

Knowledge Base with XPER3 in Palaeontology: Calculated descriptors offer new functionalities

Adeline KERNER, Regine Vignes-Lebbe

Xper3 (<http://www.xper3.com/>) is a web platform that manages descriptive data and provides interactive identification keys. Xper3 and its previous version Xper2 have been already used for various taxonomic groups. This software is well adapted for identification in paleontology, because fossils are often incomplete and unsuitable for identification by means of classical dichotomous keys. Xper2 and Xper3 are compatible with TDWG's Structured Descriptive Data (SDD) format.

In the descriptive model, three descriptor types can be used: categorical (i.e., qualitative or intervals), numerical (continuous data) and calculated. Calculated descriptors are automatically computed from other descriptors by using logical operators (boolean operators). We will focus on fossils in order to reveal new possibilities provided by calculated descriptors in Xper3 knowledge bases.

The main source of content is the archaeocyaths knowledge base (<http://infosyslab.fr/archaeocyatha/>). Archaeocyaths are the oldest calcified sponges and the first animals to build reefs. Because they are important Cambrian organisms, the archaeocyaths knowledge base is an efficient resource for scientific studies and a useful tool for non-specialists, especially with the support of calculated descriptors.

Calculated descriptors can be used to reword morphological descriptors (adapted for identification purpose) into homologous characters (adapted for phylogeny). Likewise, calculated descriptors make it possible to reconstruct the literature terminology from a set of more limited descriptors. For example, the term "cambroid pores" contains information about the shape and the repartition of pores. Cambroid pores are simple OR polygonal pores (shape) AND with a regular distribution on the wall (repartition) AND a random arrangement (repartition type).

Calculated descriptors can also be used to create a consistent multilingual interactive identification key (all translations are computed from a single language set of descriptors and taxonomic descriptions). Likewise, calculated descriptors can be used to generate descriptors adapted to different contexts and users (alternative character in DELTA are limited to a single character).

Calculated descriptors did not exist in Xper2. They do not exist in SDD format and so they are exported from Xper3 as categorical descriptors, losing the origin of the values. We are interested in discussing these with SDD users and Xper3 users in order to improve the user interface and develop new tools for the analysis of such descriptors. We will discuss too the distinctive features of paleontological data.