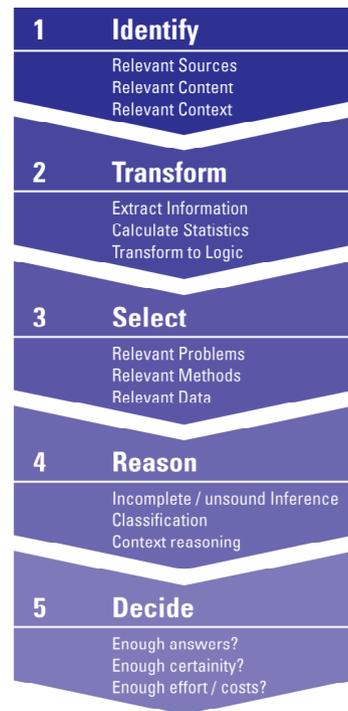


LARKC IN A NUTSHELL

The aim of the **EU FP7 Large-Scale Integrating Project LarkC** is to go beyond the limited storage, querying and inference technology currently available for semantic computing.

The fundamental assumption is that such an infrastructure must go beyond the current paradigms which are strictly based on logic. By fusing reasoning with search and taking seriously the notion of limited rationality we aim for the paradigm shift that is required for reasoning at Web scale.

THE ARCHITECTURE



MAIN INNOVATIONS

The aim of this project is to develop the **Large Knowledge Collider (LarkC)**, for short, pronounced “lark”), a platform for massive distributed incomplete reasoning that will remove the scalability barriers of existing reasoning systems for the Semantic Web. This will be achieved by:

- **Enriching current logic-based Semantic Web reasoning** with methods from information retrieval, machine learning, information theory, databases, and probabilistic reasoning.
- **Employing cognitively inspired approaches** and techniques such as spreading activation, focus of attention, reinforcement, habituation, relevance reasoning, and bounded rationality.
- **Achieve scalability through giving up completeness.** Partial reasoning results are useful in many domains of application. Significant speedups can be obtained by incompleteness in many stages of the reasoning process.
- **Achieve scalability through parallelization.** Different possibilities will be explored either through parallel processes on a high-performance computing cluster, or through wide-area distributed computing (“computing at home”).

RESULTS

We will not build just a single reasoning engine that is supposed to be suited for all kinds of use cases, but instead build a configurable platform on which different components can be plugged in to obtain different scale/efficiency trade-offs, as required by different use cases.

The resulting plug-in architecture will allow researchers and users to experiment with different forms of parallel and approximate reasoning, through building and combining their own plugins.

The platform, together with a set of pre-engineered plugin components, will be demonstrated in three use-cases:

- **Real-time analysis and interpretation of the data concerning a city infrastructure**

With the help of LarkC technology, we are developing an Urban Computing showcase. It will demonstrate the possibility to answer questions like: “Is public transport taking people to where they want to go?”, “What areas in town are attracting more people than normal?”, “Where is the traffic congestion?”.

Although the information is often available, there’s no software system capable of computing the answers and no system enables users even to issue such queries.

- **Semantic data integration for early clinical drug development**

LarkC will provide scientists with better tools to grasp the breadth and depth of relationships between data in early clinical development. The diversity of data and data sources required in early clinical development made us use a semantic knowledge platform, we call the platform “*LinkedLifeData*” (LLD). The LLD will eventually span genes-proteins-pathways-target-diseases-drug-patient data and LarkC will use LLD to facilitate “cross domain” interpretation/reasoning.

- **The production of reference works on carcinogenic risk factors**

LarkC will help carcinogenesis researchers produce definitive evaluations of all published research on a carcinogen, and to study the association between genes and cancer. LarkC will help make available background knowledge, finding mentions of relevant concepts in the biomedical literature, such as genes and diseases.

INDUSTRIAL IMPACT

Professional sectors and industry will benefit from the ability to better utilise more extensive content of greater complexity:

- World Health Organisation and AstraZeneca R&D



LarKC's two life sciences case studies contribute exactly to the goal of a better integration and reuse of crucial medical research information. European high-tech industry in this sector, and the involvement of the United Nations' World Health Organisation and AstraZeneca's research labs will show how the project's results can apply sector-wide.

- Siemens AG

SIEMENS

Siemens business activities in traffic control, intelligent resource planning, infrastructure planning, dynamic energy distribution, location-based services and medical solutions will greatly benefit from the LarKC results. Life science databases are expected to play an ever increasing role in personalized medicine. The analysis of diseases like cancer opens up new opportunities for Siemens healthcare solutions as a service provider.

- CYCORN



The research these organizations are doing with Research Cyc will benefit more or less immediately from early versions of LarKC. CYCORN will make Research Cyc available under Apache 2 or similar license. The first prototype will give the consortium access to a group of early adopters for whom participation in the Early Access Group will be very attractive.

ADMINISTRATIVE DETAILS

START	April 2008
END	September 2011
EC CONTRIBUTION	€7 million
PROJECT COORDINATOR	Dieter.Fensel@sti2.at
SCIENTIFIC DIRECTOR	Frank.van.Harmelen@cs.vu.nl



SEVENTH FRAMEWORK PROGRAMME

PARTICIPANTS



STI INNSBRUCK (Coordinator)
www.sti-innsbruck.at



AstraZeneca
www.astrazeneca.com



CEFRIEL
www.cefriel.it



CYCORN
www.cyceurope.com



WHO
www.iarc.fr



HLRS
www.hlrs.de



Max-Planck Institute
www.mpib-berlin.mpg.de



Ontotext Lab
www.ontotext.com



Saltlux
www.saltlux.com

SIEMENS

SIEMENS
www.siemens.de



WICI
www.iwici.org



University Sheffield
www.shef.ac.uk



Vrije Universiteit Amsterdam
www.vu.nl



*A configurable platform
for infinitely scalable
Semantic Web reasoning.*