SeSam4: Semi-semantic Models for Cross-Sector Portals

Robert Engels and Sesam4 Consortium
IKT-Norge
c.o Fredrik Syversen
Drammensveien 126
N-0214 Oslo, Norway
robert.engels@esis.no

ABSTRACT

Sesam4 has as overall goal to make semantic technology understandable and accessible for organisations and small/medium-sized enterprises. Through its focus on interoperability, common conceptual frameworks and standardised ways for modeling domains, semantic technology offers a competitive advantage to companies that understand how to integrate with these standards.

Information on the Internet is, if at all, often published through proprietary interfaces that require a potential user to adapt to different web-services and content representations. Automation of applications against such information sources is a pain-staking and difficult process that is both error-prone and expensive due to the unstable, dynamic nature of such implementations.

Standardisation of knowledge representation on the Internet has been a topic of research for many years. Many standards (eg. HTTP, URI, HTML, XML, SOAP) are widely adopted and have significantly contributed to the ways in which information has been exchanged on the web. With the current integration of Smart-phones and numerous devices showing good computing power, we find an increasing call for automation of many different, information-intensive tasks. Such tasks depend on a profound and well-defined amount of knowledge being digitally available. Several different knowledge representation languages and schemes have been defined over the years, probably with ISO and W3C standards as the most prominent amongst them.

Sesam4 investigates such standards and methods for how to best apply and utilize them. It does so through the implementation of different software solutions in various application domains. Methodically going through the process of application of semantic technology, the project provides tools, insights and best practice for data conversion, publishing and utilisation.

Brought together under the umbrella of ICT-Norway (IKT-Norge) the consortium consists of renown commercial partners like Computas, Cyberwatcher, TextUrgy, InfoSector, ESIS Norge, Ovitas and academic research partners such as Vestlandsforsking (Western Norwegian Research Institute), the Norwegian Computing Center (NR) and Bergen University.

Keywords

Semantic Web, Topic Maps, Ontology, Tourism

1. PROJECT RESULTS

With quite ambitious goals, including a long list of software to be delivered during the course of the project, the consortium has managed to deliver good results covering the whole process for evaluation and application of semantic technologies for use in organisations and smaller enterprises. Nearly all results of the project (not including some internal documents and published articles) are available freely through the Internet and can be found on the project portal http://www.sesam4.net.

Project results can be divided in dissemination-, research & development-, demonstration- and education results.

Dissemination results

During the course of the project the Sesam4 consortium has published and presented a diversity of work on semantic technology. This ranges from, amongst others, presentations for political groups ("reiselivspolitisk utvalg, Sogndal 2009), to academic presentations at (inter-)nationally renown conferences like Semantic Days (2010/2011), GoOpen (2008/2009), Emnekart (2009), ENTER 2010 and ONAV 2010.

Several workshops and educational events have been conducted during conferences. Some hands-on sessions open to anyone where also held by several consortium members.

In addition to the academic conferences there have been a large variety of business oriented presentations and work groups where either software or applications have been discussed and even further developed. Dissemination material (design, web, brochure & roll-ups) have been very actively used and partners have contributed to their best efforts regarding dissemination activities.

Research & Development

Several research activities have taken place within the project. The consortium has managed to take care of delivering ground research into issues of interoperability, distributed implementations and contribution to the arising SPARQL recommendations with research into the actual benefit of this paradigm and technology for business areas like Tourism, B2B information management and construction. Also the field of Natural Language Processing and automated ontology extraction and population has been in focus. A few partners have distinguished themselves within these work packages by

delivering software and complete (open-source) frameworks for use and evaluation within the project and which are now disseminated within their respective communities (Open Source frameworks like Sublima (Computas), COMPASS (Ovitas) and TextUrgy OpenBergen Tagger (TextUrgy)). A major outcome of this part of the project is really the application and implementation of academic efforts (like W3C standardisation of RDF/OWL and ISO's TM) and filling the gaps in the application process with actual software and tools. Many of these tools have gone from a research/prototype stage into an application phase through Sesam4 and contributed to an enhancement of knowledge of the state-of-the-art within the consortium and amongst the public.

Demonstrators

In order to make the technology and its benefits understandable and accessible, the original work plan contained two main demonstrators (tourism & construction). For several reasons it was decided that the construction demonstrator had to be reconsidered (main reason being the maturity of the construction end-users, which was low at the time). Instead, the consortium focused on applications and demonstrations of the software independent from business areas (SUBLIMA, COMPASS, OpenBergen Tagger, SemEx lifting tools, Folksonomy learners), and within two business areas (Tourism and Company Market Information). In addition to this, the consortium has also converted various interesting data-sets into the framework of semantic technology which are available through several so-called SPARQL-endpoints for use within third-party software and apps (CX-LOD server by Computas, the NFR project archives by ESIS). These data is than re-used in various applications and also used in the education activities so that the recognition and re-use experience for people wanting to learn about the technology is maximised.

For tourism we ended up with four good demonstrators of the technology OnBoard (Vestlandsforking), Place Companion (Computas), What's On? (Norsk Regnesentral) and Digitur (Unifob). Demonstration movies of applications running on West coast Ferries and iPhones and some live demo portals are available through the Sesam4 portal.

The Company Market Information demo, also featuring enhanced Natural Language Technologies, is also available as a live demo portal through the Sesam4 portal pages.

Educational results

On the educational side a set of tutorials and technology primers have been produced and made available free of charge on portals like Slideshare, Youtube. The consortium has focused on three areas: RDF/OWL (as representation formalisms for the Internet), SPARQL (as query language for information on the Internet) and Linked Open Data (as application area for these technologies). In addition there are white papers and academic papers, all of which are linked through the SESAM4 portal. The hands-on sessions conducted at several conferences have also contributed very positively to a greater awareness for these technologies amongst the public (e.g. a hands-on session at Software 2009 leading to more such sessions during later years)

Results after the projects' end – momentum caused by SESAM4

Very well aware that project results should not “die” at project end, the consortium has defined a strategy to get momentum up also after the project ended. This strategy consisted of three parts. The first is dissemination and easy & open access to project results. This is taken care of by publishing movies/presentations software and slide sets on portals like YouTube, SourceForge, Github and Slideshare, so that the resources are allowed to be taken up and evolved by communities and individuals. Secondly, the project contributed to and stimulated initiatives like a “hackers bootcamp” for specific parts of semantic technologies with as aim real software contributions to the global community (upcoming: The academic conference Web Intelligence, Mining, and Semantics - WIMS 2011, in Sogndal 25th - 27th of May). Finally, there is the issue of education and software re-use. To this end, several courses have been set up by partners which use/re-use project results, demonstrators and the portal content.

Taking everything together, the project has been rather successful in its main goal: making difficult technology better understood and accessible. There has been a good and fruitful cooperation between academic and business environments within the project, where partners more than once played unexpected roles (Universities showing important insights in business aspects and commercial organisations at the forefront of international R&D). This fact has led to intensive and very fruitful debates, where the outcomes have been disseminated as good as possible. Both commercial and academic partners within the consortium have signaled reuse and further development within their own organisations, or in other (inter-)national consortia.

Some of the developed software is now commercially and non-commercially in use at a variety of places (ranging from oil companies to tourism and libraries). The datasets that are made available have led to an increased demand for more datasets to be opened within the Linked Open Data paradigm, an initiative which is followed up by several consortium members.