SemaGrow

Data Intensive Techniques to Boost the Real – Time Performance of Global Agricultural Data Infrastructures

Project Vision

During the last years, the trend to open up data and provide them freely on the Internet has intensified in volume as well as quality and value of the data made available. The linked data community has grasped the opportunity to combine, cross-reference, and analyse unprecedented volumes of high-quality data and to build innovative applications. This effort has caused a tremendous network effect, adding value and creating new opportunities for everybody, including the original data providers.

But most of the low-hanging fruit has been picked and *it is time to move on to the next step, combining, cross-indexing and, in general, making the best out of all public data*, regardless of their size, update rate, and schema; accepting that centrally-managed repositories (even distributed) are not able to meet the challenges ahead and that we need to develop the infrastructure for the efficient querying of large-scale federations of independently-managed sources.

1st year of the Project is complete!

During the first year of SemaGrow we progressed towards all the major objectives that will realize the project's vision and goals:

- We focused on the development of a triple store that is capable of inferring triples from POWDER statements, as these provide a convenient formalism for succinctly storing data summaries of what data is stored where in the federation.

- We carried out ontology alignment experiments, focusing on synthesis approaches and collaborative, semi-automatic alignment methods. A first prototype of a GUI for human-assisted alignment was also developed

- We designed a detailed architecture of the SemaGrow Stack - the system that will integrate the research outcomes into a functional platform serving the envisioned Use Cases – and developed a first version of the core SemaGrow components.

- We investigated the state-of-the-art and the major trends on query decomposition and transformation technologies and identified the extensions required for serving the SemaGrow case.

For more information: **www.semagrow.eu**

Heterogeneous Data Collections & Streams
Use Case

The perspective from which extremely large and very complex agriculture-related data sets are considered is the one of **research activities**, during which the users need to cope with **heterogeneous data collections & streams** in order to **achieve new scientific investigations** that may help **forecast and address societal challenges such as food production in changing climate conditions**.

The perspective from which extremely large and very complex agriculture-related data sets are considered is the one of **information management**, during which the users need to cope with **reactive analysis of the data within the time scale and processes that they need to support** in order to **create value through extensive data collection and analysis** that may help **timely and better decision making related to societal challenges like food security**.

Reactive Data Analysis
Use Case

Reactive Resource Discovery
Use Case

The perspective from which extremely large and very complex agriculture-related data sets are considered is the one of **education**, during which the users need to cope with **reactive resource discovery** in order to be able to **find, reuse and exploit data resources created in one environment in very different contexts**.

Contacts

Project Coordinator	*Scientific Manager*	*Technical Manager*
Prof. Miguel A. Sicilia	Dr. Vangelis Karkaletsis	Dr. Pythagoras P. Karampiperis
Universidad de Alcalá (UAH)	*National Centre for Scientific Research "Demokritos" (NCSR-D)*	*National Centre for Scientific Research "Demokritos" (NCSR-D)*
msicilia@uah.es	vangelis@iit.demokritos.gr	pythk@iit.demokritos.gr
http://www.cc.uah.es/msicilia/	http://users.iit.demokritos.gr/~vangelis/	http://users.iit.demokritos.gr/~pythk/