CUBIST 2010 Annual Report

Project Summary

Constantly growing amounts of data and an emerging trend of incorporating unstructured data into analytics is bringing new challenges to Business Intelligence (BI). Contemporary BI solutions fall short in the following aspects: First, they focus on structured data only and disregard the increasing amount of information hidden in unstructured data. Secondly, the complexity of BI tools becomes the biggest barrier to their success.

CUBIST copes with the above mentioned problems by combining essential features of Semantic Technologies, Business Intelligence and Visual Analytics. CUBIST aims to

- support federation of data from unstructured and structured sources,
- persist the federated data in a BI enabled triple store,
- provide novel ways of applying visual analytics in which conceptually meaningful diagrammatic representations of the data will be used.

Summary of Activities

As CUBIST started in 2010/10, it is in its requirement analysis phase, which will be the main activity for the first six months of the project. In 2010, the following main activities have been conducted:

- **Dissemination:** To inform the public about the start and content of CUBIST, several dissemination activities have been carried out.
- **Kickoff meeting and tools:** A two days kickoff meeting took place at the beginning of October. Besides discussing general project planning issues, the meeting has been used to achieve a better joint understanding of the concepts, technologies and tools to be used in CUBIST on the on hand and of the three CUBIST use cases on the other hand.
- **Requirement analysis:** To initiate the requirement analysis, the use case partners have been provided guidelines and means to trawl and describe requirements.

With the expected outcome of the requirement analysis and the partner tools which are either already available or which are to be implemented, the CUBIST consortium is looking forward to develop first non-integrated prototypes in 2011. Below, some more details are provided.

Dissemination activities: The CUBIST website, including two factsheets for downloading, has been set up. A majority of CUBIST partners have announced CUBIST on their websites.
Actively pushing information about CUBIST to the press has particularly been carried out by Sheffield Hallam University (SHU). Due to these activities, several online magazines have posted articles and interviews on their sites. To name the most important ones: semanticweb.com, www.zdnet.co.uk, www.businesscomputingworld.co.uk, www.c3.co.uk, www.computerweekly.com and kntheiet.org. In addition to these activities, an article about CUBIST has been published in Headline, a print magazine from Sheffield Hallam University.

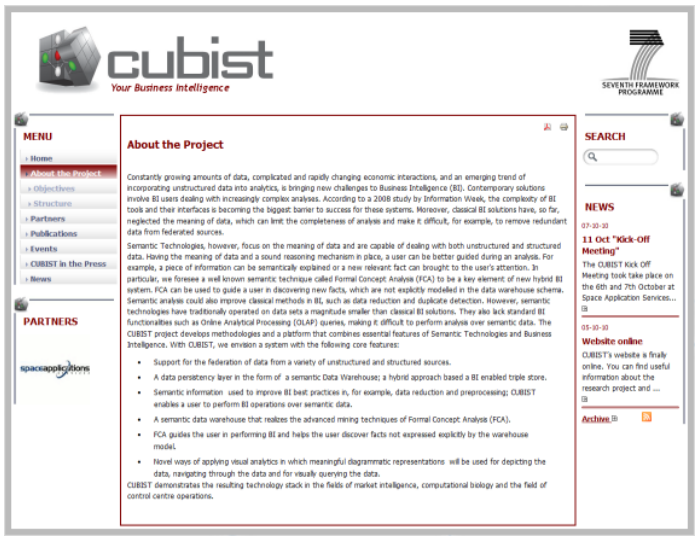
Requirement Analysis: Besides standard means like personas, mock-ups or a template for taking down and categorizing atomic requirements, other means like so-called utilization scenarios which serve better the research nature of CUBIST will be employed as well. For February 2011, the next CUBIST meeting which is dedicated to the requirement analysis is planned.

Technologies and tools:

- The persistency layer in CUBIST will be Ontotext's high-performance semantic repository BIGOWLIM. In 2010/11, Ontotext has released version 3.4 of their OWLIM[1] family of semantic repositories. This release is aimed at improving the integration possibilities for BigOWLIM by including support for the popular Jena RDF framework[2]. Due to this, Jena's ARQ engine[3] allows BigOWLIM to handle the latest SPARQL 1.1 extensions[4], e.g. aggregates, which is extremely relevant for CUBIST.
- The underlying theoretical framework for the visual analytics is "formal concept analysis" (FCA): a mathematical approach for conceptually clustering objects and arranging these clusters in a hierarchy called "concept lattice". CUBIST partner Sheffield Hallam University will contribute two FCA tools which will be extended and improved within CUBIST, namely FCABedrock[5] for converting data sets into formal context (the underlying data structure of FCA) and InClose[6] (for transforming formal contexts into concept lattices).
- The front-end of VA will be developed by CRSA and will incorporate existing techniques and algorithms they have developed on lattice filtering, visualization, tree extraction and exploration.

User Involvement, Promotion and Awareness

The above mentioned dissemination activities have been the first important steps in order to raise and increase community awareness and involvement. Different kinds of information for the diversity of potentially interested parties are provided by the website. It offers general information about the projects aims and goals, its setup and its partners, a publication list, press activities, as well as two project factsheets. Due to the dissemination activities of Sheffield Hallam University, despite being in it's ramp-up phase, CUBIST already raised considerable attention in the public. For 2011, further steps and dissemination channels are targeted.



Future Work in 2011

In 2011, a first non-integrated prototype will be developed. This prototype will be based on already existing components like BIGOwlim, FCAbedrock and InClose, as well as on further components (particularly for the visual analytics) which are to be developed in 2011.

The public will be informed about the project's progress on the existing dissemination channels. Further dissemination means like a youtube channel and a blog will be set up in order broaden CUBIST's audience. In order to particularly target the research community, a dedicated CUBIST workshop is planned for the International Conference on Conceptual Structures.

[1] http://www.ontotext.com/owlim/index.html
[2] http://jena.sourceforge.net/
[3] http://jena.sourceforge.net/ARQ/
[4] http://www.w3.org/TR/sparql11-query/
[5] http://sourceforge.net/projects/fcabedrock/
[6] http://sourceforge.net/projects/inclose/