5. Validation and refinement

We experimented on corals to test the reliability of IKBS identification with different users. We tested two consecutive descriptive models in a sub-domain of *Pocilloporidæ*: the genus *Pocillopora* (9 species and ecomorphs). The validation of both descriptive models was *qualitative*. It led to modification of the initial descriptive model (dm_1) and case base to the one shown in this paper (dm_2). The first test with dm_1 is called A. Later, another test B on dm_2 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 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.

Table 1. Number of good identifications with IKBS

	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

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 20% (3/15) of improvement in identification process (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:

1. 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.

2. 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, frames of reference, warning signals, enhanced illustrations.

6. Discussion

IKBS has been implemented in Java language and is fully operational on (http://www.univ-reunion.fr/~ikbs). Experts unaccustomed to computers are able to model and describe, and any non-specialists interested in the field can describe and identify new observations. IKBS is used directly by experts for creating descriptive models and filling 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, 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 *Pocilloporidæ* is concerned, this family is one of the sixteen families of corals containing the greatest intra-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 *Pocilloporidæ* requires nearly a whole day's work!

7. Related work

In other domains such as botany and zoology, some researchers have come up with solutions for coding descriptions [5]. Their programs enable to compare descriptions and facilitate identification process from databases [10], [15].

In Case-Based Reasoning methodology, 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 [3]. In the life sciences, our objective is to deal with more complex descriptions and less data (cases) by class.

8. Conclusions and future work

In collaboration with three experts, we are presently experimenting with IKBS on three other families of corals of the Mascarene archipelago (*Fungiidæ, Poritidæ, Thamnasteriidæ)*. 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 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 IKBS. At ATM Developments'98, experts were able to share their interpretations of observations of specimens under a microscopic examination synchronously between La Reunion (South-West of Indian Ocean) and Rennes (France).

Nowadays, expertise in natural sciences is precious (it becomes 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.

Acknowledgement

We would like to thank the French experts G. Faure, M. Pichon and M. Guillaume for their valuable contributions on the applications on corals. We are also grateful to Philippe Sills for English improvement. This work is supported by the Conseil Régional of La Reunion Island.

References

1. Aamodt A., Plaza E., Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, AI Communications 7(1): 39-59, 1994.
2. Allkin R., Handling taxonomic descriptions by computer, In; Allkin R. and Bisby F.A. (eds.), Databases in Systematics. Systematics Association London, Academic Press, 26: 263-278, 1984.
3. Althoff K. D., Auriol E., Barletta R., Manago M., A review of Industrial Case-Based Reasoning Tools, AI Intelligence, Oxford, 1995.
4. Conruyt N., Grosser D., Faure G. Ingénierie des connaissances en Sciences de la vie: application à la systématique des coraux des Mascareignes. Journées Ingénierie des Connaissances et Apprentissage Automatique (JICAA'97), Roscoff, pages 539-566, 1997.
5. Dallwitz M.J., Paine T.A., Zurcher E.J., User's guide to the DELTA System. A general system for processing taxonomic descriptions, Canberra: CSIRO, Div. Entomol., 4th ed., 1993.
6. Diederich J.R., Milton J., Creating domain specific metadata for scientific data and knowledge bases, IEEE Trans., Knowledge Data Engineering 3(4): 421-434, 1991.

7. Faure G., Recherche sur les peuplements de scléractiniaires des récifs coralliens des Mascareignes. Thèse es sciences, Univ Aix-Marseille II, 1982.
8. Fayyad U., Piatetsky-Shapiro G., Padhraic S., From Data Mining to Knowledge Discovery in Databases, AI magazine, 17(3): 37-54, Fall 1996.
9. Kodratoff Y. L'extraction de connaissances à partir des données. Journées Ingénierie des Connaissances et Apprentissage Automatique (JICAA'97), Roscoff, pages 539-566, 1997.
10. Lebbe J., Systématique et informatique. Systématique et biodiversité, Bourgoin T. (Ed), Biosystema, 13:71-79, Paris, 1995.
11. Le Renard J., Conruyt N. On the representation of observational data used for classification and identification of natural objects. IFCS'93, Lecture notes in Artificial Intelligence, Springer-Verlag, pages 308-315, 1994.
12. Manago M., Conruyt N. Using Information Technology to Solve Real World Problems, Lecture Notes in Computer Science subseries, 622: 22-37, Springer Verlag, 1992.
13. Manago M., Althoff K.D., Auriol E., Traphoner R., Wess S., Conruyt N., Maurer F., Induction and reasoning from cases, First European workshop on case-based reasoning (EWCBR-93), MM Richter, S Wess, KD Althoff and F Maurer (Eds.), Springer Verlag, (2), 1993.
14. Mingers J. Expert Systems – Rule induction with statistical data. Journal of the operational research society. 38(1): 39-47, 1987.
15. Pankhurst R.J., Practical taxonomic computing. Cambridge Univ. Press, Cambridge, 1991.
16. Popper K.R., La logique de la découverte scientifique. Payot (Eds.) Press, Paris, 1973.
17. Quinlan J.R., C4.5: Programs for Machine Learning, Morgan Kaufmann, Los Altos, CA, 1993.
18. Veron J.E.N., Pichon M., Scleractinia of eastern australia, vol. I, Part I, Australian Institute of Marine Science Monograph Series, 1976.