INFOMIX: Boosting Information Integration

Why INFOMIX:

Recent developments in Information Technologies such as the expansion of the Internet and the World Wide Web, have made available to users a huge number of information sources, generally autonomous, heterogeneous and widely distributed: as a consequence, information integration has emerged as a crucial issue in many application domains, e.g., distributed databases, cooperative information systems, data warehousing, or on-demand computing. Recent estimates view information integration to be a $10 Billion market by 2006. However, information integration is in general an extremely complex task. Both state-of-the-art commercial software solutions and academic systems fulfill only partially the ambitious goal of integrating information in complex application scenarios.

What is INFOMIX:

INFOMIX is an advanced system for information integration which has been developed within an Information Societies Technology project (IST-2001-33570). INFOMIX offers transparent access to the data dispersed over different sources by providing users with a unified view of them, called global schema. Users formulate queries over the global schema in order to extract data relevant for their aims. Then, the system is in charge of accessing each source separately, combining local results into the global answer, and supplying the answer to the user, who is not obliged to have any information about the sources. To this aim, INFOMIX implements recent results in database theory and artificial intelligence, and make use of the advanced capabilities provided by the reasoner DLV.

INFOMIX is innovative in the following respects:
- Advanced reasoning capabilities
- Advanced information integration algorithms
- Sophisticated optimization techniques
- A rich data acquisition and transformation framework
- Handling of semi-structured data
- Declarative user-interaction
- Handling of incomplete and inconsistent data sources.

INFOMIX architecture and functionalities:

The INFOMIX system supports two modes: a design and a query mode. In the first, the global schema, the source schema, and the mapping between them are specified. Furthermore, wrappers for the data sources are created or imported. In the query mode, the system provides query answering facilities at run time, including data acquisition, integration, answer computation, and presentation to the user. In both the design and query mode, INFOMIX is conceptually divided into three levels:

1. **Information Service Level.** This level serves as a direct interface to the user (at run time) and the designer (at design time). It deals with global data, and provides the necessary interfaces (e.g., for global schema definition and for query formulation).
2. **Internal Integration Level.** In this level, the actual integration of data is performed. It receives data from the sources, and exploits the mapping and the information of the global schema to effectively compute answers to the queries posed by the user at run time, relying on a the DLV system at its core.
3. **Data Acquisition and Transformation Level.** Typically, information sources do not provide homogeneous data, which therefore can not be uniformly accessed. This level controls the way in which these data are acquired by the system, performs data transformations, alignment and cleaning, allowing upper levels to deal with source data in a uniform way.

Who contributed to INFOMIX:

The system has been developed by a consortium comprising partners that come from three different European countries. The University of Calabria (Italy), the Vienna University of Technology (Austria), the University of Rome “La Sapienza” (Italy), and Rodan Systems (Poland) covered skills and expertise in the fields of information integration, database theory, computational logic and deductive databases, database implementation and optimization.