

CROSSMiner: a deductive system to discover biological relations

ID - 188

Licciulli Flavio¹, Grillo Giorgio¹, Turi Antonio¹, Delle Foglie Gianfranco¹, Catalano Domenico¹

¹Institute for Biomedical Technologies, ITB-CNR, BARI, Italy

Motivation

Methods

CROSSMiner uses cross-references between biological database pairs to find out relations among biological entities not directly linked without any a priori knowledge of the database content and structure. The system architecture is composed by a repository containing direct 'cross- reference' between biological database identifiers (IDs or Accession number) and a software engine to navigate a graph representing all relations and paths existing among different databases. The repository is implemented in a relational schema and the software engine uses BreathFirst and BackTracking techniques to explore the graph. The tool we developed, is free from any data annotation and source database format/structure since it exploits only IDs to link different database entries and this yields an easy and efficient repository updating. Due to this approach no change to the CROSSMiner schema and structure is required so it can be considered a simple, scalable and dynamic system.

The engine builds a graph starting from the pool of direct cross-references. A relation between every implemented database is represented by one or more navigational paths in the graph which relate a set of source database IDs to a correspondent set of one or more destination databases.

In the initialization step the software engine sets all possible paths among different databases up using the navigational graph; then, in the query step, it iteratively explores the graph to perform all the expected paths.

To avoid the overloading in the path navigation process the engine that uses the graph approach limits the result to the most affordable path based on heuristics (see below) and therefore it has been joined with a deductive approach making use of an engine based on logic programming (Prolog). This solution is more efficient because it dynamically computes the really feasible paths based on the run-time input query and knowledge-base (IDs cross- references), dramatically reducing the amount of explored paths.

CROSSMiner adopts a set of heuristics to estimate the reliability of results (navigational path) through the assignment of an a priori weight to each cross- reference available in the repository, the number of feasible paths between two databases, the path length (crossing nodes or databases). The resulting paths that relate a database A with database B are ranked by means of the calculated assessment.

Results

CROSSMiner is a deductive system to discovery new relations among data available in databases not directly linked or connected by complex cross- reference paths.

The tool is an effective solution to help a researcher to easily integrate his own data with public databases empowering the data analysis; it also facilitates the development and updating process of new specialized databases.

An example of useful application is the comparison of expression data coming from microarray experiments carried out on different microchip platforms. We successfully tested the tool in the development process of the p53FamTaG database in which experimental data from in silico and in vitro analysis have been annotated through cross-refences to the main public databases.

Currently a prototype showing CROSSMiner features is available as a web- service and a web interface.

Email: flavio.licciulli@ba.itb.cnr.it