions. The iterative methodology which is presented in this paper enables specialists to take into account the evolution of their observable and observed knowledge and their ability to transmit this knowledge to other biologists.

As an example, a knowledge base of the scleractinian corals of the Mascarene Archipelago (SW Indian Ocean) is being developed using 3000 specimens from the collections made by [4] belonging to 185 species of 58 genera. To date 5 families (*Pocilloporidae*, *Fungiidae*, *Poritidae*, *Siderastreidae*, *Astrocoeniidae*) covering 18 genera and 52 species have been entered (or are in the process of being entered) in the knowledge base. This represents more than 200 specimen descriptions related to 5 descriptive models of the families (16 families are described in the Mascarene Archipelago). In this paper, we will present the work carried out on the *Pocilloporidae* family, which serves as a model for the development of any other knowledge base in Systematics.

2 METHODOLOGY

Knowledge is extracted with $\mathcal{IKBS}$ by a cyclical process, divided into three parts (see Fig. 1):

1. Knowledge acquisition:
   - Acquire a descriptive model (domain knowledge or observable facts),
   - Acquire a multimedia questionnaire on the web,
   - Acquire descriptions (observed facts or cases),
2. Knowledge processing:
   - Generate classification rules with decision tree induction,
   - Identify new observations (unknown specimens) with case-based reasoning,
3. Knowledge validation and refinement:
   - Verify the origin of misidentifications by analyzing differences of interpretation between the expert and the users of the knowledge base,
   - Iterate on the definition of the descriptive model (characters), update old cases.

For experts in biology (taxonomists), this approach is well suited to the natural process of their knowledge acquisition, conjecture and test [10]:

1. Observe and familiarize oneself,
2. Represent observations, i.e. make descriptions,
3. Build hypotheses from descriptions (pre-classified), i.e. generate identification keys,
4. Test and use them with new observations, i.e. identify new specimens,
5. Refine the initial knowledge (new characters, cases and classifications).

The last point of the method is fundamental because it is hard for experts to define the best representation of reality at once in a descriptive model. The challenge is to acquire the best character definitions and illustrations leading to interpretations of observations understood by anyone consulting the knowledge base. After the identification of a new specimen, the end-user can be connected to distributed databases (bibliographic, photographic, geographic, taxonomic, etc.) that yield contextual information on species. We are currently evaluating an object-oriented Systematics information system (OSIS project) to develop an integrated knowledge base and database on the Systematics of the Mascarene coral reef ecosystem [5].

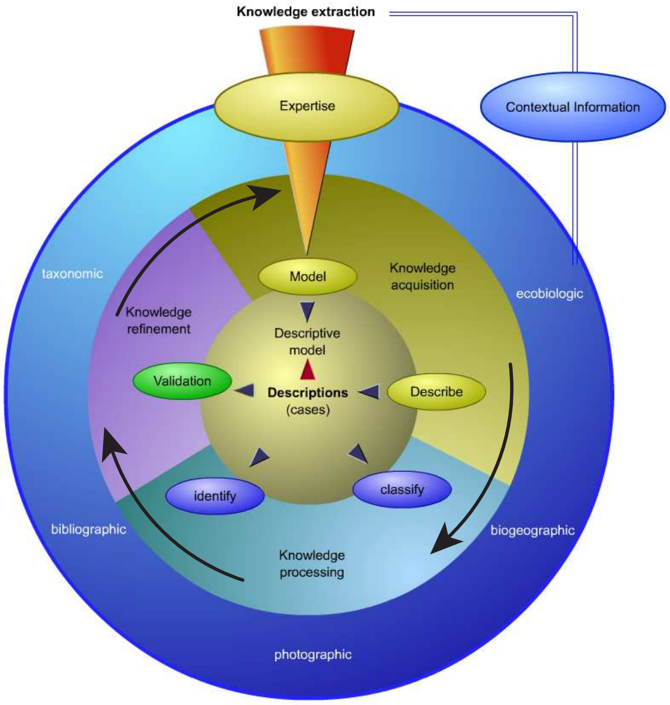

Figure 1. Knowledge management cycle of $\mathcal{IKBS}$

3 KNOWLEDGE ACQUISITION

The knowledge acquisition part of the methodology is very important in order to acquire a case base with good quality descriptions, i.e. that are well structured in different dimensions with all the required information, with characters, illustrations and comments that are easily comprehensible for other biologists.

3.1 The descriptive model

The descriptive model (Fig. 2) represents all the observable characteristics (objects, attributes and values) pertaining to individuals belonging to a particular domain. It is organized in a structured scheme, the name of the domain being at the root of a description tree. Each node of the tree is an object (a component of the individual) defined by a list of attributes with their respective possible values. Designing a descriptive model is essentially an expert task.

To help them, we have set up logical rules for case description covering: decomposition, viewpoint, iteration, specialization, and contextual conditions [6]. These rules were constructed from the analysis of the process followed by the experts to create monographs of organisms or diseases.

To serve as an example in coral Systematics, we present the descriptive model of the family *Pocilloporidae* (see Fig. 2). The expert has defined 51 objects and 120 attributes. With them, biologists are

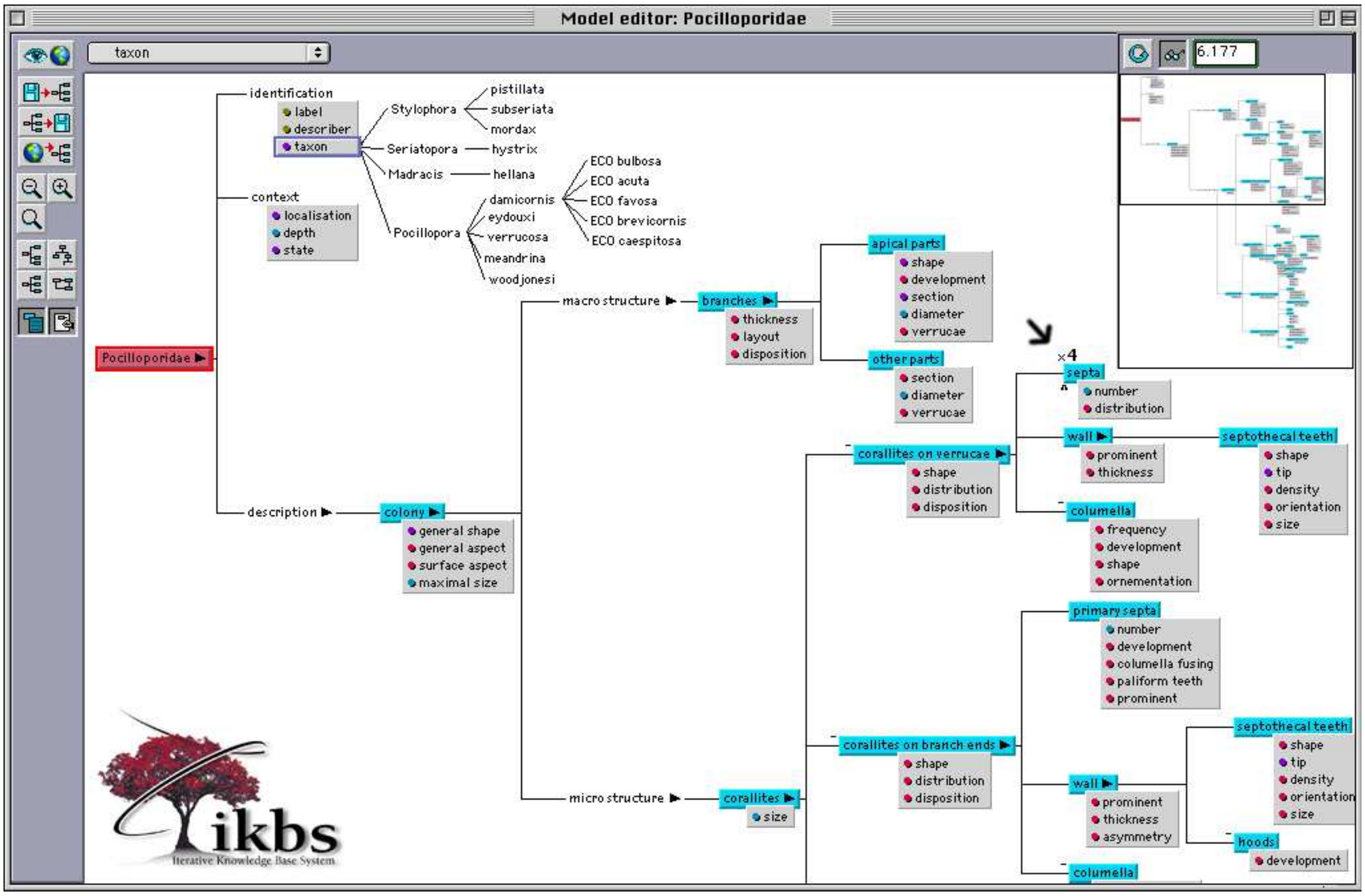

Figure 2. Part of the descriptive model of the Family *Pocilloporidae*

able to describe 4 genera and 14 species and ecomorphs (see attribute called “taxon” in Fig. 2).

There are multiple benefits in such a representation. Viewpoints divide the descriptive model into homogeneous parts, thus giving a frame of reference for describing organisms at a particular level of observation (see object “identification”, “context”, “description”, “macro” and “micro structure” in Fig. 2).

Sub-components introduce modularity into the descriptions making it possible to structure the domain from the more general to the more specific parts. This object representation of specimens is semantically better than the flat feature-value representation: in the former, local descriptions of attributes depend on the existence of parent objects, although in the latter the defined characters are independent of one another. Some of the possibly missing objects are marked with a minus sign (e.g. columella).

Fig. 2 shows the partitioning dimension of objects (subpart links for disjoint classes). For some of them (i.e. septa), other dimensions such as multi-instantiation (x symbol) and specialization (ŝymbol) of objects can be seen. The former enables users to describe several sorts of the same object by descriptive iteration (there are 4 possible instances for septa in Fig. 2) and the latter lets users name each sort with the help of the following classification tree of objects (specialization links in Fig. 3).

In fact, one of the roles of the descriptive model is to bring an observation guide to the end-user. The objects are linked together by relations that go from the most general to the most specific (from left to right), making the next description process easier for the non-specialist (see below).

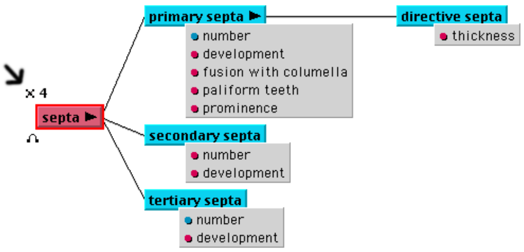

Figure 3. Classification tree of object “septa”

3.2 The questionnaire

Starting from the selection of a descriptive model, the program automatically generates a HTML questionnaire with illustrations, comments and animations (see Fig. 4). Each object of the descriptive model is illustrated by an HTML page with its corresponding attributes and values (see Fig. 5). This gives the expert the ability to index pictures, animations, and videos to each object and attribute of the descriptive model. The same functionality is offered to any end-user that wants to illustrate the specimens to describe, and thus justify his interpretation of the observation.

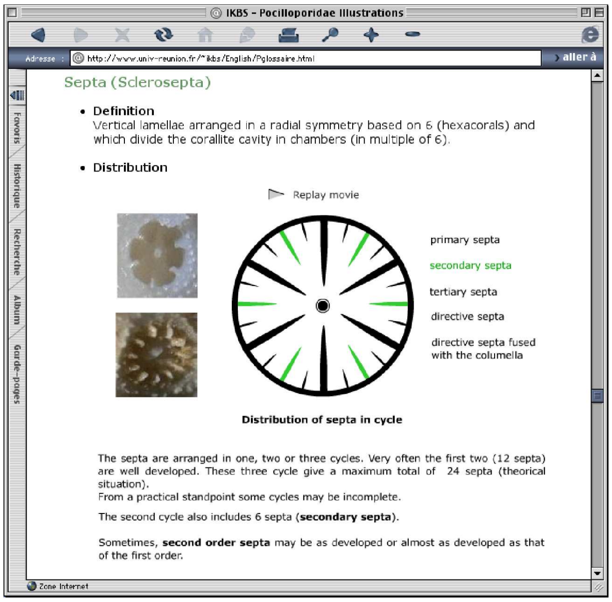

Figure 4. Definition of object "septa" in the glossary

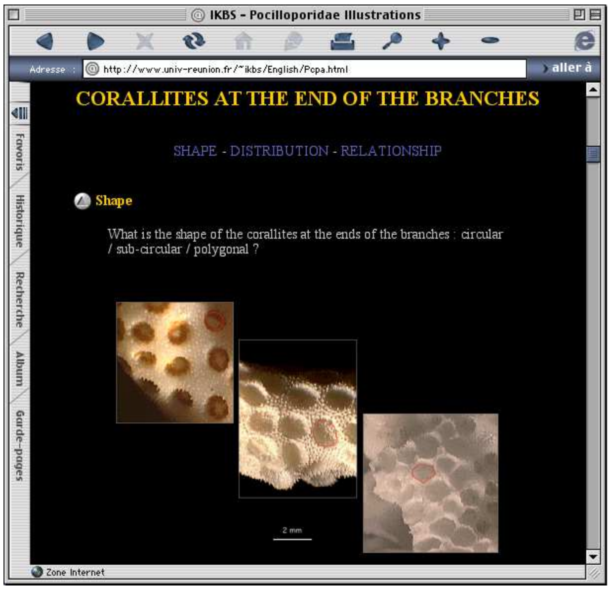

Figure 5. An HTML page for illustrating the object "corallites at the end of branches"

3.3 The descriptions

The second phase of knowledge acquisition allows biologists less informed than the experts to acquire personal descriptions and create a case base. An identification name is associated to each specimen observation in order to form a description or a case (Fig. 6).

The description process generates sub-trees of the descriptive model (see Fig. 2 and Fig. 6). Therefore, observed descriptions can be directly compared to one another by leafing through page by page: this navigation process is easier than viewing different lists of attribute-value pairs. In Fig. 6, we illustrate possibilities of $\mathcal{IKBS}$ for rendering complete and comprehensive descriptions of a given sample. Different types of attribute are used: taxonomic (e.g. general shape of object colony), numerical intervals (e.g. diameter of apical parts) and multi-nominal values (e.g. section of apical parts). The latter shows variation in objects displaying a set of multiple elements.

The visualization of objects differs graphically according to their status: black if present, black with a cross if absent, symbol **?** if unknown (see object "hoods" at the bottom-right side of Fig. 2 and Fig. 6). And last, an object can be specialized (e.g. the septa of corallites on branch ends, see Fig. 2): the result is a substitution of its name by a more precise one (e.g. primary septa, see Fig. 6) with its associated attributes (inherited or not, see Fig. 3). It is important for the user to visualize structured descriptions: so doing brings better clarity and comprehensibility to the acquisition phase.

4 KNOWLEDGE PROCESSING

In Systematics, data to be processed may be more complex than those considered in conventional data analysis. This complexity cannot be captured by a simple data matrix representation composed by a set of attributes and values. Diversity and incompleteness must be taken into account, and the exception is the only valid rule. The descriptions of specimens are often highly structured (composite or specialised objects), noisy (erroneous or unknown data) and polymorph (variable or imprecise data). Consequently, the design of new symbolic/numeric methods of data analysis that masters this complexity is a challenge for $\mathcal{IKBS}$.

From the computer science viewpoint, we have adapted learning methods from inductive learning algorithm C4.5 [9], software engineering integration [21] and case-based reasoning [1] fields. $\mathcal{IKBS}$ can be compared with AcknoSoft's KATE, Isoft's RECALL and TecInno's CBR-Works. These decision support systems have been designed to cope with industrial fields and very large databases [11]. For example, these tools are used for diagnosing troubleshooting Boeing 737 engines: the system contains 23,000 reference cases [19]. Our contribution was to develop new algorithms that exploit background knowledge to facilitate classification (class definition) and identification of natural organisms with the representation and processing of such reality.

From the biological viewpoint, identification methods have been well studied for thirty years in computer-aided Systematics, and many powerful tools have been developed [20]. Some biologists have come up with solutions (DELTA format) for coding descriptions [13]. Their programs (i.e. Pandora, Intkey) enable one to compare descriptions and facilitate identification process from databases [7]. [14] showed recently the need for interactive keys during the identification process (with unrestricted character use, easy updating, character deletion and changing, etc.). Other biologists emphasize multimedia and ergonomic solutions for education: the Linnaeus II

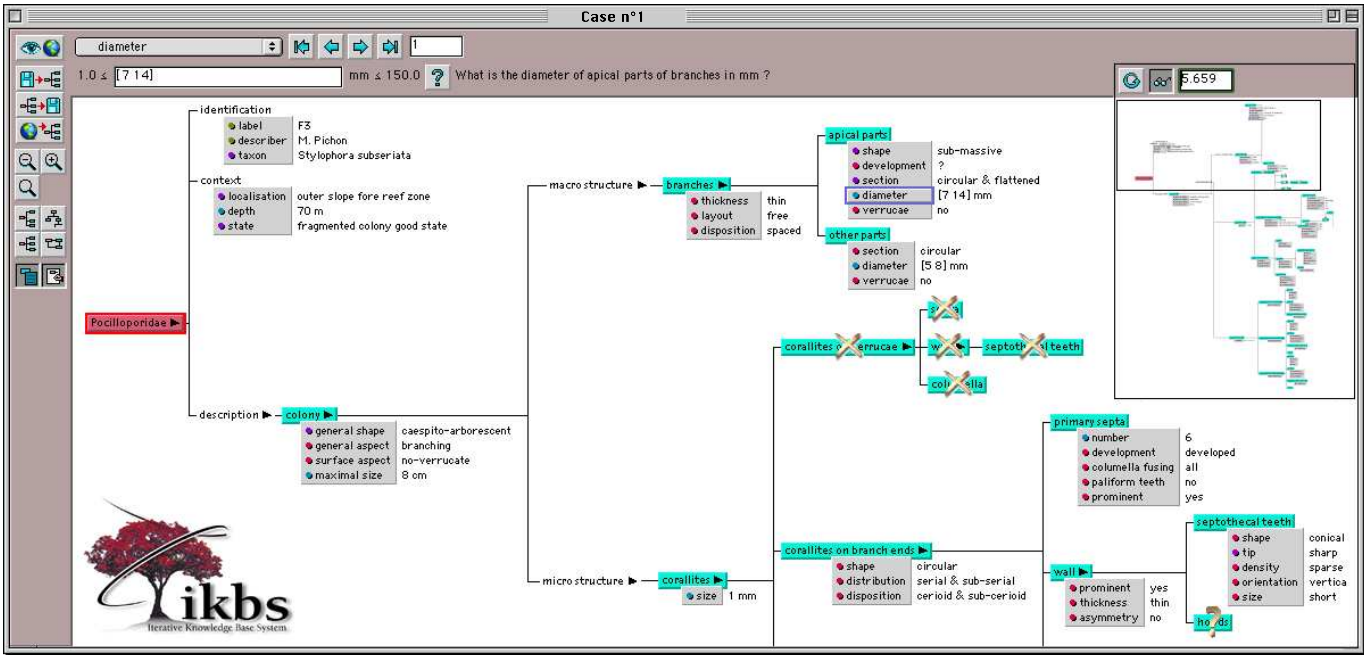

Figure 6. Part of the description tree of a case of the Family *Pocilloporidae*: *Stylophora subseriata*

software [18] proposes an electronic version of written dichotomous keys, a picture based and multiple-entry key based identification process.

All these works show clearly that there is a need for integrated tools that combine efficiently different approaches in order to satisfy different kind of end-users. $\mathcal{IKBS}$ proposes an easy to use and on-line identification and classification tool made in the Java programming language. It integrates two main approaches for finding the class (taxa) to which a specimen belongs. These approaches are based on decision tree method (monothetic selection of characters) and case-based reasoning method (polythetic selection of characters).

4.0.1 Decision tree

An accurate decision tree is generated from a descriptive model and a set of pre-classified cases. This is also called supervised classification in computer science. The decision tree is used to: 1) generate classification rules that correspond to conventional identification keys or 2) used by itself as an interactive process. It takes into account the complexity of previous descriptions (multiple, unknown, hierarchical values). The non-applicable values problem is managed by using the structure of the model and the absence/presence of objects [16]. The best character selection procedure is based on character weighting depending on two factors: 1) the observation cost and 2) the discrimination power of that character (information gain). The user who consults the decision tree can come back to a previous answered question. Another sub-tree is proposed when selecting another character or answering unknown. This dynamic aspect is achieved by the indexing of a sub-set of cases at each node of the decision tree. At each node, the set of indexed cases can be viewed and the case-based strategy can be used (see below).

An illustration of a decision tree built with 30 cases is shown in Fig. 7. The numbers refer to previously discussed classifier types: taxonomic attributes (1 and 2 highlight the use of the same classifier at two different levels of the hierarchy of values), multi-valued attributes (case bulbosa with size of spines "short&long" goes in three branches (3) and structured objects (4). For classification purposes, we use domain knowledge to generate accurate and meaningful decision trees (from pre-classified examples). Starting from the well known decision tree builder algorithm C4.5 [9] which works on discrete and continuous attributes, $\mathcal{IKBS}$ adds some functionality to this algorithm to deal with structured objects, taxonomic attribute-values and multi-valued attributes [16].

4.0.2 Case-based reasoning

This method allows the user to choose any characters in any order (random-access keys). Characters which are not available for the specimen observation, or whose interpretation is not clear, can be avoided. Then an interpretation of this incomplete description is made by selecting the k-nearest-neighbours (retrieval phase) in the case base and adapting them (adaptation phase) by maximising the probability of obtaining a correct identification, or by generating a decision tree as seen above from the k selected cases.

4.0.3 Classification

$\mathcal{IKBS}$ offers also unsupervised classification methods (from unclassified examples) that can be useful for experts: hierarchical ascend clustering and conceptual clustering. These methods are based on the use of a dissimilarity function for structured descriptions [3] that takes into account continuous, unknown and multiple character state values. The paper shows that these methods are particularly well suited for data pertaining to the biological domains.

5 VALIDATION AND REFINEMENT

We experimented on corals to test the reliability of $\mathcal{IKBS}$ identification with different users. We tested two con-

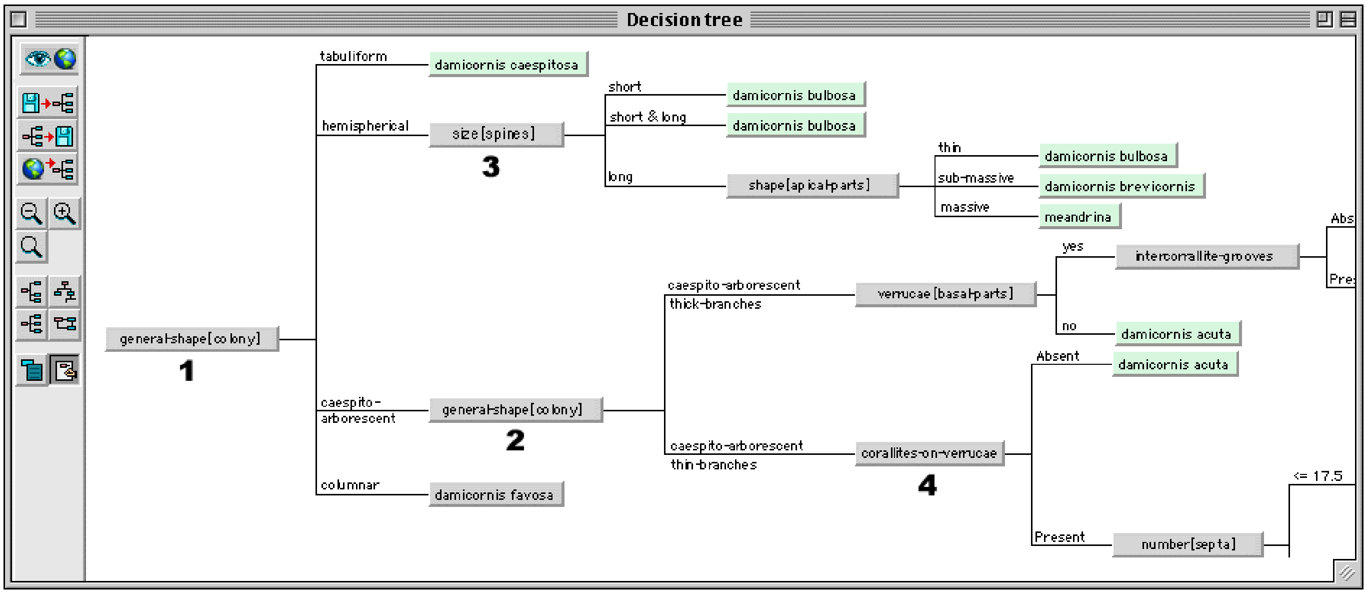

Figure 7. Part of a decision tree that makes use of domain knowledge

secutive descriptive models in a sub-domain of *Pocilloporidae*: the genus *Pocillopora* (9 species and ecomorphs). The validation of both descriptive models was qualitative. It led to modification of the initial descriptive model (dm1) and case base to the one shown in this paper (dm2). The first test with dm1 is called A. Later, another test B on dm2 was carried out. The experiments were made with a sample of 15 specimens of the Genus, each one of them being described completely in both descriptive models by 3 different biologists (x_1, x_2, x_3), and the expert (E). With this training set of 60 cases, the expert added 22 other descriptions of *Pocillopora* (37 expert cases). The four experiments that were led with $\mathcal{IKBS}$ are the following:

- A_1: 15 cases of x_n tested against 37 cases of E.
- A_2: 15 cases of x_n tested against 67 cases (E+other x_n).
- B_1: 15 cases of x_n tested against 37 cases of E.
- B_2: 15 cases of x_n tested against 67 cases (E+other x_n).

The results on 15 consultations are shown in table 1.

	x_1	x_2	x_3
A_1	6	7	8
A_2	9	8	10
B_1	9	10	11
B_2	11	10	12

Table 1. Number of good identifications with $\mathcal{IKBS}$

The results show that updating the first descriptive model and case base on *Pocillopora* brought better results. When testing the expert training set, it gave a 20% (3/15) of improvement in identification (from 46% to 66%). If we integrate other biologists' descriptions of the same specimens in the reference case base, the score goes up from 60% to 73%.

The reasons of these improvements are principally:

- The expert was able to detect inconsistencies in the first case base (omissions or errors in descriptions) and descriptive model (misunderstood characters, faulty illustrations). He could verify the answers of other biologists in regards to decision tree questioning that lead to misidentifications. He noticed the difficulties of interpretations of observation of specimens on some noisy comparative attributes and refined them into a new descriptive model.
- Consequently, the expert, aware of the importance of transmitting his knowledge to other biologists, postulates more precise and relevant characters that may be easier to observe and/or offer less ambiguous values (easier to interpret) in his descriptive model. For example, he will refine on the basis of mutually exclusive values, monosemic attributes, position of objects' attributes, glossary of terms, warning messages and observation help, enhanced illustrations and animations.

6 DISCUSSION

For end-users, a robust knowledge base is an application that enables reliable identifications of species. For experts and knowledge engineers, a robust knowledge base must satisfy qualitative descriptions and classifications. To achieve these goals, the problem we are faced with in Systematics is complex: good identifications depend on previously good classifications from experts, and also good descriptions from both experts and end-users. For classification purposes, Nature is so conceived that giving a name to organisms can also be difficult for experts (problems of synonymies), especially when there is a large intra-specific variation. This is the case in coral taxonomy where the number of named species in the world is uncertain [8]. Thus, managing complex knowledge in natural sciences means to cope with such evolving knowledge.

The objective of $\mathcal{IKBS}$ is to offer more robust descriptive work in Systematics in order to facilitate the identification of species by non specialists. We are faced with the problem of knowing, defining and reaching a consensus on: 1/ what to describe (taxonomic level of descriptions), 2/ how to represent descriptions (the choice of descriptive logics), 3/ what terms to use to develop an illustrated thesaurus to be presented as a reference for specialists of a domain.

The main goal of $\mathcal{IKBS}$ is to help produce quality descriptions from collection specimens, which is a key factor for getting better results in identification and avoid future revisions [15].

With such help, experts unaccustomed to computers are able to model and describe, and any non-specialists interested in the field can describe and identify new specimens. $\mathcal{IKBS}$ is used directly by experts to produce descriptive models and to fill cases without any help from a computer scientist. They find the interface very pleasant and enjoy the effectiveness of the tool.

In our methodology, it is important that the case base contains descriptions of specimens made by biologists other than the expert. This is made in order to counterbalance his interpretation of observations (inter-observer variation) when consulting the knowledge base. The results of the identification process are more dependable when we mix descriptions of different users for the same specimens (shown in Table 1). As they are labeled with the correct identification name from the expert, we can integrate the noise due to misinterpretations from end-users directly into the case base.

Similarly, because of the intra-specific variability, the number of described specimens by species must be increased. Insofar as *Pocilloporidae* is concerned, this family is one of the sixteen families of corals containing the largest intra-specific variability, and its complex diversity was covered with detailed precision. The difficulty arises due to the number of attributes applicable to each case. Thus, the building of an exhaustive knowledge base is time-consuming for describers: updating a case with the latest descriptive model on *Pocilloporidae* requires nearly a whole day's work!

7 CONCLUSION

In collaboration with four experts (taxonomists), we are presently experimenting with $\mathcal{IKBS}$ on four other families of corals of the Mascarene archipelago (*Fungiidae*, *Poritidae*, *Siderastreidae*, *Astrocoeniidae*) and one family of hydroids (*Sertulariidae*). The meticulous choice of terms, drawings and images seems decisive for generating a dependable knowledge base and managing the complexity of natural objects.

This is why we are designing $\mathcal{IKBS}$ to build cooperative knowledge bases. The aim is to encourage experts to draw up a common thesaurus of vocabulary and illustrations (i.e. the questionnaire) on the same Family.

Collections of specimens, like experts, are distributed around the world. Thanks to satellite high-speed broadband networks, we have been able to demonstrate TeleSystematics using video-conferencing and $\mathcal{IKBS}$. At MILIA'97 (International Market on Multimedia in Cannes, France) and at ATM-Developments'98 (Asynchronous Transfer Mode Telecommunications Conference in Rennes, France), experts were able to share their interpretations of observations of specimens under a microscopic examination synchronously between La Reunion (Southwest of Indian Ocean) and the French metropolitan town.

Nowadays, expertise in natural sciences is precious (it is becoming very rare). It is therefore urgent to develop tools that will ensure that expertise be collected and safeguarded for transmission to future generations. If this is not done, we will be left only with monographic descriptions and museum collections. The Reengineering of Systematics with $\mathcal{IKBS}$ is our response to this problem of enhancing scientific databases and museum collections.

ACKNOWLEDGEMENTS

We would like to thank the Conseil Régional of Reunion Island for supporting this work.

REFERENCES

[1] Aamodt A and Plaza E, 'Case-based reasoning: Foundational issues, methodological variations, and system approaches', *AI Communications*, **1**(7), 39–59, (1994).

[2] Grosser D., *Construction itérative de bases de connaissances descriptives et classificatoires avec la plate-forme à objets IKBS : application à la systématique des coraux des Mascareignes*, Thèse de doctorat, Université de la Réunion, January 2002.

[3] Grosser D., Diatta J., and Conruyt N., 'Improving dissimilarity functions with domain knowledge', *Proc. of the 4th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'2000)*, 409–415, (2000).

[4] Faure G., *Recherche sur les peuplements de scléractiniaires des récifs coralliens des Mascareignes*, Thèse es sciences, Université Aix-Marseille II, 1982.

[5] Rousse G., Conruyt N., Grosser D., and Vignes R., 'Object-oriented information system: toward systematics domain objects', *Poster presentation of the Ninth International Coral Reef Symposium*, (2000).

[6] Le Renard J. and Conruyt N., 'On the representation of observational data used for classification and identification of natural objects', *LNAI IFCS'93*, 308–315, (1994).

[7] Lebbe J., 'Systématique et informatique', *Systématique et biodiversité, Biosystema*, **1**(13), 71–79, (1995).

[8] Veron J.E.N. and Pichon M., *Scleractinia of eastern Australia, vol. I, Part I*, Australian Institute of Marine Science Monograph Series, 1976.

[9] Quinlan J.R., *C4.5: Programs for Machine Learning*, Morgan Kaufmann (Eds.), Los Altos (CA), 1993.

[10] Popper K.R., *La logique de la découverte scientifique*, Payot (Eds.) Press, Paris, 1973.

[11] Manago M., Althoff K.D., Auriol E., Traphoner R., Wess S., Conruyt N., and Maurer F., 'Induction and reasoning from cases', *in First European workshop on case-based reasoning (EWCBR-93)*, (1993).

[12] Manago M. and Conruyt N., 'Using information technology to solve real world problems', *Lecture Notes in Computer Science subseries, Springer-Verlag 622*, 22–37, (1992).

[13] Dallwitz M.J., Paine T.A., and Zurcher E.J., *User s guide to DELTA. A general system for processing taxonomic descriptions*, CSIRO, Div. Entomol., 4th ed., Canberra, 1993.

[14] Dallwitz M.J., Paine T.A., and Zurcher E.J. Interactive keys. http://biodiversity.uno.edu/delta/, 2000.

[15] Conruyt N., *Amélioration de la robustesse des systèmes d'aide la description, à la classification et à la détermination des objets biologiques*, Thèse de doctorat, Université Paris-IX-Dauphine, 1994.

[16] Conruyt N. and Grosser D., 'Managing complex knowledge in natural sciences', *Proc. 3rd International Conference on Case-Based Reasoning (ICCBR-99)*, 401–414, (1999).

[17] Conruyt N., Grosser D., and Geynet Y. Ikbs, an iterative knowledge base system for corals identification. http://www.univ-reunion.fr/∼ikbs, 2000.

[18] Schalk P.H. and Heijman R.P. Etis linnaeus ii taxonomic software : A new tool for interactive education. http://www.eti.uva.nl/, 1996.

[19] Heider R., *Troubleshooting CFM 56-3 engines for Boeing 737 using CBR and data mining*, 512–518, Transputing for Numerical and Neural Network Applications, In Smith and Faltings 1996, 1996.

[20] Pankhurst R.J., 'Computerized identification methods', *Practical taxonomic computing*, (1991).

[21] Schach S.R., *Implementation and integration phase. Software Engineering with Java*, Irwin Press, 1998.