# The Aquarelle Terminology Service

by Martin Doerr, Irini Fundulaki

{martin, fundul}@ics.forth.gr

*ICS-FORTH*

Thesauri and other kinds of authority data, place names, artist names, periods etc., are important for the intellectual access to information assets. It is widely accepted that the use of thesauri for the classification of assets and as search aid considerably improves precision and recall of retrieval methods. This holds in particular for database records with few words per field - the typical form of records about museum objects. Even more in a multilingual environment there should be support for translation - either transparent transformation of the valid terminology a user refers in his request into the languages of the addressed databases, or at least a guidance of the user to the local terminology in use. The AQUARELLE users required from an early stage of the project that this support must be based on high quality information managed by human experts.

For optimal results, the terms used for asset classification, in the search aid thesaurus and in the experts' terminology should be consistent. This led us to a three level architecture of components cooperating with the IT environment of an Aquarelle installation: vocabularies in local databases, local thesaurus management systems of wider use and central Term Servers for retrieval.

Typically, local databases have a more or less idiosynchratic way to enforce vocabulary control. For reasons of standardization of format and centralization of handling, we foresee an independent thesaurus manager to which the vocabularies of several local databases are loaded, and in the sequence organized as thesauri ("authorities") by an expert, following variations of the ISO2788 semantic structure. In addition, standard external vocabularies can be loaded. These authorities may be specific to one database, a user organization, or a whole language group. The local vocabularies and terms already used for classification may need updating with changes done at the thesaurus manager.

The Aquarelle Access Server, which is responsible for the distribution and transformation of user requests, needs knowledge of the authorities in local use, at least of the higher level terms. Therefore it communicates with one or more Term Servers, which hold released versions or extracts of the local authorities. Moreover, a Term Server must be fed with equivalence expressions between the meaning of terms in different authorities, either by an expert team or by linguistic methods and subsequent human control. These expressions are used to replace the terms in a user request with equivalent terms of the target system - automatically or in a dialogue with the user.

This three stage architecture reflects ideally the practice and needs of classification, expert agreement, user organization and search aids. It is a fully scalable solution and a flexible approach to standards enforcement. The SIS-Thesaurus Management System (SIS-TMS) developed in the past by FORTH was extended for AQUARELLE in order to support the above scenario. It will be a product of FORTH by summer 1998. It implements a client-server architecture. There is a client for reading and one for manual editing. Another client on the same base, the Term Server, was developed by ILSP.

The system has several innovative features. It allows to maintain multiple, multilingual thesauri and their interrelations in one logical database. Different teams can cooperatively maintain multiple systems of semantic relations on a shared body of terms and concepts. User groups can further specialize the semantics of ISO2788 and ISO5964 (multilingual thesauri) links and add custom fields.

The **SIS-TMS** graphical user interface allows for the unconstraint navigation within and between different thesauri; the execution of predefined queries and graphical views to identify concepts for cataloguing or database queries; to identify translations or equivalent expressions for information access in a heterogeneous environment; and to control the quality and the logical consistency of a system of interlinked thesauri.

The editing system maintains a history of changes and provides release operations for a set of changes done. Referential integrity and vocabulary control is maintained throughout the system. These mechanisms are the prerequisite to update incrementally the local databases and the Term Servers from the thesaurus development units at regular release intervals with minimal possible human intervention.

The local integration of the user databases with the terminology system exceeds the AQUARELLE framework, but it can be done now with minor effort by any skilled programmer. AQUARELLE will end with the evaluation of the terminology management system and the Term Server, using as test data the AAT, RCHME and the multilingual MERIMEE thesaurus. With this terminology service, AQUARELLE provides the technical means to solve one of the major problems of the semantic interoperability. Further work must concentrate on means to make the creation of authorities cheaper, and on the social organization of the cooperative development and use of authorities.