KW Partner: FU Berlin

1. Overview

Challenge
Finding data in a large institution can be difficult and time consuming
Solution
Knowledge management solutions allow to classify data and improve precision and recall.
Why a Semantic solution
This classification can be made according to a shared conceptualization of the domain, i.e. an ontology. Additional expressability through rules allows for more powerful search and retrieval based on reasoning.
Key Business Benefits
Employees find relevant or required data more quickly, which leads in total to significant time and cost savings for the enterprise.

Keys components

Existing Software
Underlying storage e.g. relational database

Research and development
Ontology development
Semantic portal technology
Access rights
Rules engine

Technology locks
Ontology creation tools supporting good ontology modelling
Integration of semantic technologies within the business IT infrastructure
Rule creation and testing

DaimlerChrysler is a large enterprise with departments spread worldwide and over 360 000 employees in total. A significant volume of data is being generated daily through business activities and this data may not only be relevant to the originating department at the time of generation but also to other sectors of the enterprise at any time in the future.

Knowledge management is a requisite for employees to locate the data that is relevant to their search. However currently much time is spent in locating that data from the large body of corporate data that is available. There is a need for higher precision and better recall so that employees find the required data more quickly.

As a result, an internal Semantic Portal is being implemented which organizes the data according to a shared conceptualisation of the domain, i.e. an ontology, of the different aspects of the enterprise (e.g. employees, products, processes).

2. Current Practices and Technologies
2.1 Typical Business Practices

The DCVD Semantic Portal is implemented on the company intranet as a research prototype. Later it will be available to employees of Daimler Chrysler to insert, navigate and search for corporate data.
It contains a navigation structure for manual location of data based on the enterprise ontology as well as a semantic search facility. Navigation is along the taxonomic structure of the ontology (sub/superclasses) and for a selected class it lists the known instances. Individuals can be selected and examined in terms of their properties and relations to other individuals.

The search returns both direct matches as well as 'related' matches. The semantic matching and ranking is based on a direct match to classes or individuals and then a ranking based on subsumption (i.e. other relevant search results are returned based on their relationship to a superclass or subclass of the queried concept). In addition, rules have been modelled to describe knowledge that is not expressible at the ontological level. These rules form the basis for additional inferences so that other relevant matches are determined for the query.

2.2 System requirements Analysis

The portal implementation involved the building of an enterprise ontology from scratch. There is a need for better ontologies in the sense of maintaining consistency and following modelling guidelines, which in turn requires ontology development tools which are able to guide the ontology developer through the process. Correspondingly, the effort required to produce ontologies needs to be reduced.

Another requirement identified from the portal development is the need to integrate semantic technologies transparently into the business IT infrastructure. Rather than requiring a new set of technologies, existing systems should be extended as transparently as possible to use semantic technologies with the end user benefiting from the introduction of the use of semantics without being required to learn new tools or processes.

A key need is for access rights. This must be integrated into the business infrastructure in both semantic and non-semantic systems. Semantics could be very useful in this case in that the appropriate access of a user to some data could be inferred from the user's position and the data's status e.g. a project manager would reasonably be given read and write access to project data while other users who are involved in similar or related projects may be granted read access. Such 'intentional rights' would also need to take into account new access rights when a user changes their position within the enterprise (transferred to another department, taken off a project, etc).

The determination of relevant matches needs to be extended to take into account the complexities of the real world domain. For example, as matches on superclasses can be included within the bounds of relevant matches, a query on who is responsible for a particular problem may return not only the individual directly responsible but also his boss or the bosses' boss – yet it is clear that actually the big boss is unlikely to be interested in this small problem!

These complexities can be further modelled using rules. However there is a need for an easier means to model such rules for non-logicians and better tools to test and debug rule bases (collections of rules applied to some knowledge).

2.3 Review of the current systems

The current corporate Intranet has some structure for the navigation to data and a traditional search engine. In its current form data is linked to at most one category meaning that the user's location of the desired data is heavily dependant upon choosing the right category and other (relevant) data which has been placed into other categories is not found. Correspondingly, a search for data which transcends a simple flat categorization is limited to the traditional keyword search with its ambiguities and inconsistencies (e.g. when a keyword is input but spelt incorrectly). Finding e.g. 'all data related to Mercedes-Benz' is restricted to those documents which specifically mention the car make, and excludes others which are related but do not directly mention the car make.