

## A knowledge-based interface for accessing biological databases

Paolo Bresciani(1), Paolo Fontana(2), Stefano Toppo(3) and Riccardo Velasco(4)

(1) ITC-irst (Istituto per la Ricerca Scientifica e Tecnologica) - Trento

(2) DIT, Universit? di Trento

(3) CRIBI, Universit? di Padova

(4) IASMA (Istituto Agrario di S. Michele all'Adige)

An important task faced by bioinformatics is data and scientific knowledge management. This activity involves several aspects, like, for example, storing, extracting, organizing, analyzing, and make data and knowledge assets easily usable and accessible.

The quickly increasing quantity of biological data, that far overpass the capability of analysis and use, requires a clever use of the best of information and data technology, in order to try to maximize the digestion of the ever-growing mass of genetic information. Although many efforts have been done in this direction, still one aspect has been, in our opinion, under-explored, that is the realization of easy, intuitive, and flexible user-interfaces for complex and structured data access.

The availability of intuitive interfaces that really help the user represent a necessary condition to gain user's acceptance and confidence in the provided data and knowledge resource.

We tried to tackle this problem by means of a first prototypical knowledge-based flexible tool for query formulation against relational and object-oriented databases [BNP2000]. Its most interesting features are based on its capability of reasoning on the semantics of the queries. These features are aimed at improving the interaction between inexperienced users and complex databases, thanks to advanced knowledge reasoning algorithms. In particular, it provides a simple, flexible, and intuitive visual interface that allow the user to:

- <br> - interactively and iteratively build queries, exploring the semantic of database schemas via refinement and generalization of terms;
- <br> - be prevented from building semantically inconsistent queries;
- <br> - be gradually introduced only to those part of the conceptual model relevant for the query formulation;
- <br> - be provided with simple, but effective, features for query refinement and query generalization.

All these characteristics are desirable in all the situations in which "unskilled" users (i.e. non DB experts) would like to access complex and highly structured data by formulating sophisticated and extemporary queries. These users are likely to have no familiarity with structured query languages like SQL, and probably they are just occasional users, unwilling to spend many efforts in trying to understand the way in which the information is logically organized in the database.

This is the situation faced, in most of the cases, also by the normal users of biological databases.

In the present communication we will explain how the concepts introduced above are being applied to two cases of biological databases.

The first case, currently in an advanced state of integration, refers to the [Muscle TRAIT](#) maintained by C.R.I.B.I. (Interdepartmental Research Center of Biotechnology) of Padua University. Transcripts expressed in skeletal muscle are annotated in this database; the whole work is supported by Telethon Foundation.

The second case is the database on the genoma of *Vitis vinifera*, in particular, set of genes expressed in different tissues at several developmental stages and under different biotic stress. This *Vitis* database, once is organised, could be proposed as a model for the international *Vitis* Consortium. The data are being acquired as part of the activities of the project "GENOMA VITE", started in year 2001 and ending in year 2005 at IASMA, sponsored by the CA.RI.T.RO Foundation.