Knowledge Management in Environmental Sciences with \textit{IKBS}:
Application to Systematics of Corals of the Mascarene Archipelago

Noel Conruyt and David Grosser

Laboratoire IREMIA, Université de la Réunion,
15 avenue René Cassin, BP 7151,
97715 Saint-Denis Msg. Cedex 9, France
Web: http://ikbs.univ-reunion.fr/
conruyt,grosser@univ-reunion.fr

Abstract. Systematics, the scientific discipline that deals with listing, describing, naming, classifying and identifying living organisms is a central point in environmental sciences. Expertise is becoming rare and 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, 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 taxa and descriptions of specimens. We have designed an Iterative Knowledge Base System – \textit{IKBS} – for achieving these goals. It applies the scientific method in biology (conjecture and test) with a natural process of knowledge management. 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.

This paper presents an overview of the methodology, the methods (identification tree and case-based reasoning) and the validation process used to build knowledge bases in Systematics. An application on corals of the Mascarene Archipelago is given as a case study.

1 Introduction

Today around the world, scientific databases are increasingly delivered on CD-ROM or through Internet (e.g. World Biodiversity data-base from ETI in Netherlands, Reefbase and Fishbase from ICLARM in the Philippines, Hawaii Biological Survey databases, Coral Id at AIMS, 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 collections 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.

We have developed a type of knowledge base that supports the above methodology. The tool that generates such applications is called *IKBS* (*Iterative Knowledge Base System*; Grosser (2002)). *IKBS* is a knowledge management system available on the Internet which is developed in the object-oriented language Java. This tool was co-designed with specialists and end-users for 15 years in different domains such as plant pathology diagnosis, Manago (1992) and computer aided Systematics, Conruyt (1994). 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).

2 Knowledge acquisition

Three points have to be addressed for the knowledge acquisition process: descriptive model definition, questionnaire generation and case acquisition.

2.1 The descriptive model

The descriptive model 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, Le Renard et al. (1994). These rules were constructed from the analysis of