About the Project

Constantly growing amounts of data, complicated and rapidly changing economic interactions, and an emerging trend of incorporating unstructured data into analytics, is bringing new challenges to Business Intelligence (BI). Contemporary solutions involve BI users dealing with increasingly complex analyses. According to a 2008 study by Information Week, the complexity of BI tools and their interfaces is becoming the biggest barrier to success for these systems. Moreover, classical BI solutions have, so far, neglected the meaning of data, which can limit the completeness of analysis and make it difficult, for example, to remove redundant data from federated sources.

Semantic Technologies, however, focus on the meaning of data and are capable of dealing with both unstructured and structured data. Having the meaning of data and a sound reasoning mechanism in place, a user can be better guided during an analysis. For example, a piece of information can be semantically explained or a new relevant fact can brought to the user's attention. In particular, we foresee a well known semantic technique called Formal Concept Analysis (FCA) to be a key element of new hybrid BI system. FCA can be used to guide a user in discovering new facts, which are not explicitly modelled in the data warehouse schema. Semantic analysis could also improve classical methods in BI, such as data reduction and duplicate detection. However, semantic technologies have traditionally operated on data sets a magnitude smaller than classical BI solutions. They also lack standard BI functionalities such as Online Analytical Processing (OLAP) queries, making it difficult to perform analysis over semantic data. The CUBIST project develops methodologies and a platform that combines essential features of Semantic Technologies and Business Intelligence. With CUBIST, we envision a system with the following core features:

- Support for the federation of data from a variety of unstructured and structured sources.
- A data persistency layer in the form of a semantic Data Warehouse; a hybrid approach based a BI enabled triple store.
- Semantic information used to improve BI best practices in, for example, data reduction and preprocessing; CUBIST enables a user to perform BI operations over semantic data.
- A semantic data warehouse that realizes the advanced mining techniques of Formal Concept Analysis (FCA).
- FCA guides the user in performing BI and helps the user discover facts not expressed explicitly by the warehouse model.
- Novel ways of applying visual analytics in which meaningful diagrammatic representations will be used for depicting the data, navigating through the data and for visually querying the data.

CUBIST demonstrates the resulting technology stack in the fields of market intelligence, computational biology and the field of control centre operations.

CUBIST is funded by the European Commission under the 7th Framework Programme of ICT, topic 4.3: Intelligent Information Management.

The CUBIST promotional video highlights the visual analytics capabilities of CUBIST:

The following screencast shows the integrated CUBIST prototype in action. The video is structured as follows:

- Starting at 0:00 an overview of the project is given.
- Starting at 2:53 navigation along semantic relationships is demonstrated.
- Starting at 6:30 the semantic search facilities are shown.
- Starting at 14:00 exploring the data with a graph exploration is shown.
- Starting at 17:14 we demonstrate the Visual Analytics capabilities of CUBIST.