



Graphs processing is a challenge in the most areas of sciences (biology, engineering and humanities) since graphs allow us to represent and, later, to process several types of data structures and relations. Nevertheless, the analysis of the information related to these problems, represented by means of graphs, becomes less useful when there exists a high volume of information or it is necessary to extract a variety of possible relations, similarities, etc., of the graphs stored in large databases. Thus, the problems related to the maintenance, processing and data mining of large databases and problems related to graphs and subgraphs isomorphism are a challenge for the researchers since new models and new algorithms are required to develop useful solutions for information retrieving with a low computational cost.

The objective of our project is the proposal and development of new computational solutions using graphs and subgraphs isomorphism for the construction, classification and mining of large databases in which the information is modelled using graphs.

Since the graph topology determines real characteristics and properties of the information represented, and based on previous works in which we have development models and suitable algorithms for the extraction of structural characteristics of graphs and the measure of similarity based on graph isomorphism, in this project we will extend and improve the existing proposals using techniques of soft computing, agents and parallelization of algorithms, since: (a) the development of new algorithms based on fuzzy values of subgraph isomorphism will allow us to obtain more trustworthy measures of similarity, (b) the use of approximate similarity measurements will allow the construction of more effective clustering and screening methods applied to medium and large databases, c) these approximate isomorphism measurements will allow to obtain improved models of properties prediction, d) the use of these measurements and topological descriptors will allow the construction of combinatorial (virtual) databases which they have a high interest in pharmacological research, and e) the new algorithms will allow to diminish the computational cost in these combinatorial and NP-complete problems.

Our developments will give to a set of software systems that they will be published in a Web portal for their use, under registration, for the scientific community. These systems

(isomorphism calculation, clustering, screening and construction of combinatorial databases) built with Java language, Oracle DBMS and procedures in native languages (C, PL/SQL, etc.) they will be able to be susceptible to trade or patent.

The projects goals are:

- Estructurate similarity calculations with softcomputing algorithms of approximate similarity subgraphs and graphs.
- Applications QSPR / QSAR based on approximate similarity subgraphs and graphs.
- Development algorithm based on requierement satisfaction for generation of combinatorial data base of graphs.
- Deveopment multidimensional and evolutive classification model based on approximate similarity measurements.
- To make a tool for data base HTS.
- Documentation. Results diffusion. Coordination.