SEKT – Semantically Enabled Knowledge Technologies

York Sure[1], John Davies[2], Rudi Studer[1], and Paul Warren[2] *on behalf of the SEKT consortium*
[1] Institute AIFB, University of Karlsruhe, Germany (*Technical Coordinator*)
{sure,studer}@aifb.uni-karlsruhe.de
[2] BT Exact Technologies, Adastral Park, Ipswich, UK (*Coordinator*)
{john.nj.davies,paul.w.warren}@bt.com

Abstract

Progress in the development of the Information Society has seen a truly revolutionary decade, with the advent of widespread networked information services. A new social revolution is imminent, involving the transition from Information Society to Knowledge Society. The EU IST project "Semantically Enabled Knowledge Technologies (SEKT)" (see http://www.sekt-project.com/) aims to position the EU IT industry at the core of this, by developing the essential semantic knowledge technologies for realizing the European Knowledge Society.

1 Introduction

SEKT is a European funded integrated project (IP) in the 6th Framework Program (FP6) which runs from 01/2004 to 12/2006.

The vision of SEKT is to develop and exploit the knowledge technologies which underlie Next Generation Knowledge Management. We envision knowledge workplaces where the boundaries between document management, content management, and knowledge management are broken down, and where knowledge management is an effortless part of day to day activities. Appropriate knowledge is automatically delivered to the right people at the right time at the right granularity via a range of user devices. Knowledge workers will be empowered to focus on their core roles and creativity; this is key to European competitiveness.

The SEKT strategy is built around the synergy of the complementary know-how of the key European centres of excellence in Ontology and Metadata Technology, Knowledge Discovery and Human Language Technology, besides major European ICT organisations.

Specifically, SEKT is delivering software to: semi-automatically learn ontologies and extract metadata, and to maintain and evolve the ontologies and metadata over time; to provide knowledge access; besides middleware to effect integration of all the SEKT components. SEKT is also developing a methodology for using semantically-based knowledge management. The software components and the methodology are evaluated and refined through three case studies, in the legal, telecoms and consultancy industries.

SEKT is also undertaking a programme of dissemination. This is aiming not just at specialist technical communities, but also at wider information technology management and general management communities. Some of this programme will be in concertation with other projects in the semantically-enabled knowledge systems strategic objective such as the SDK Cluster[1].

2 Project objectives

2.1 Overall Scientific Goal

The vision of this project is to develop and exploit the knowledge technologies which will underlie Next Generation Knowledge Management (NGKM). These NGKM systems will include means for automated knowledge extraction, knowledge packaging and delivering according to user profiles as well as semantic-based knowledge analyses and matching for user-driven knowledge push. A major barrier to a widespread use of such NGKM systems in industry and organizations arises from the necessary overhead of knowledge modelling and annotation.

SEKT will address these and other NGKM challenges by an interdisciplinary approach focussing on substantially reducing the overhead of knowledge modelling and annotation of sources. This will be done by integrating Ontology and Metadata Technology (OMT), Human Language Technology (HLT), and Knowledge Discovery (KD) into a uniform and scalable framework that supports the integrated learning and management of ontologies and metadata in a (semi-) automatic way.

The reduction in overhead time will be measured by comparing manual approaches with the use of the semi-automatic tools developed in SEKT.

In brief, the technologies will be combined to produce semi-automatic tools for the creation of ontologies, the population of those ontologies with medadata, and the maintenance and evolution of the ontologies and associated metadata. A knowledge access tool will also be developed to make use of the ontologies and associated metadata to meaningfully access the knowledge. Middleware software, with published APIs, will enable the suite of tools to interoperate.

Specifically, the use of ontologies and metadata underlies the SEKT components, and the whole approach; human lan-

[1] http://www.sdk-cluster.org/

guage technology will be used to extract metadata; knowledge discovery will be used to semi-automatically learn and evolve ontologies.

Further RTD objectives of SEKT are

- Providing seamless context-aware access to knowledge, including innovative visualization techniques and context-aware personalised push services.
- The integration of information from different, heterogeneous sources based on ontology mediation techniques. Searches defined with respect to one ontology will need to identify information annotated using a different ontology; whilst searches will identify information sources annotated using different ontologies and it will be necessary to make inferences from these different information sources.
- The development of methods for reasoning under inconsistency.
- The integration of knowledge management with normal business processes in a seamless way thus making knowledge input a side-effect of doing normal business tasks.

SEKT will meet these technological objectives by an integrated approach that ranges from foundational research (the development of new well-grounded (formal) methods), to component-level research (the development of next generation tools based on these methods), and finally to system-level integration (the realization of new semantic-based KM solutions in several case studies). Thus, SEKT is based on the integration of fundamental research, component development and integration driven by real world use cases, that span the three complementary technology areas of SEKT.

At the end of the project the SEKT technologies will have been applied and evaluated in in several case studies. The technologies will be in a state ready to be incorporated into commercial products. There will also be a comprehensive methodology which includes guidelines and best practices from the case studies. The SEKT technology will be ready to be extended by other technology being developed in the meantime, e.g. semantic web services as well as security and trust. On the other hand, the SEKT technology will be available for transfer into other technology areas, like semantic web services, and to further application scenarios.

2.2 Technologies

NGKM solutions will be built upon ontology-based metadata and thus the creation and management of machine-interpretable information and the consequent use of ontologies. The integrated management of ontologies and metadata, especially their generation, mediation and evolution, is fundamental to this development and relies in part on innovative HLT and KD methods. Experience shows, e.g. with the UNSPC product classification ontology (`http://www.daml.org/ontologies/106`) which consists of 10.000 concepts, that the effort for generation, mediation and evolution – and thus the complexity – is directly related to the size of the ontology.

Advanced reasoning capabilities will strongly support the evolution of ontologies and metadata and greatly reduce the overhead for maintenance. Work in advanced reasoning will include the development of techniques for robust reasoning, i.e. reasoning in the presence of inconsistencies, i.e. in order to give meaningful results even when the overall ontology has conflicts. It will also include flexible reasoning which can cope with changes and conflicts in a given model and can fall back to old versions or change the scope of reasoning to a consistent set of statements. The advanced reasoning work will support the evolution of ontologies and meta-data, in order to reduce maintenance overhead.

In this project HLT has two key contributions that will both include the handling of multilinguality. In the first place it will be used to semantically annotate informal and unstructured knowledge. Thus, the automatic or semi-automatic extraction of metadata from legacy data will be achieved. Secondly natural language processing will be used to generate natural language based on formal knowledge (ontologies and metadata). Here, HLT will be strongly integrated with methods from KD and OMT. The ontologies that structure metadata are in many cases language-independent to a significant degree.

Knowledge discovery is also a key technology in its own right for knowledge management, helping to provide knowledge workers with the knowledge they need from large volumes of data. At the same time, like natural language processing, knowledge discovery has a role in metadata extraction to enable the automatic or semi-automatic mark-up of data, and in ontology learning and evolution. Obviously, KD interacts closely with HLT for e.g. metadata generation, and ontologies provide the background knowledge for improving KD results.

3 Mission: Explore Synergies

In order to fully benefit from these technologies, they must be used together. This convergence is now timely because of the maturing of the 3 separate disciplines, particularly ontology technology, which has received much attention over the last 2-3 years. Thus, in SEKT, the ontology learning software based on knowledge discovery techniques will develop ontologies which will be populated with metadata using software employing human language technology. These ontologies and their corresponding metadata will be managed and evolved using enhanced ontology and metadata technology developed in SEKT.

In turn, the ontology evolution software will benefit from knowledge discovery techniques. Inferencing technology will be developed to inference over the ontologies and mediation software will be used to align, merge and translate between different but related ontologies. The use of KD and HLT to generate ontologies and annotate data respectively, will lead to faster (and therefore lower cost) production of semantic-based systems. The use of semi-automatic approaches will also lead to more accurate and complete ontology and metadata descriptions since the human view will be augmented by a machine-aided process which will help to neutralise any effect of a specific human perspective on the domain and data.