Infraestructura tecnológica de servicios semánticos para la Web Semántica
TIN2004-02660

Gómez Pérez, A.; Aguado de Cea, G.; Álvarez de Mon Rego, I.; Barrasa Rodríguez, J.; García Castro, R.; González Cabero, R.; López Cima, Á.; Pareja Lora, A.; Ramos Gargantilla, J.A. and Suarez-Figueroa, M.C.; Villazón Terrazas, B.
Universidad Politécnica de Madrid

Corcho, O.
University of Manchester

Fernández-López, M.
Universidad San Pablo CEU

Manuel Lama Penin
Universidad de Santiago de Compostela

Abstract

This project aims at creating a network of distributed interoperable semantic services for building more complex ones. These services will be available in semantic Web service libraries, so that they can be invoked by other systems (e.g., semantic portals, software agents, etc.). Thus, to accomplish this objective, the project proposes:

a) To create specific technology for developing and composing Semantic Web Services.

b) To migrate the WebODE ontology development workbench to this new distributed interoperable semantic service architecture.

c) To develop new semantic services (ontology learning, ontology mappings, incremental ontology evaluation, and ontology evolution).

d) To develop technological support that eases semantic portal interoperability, using Web services and Semantic Web Services.

The project results will be open source, so as to improve their technological transfer. The quality of these results is ensured by a benchmarking process.

Keywords: Ontologies and Semantic Web

1 Project objectives

This project aims at creating a network of distributed interoperable semantic services and thus to build more complex ones. These services will be available in semantic Web service libraries, so that they can be invoked by other systems (e.g., semantic portals, software agents, etc.). In order to accomplish this objective, the project proposes to achieve the following sub-goals:

1. Methodological and technological support for the creation of semantic Web services.

Specifically, we will have to work on the design, discovery, and (semi)automatic composition
of semantic Web services and in the management and building of distributed semantic Web
service libraries.

2. Interoperability and management of semantic portals. We propose an architecture that
facilitates the interoperability between semantic portals through Web services and semantic
Web services.

3. Basic research and development of an infrastructure in the ontology area. We will
develop new mapping and evolution services, extend the ontology learning services, and
transform the evaluation service.
   a. Creation of a mapping service between heterogeneous knowledge models. It will be
      created a mapping specification language, tools for mapping development and
      management, and a method for automating mapping discovery.
   b. Creation of the ontology evolution management service, by identifying the
      components to be managed and providing software support.
   c. Extension of the automatic ontology learning service resulting from the TIC-01-
      2745 project. It will use standardized linguistic sources and information from
      distributed databases that will be accessed via ontology mappings. The extensions will
      consist on the improvement of concept and taxonomy learning, and on the learning of
      attributes, relations, and instances.
   d. Transformation of the evaluation service into an incremental evaluation service
      that analyses the changes performed in the ontologies and determines their consistency.

4. Improvement of the METHONTOLOGY ontology development methodology and
adaptation of this methodology to the new service paradigm. To be precise, the methods
to perform knowledge acquisition, integration, evaluation, configuration management, and
maintenance will be defined in detail and METHONTOLOGY will be extended, improved,
and adapted to the new distributed semantic service paradigm.

5. Decentralization of the current services of the WebODE platform and its transformation to
semantic services that can be used by other platforms and/or applications. Migration of the
WebODE platform to open source to favor the development of the Semantic Web and of
semantic-based applications, and therefore to improve technology transfer to other national
and international research groups as well as to companies and organizations. Concretely, we
will redesign and reimplement some parts of the source code, obtain a continuous quality
improvement through the establishment of a continuous benchmarking process that involves
the development of supporting technology, and we will generate documentation, installation
and user manuals, a list of most frequent errors, management of mailing lists, etc.

2 Main achievements of the project

2.1 Methodological and technological support for the creation of semantic Web services

At the beginning of the project, we had a conceptual model for describing semantic web and
semantic grid services. We have specified, designed and developed an open source graphical user
interface for annotating and designing semantic services for the Grid and for the Web that uses
ontologies and Problem-Solving Methods for describing the features of service operations.
Choreography and orchestration of semantic services are represented by workflows that model the
coordination between the services involved in the choreography and/or orchestration. In our
approach, these workflows are described with High-Level Petri Nets (HLPN): each one of the
workflow primitives (such as and/or-split, and/or-join, etc.) are modelled by a given HLPN.
As semantic services are described by means of ontologies, we have developed a workflow ontology (WO) to represent service choreography and orchestration. The core of the WO is an ontology of HLPN, which describes the semantics of both the dynamic and static models of the HLPN formalism: we have developed a concept taxonomy where the HLPN components are described and a set of axioms for constrain checking between the taxonomy concepts. This HLPN ontology is based on the ISO/IEC 15909-1 standard that compiles the shared knowledge of the Petri net community about HLPN.

2.2 Interoperability and management of semantic portals

The starting point at the beginning of the reporting period was the design of a conceptual architecture for semantic portal interoperability and data exchange in ODESeW. During this period, the ODESeW architecture has been refined.

The implementation efforts were devoted to the external gateway component, whose goal is to access an external resource that contains information according the domain ontologies of the portal and to annotate or translate this information according to the domain ontologies. This activity makes use of results from other parts of the project. As ODESeW employs the WebODE ontology development platform for storing and managing multiple ontologies, data exchange in ODESeW strongly depends on WebODE import and export facilities, which are being benchmarked in this project as seen below.

2.3 Basic research and development of an infrastructure in the ontology area

Regarding the creation of a mapping service between heterogeneous knowledge models, the tasks that have been performed cover the following types of mappings: mappings between two ontologies and mappings between ontologies and databases.

In the case of mappings between two ontologies, the different casuistry has been analysed and the actions for carrying out the merge process have been established, though only for the limited case of concepts, attributes and relations. Research has been carried out in the use of linguistic tools for the automatic discovering of mappings between ontologies, as appears below in this paper. Furthermore, for the future mapping representation the existing representations have been analysed and a first version of a mapping representation ontology has been developed.

A WebODE Mapping service has been created for browsing and editing mappings. Currently, the service only shows mappings from the automatic discovery implemented.

In the case of the mappings between ontologies and databases, the group developed R2O, an extensible and fully declarative language that describes mappings between relational database schemas and ontologies. Once mappings are defined in a R2O mapping document, they are processed automatically by the ODEMapster mapping processor to populate an ontology (batch process) or to translate a query (on demand).

The work on the ODEMapster engine has been extended to fully cover the on-demand mapping scenario. A query defined in terms of the elements in the ontology (expressed in ODEMQL) is processed and translated in terms of the underlying database and the results of the dynamically generated relational query are translated into instances of the ontology (expressed in RDF). Compliance with standard query languages such as SPARQL is under development.

A module for implementing the Semantic Mapper algorithm for automatic database schema to ontology matching discovery has also been developed. Such module takes as input the relational schema of an existing database and an OWL ontology and delivers as output an skeleton of an R2O mapping document containing the matches found.
We are also extending ODEMapster and enriching R2O language to describe mappings between XML schemas and ontologies.

As for the extension of the automatic ontology learning service resulting from the TIC-01-2745 project, the tasks performed are not just related to the ontology learning but also to the ontology mapping discovery.

We have studied the use of the SUBCLASS-OF relation in texts. More than 60 linguistic patterns have been identified and systematized according to the main lemma, taking into account that each lemma can present different word-forms. In this case the patterns have been analyzed according to their word-forms.

Furthermore, 30 lexical collocations that provide semantic information related to concept extraction and 10 metalinguistic devices that help extract the SUBCLASS-OF relation have been identified; in addition, 25 linguistic pattern-based rules have been defined from the patterns identified for each lemma, and for each variant in these lemmas, in order to formalize them and achieve a (semi)automatic extraction of the concepts and relations in the texts found in the Web.

We have identified the localization of a new component in the architecture that can exploit the use of rules for knowledge extraction, and that can be used for ontology learning and for discovering mappings between several ontologies. We have started the implementation of this information extraction component in ODELinger. We must add that the OntoTagger-ODELinger communication was outdated. In the last few months, this communication has been updated and the ODELinger database model has been extended for covering semantic annotation information.

Meanwhile, an automatic mapping discovery tool has been implemented into WebODE that uses EuroWordNet to create mappings between concepts with the synonymy semantic relation.

2.4 Improvement of the METHONTOLOGY ontology development methodology and adaptation of this methodology to the new service paradigm

The work performed is focused on the notion of networked ontologies, analyzing different dependencies between the activities of the development process of several ontologies. Based on such analyses, the development process proposed in METHONTOLOGY has been modified. The activities involved in the development of networked ontologies (conceptualization, merging, documentation, evaluation, etc.) are being defined within the new notion. We have created a consensuated glossary of activities to be used in the improved development process.

Research has also been carried out on the use of formal ontologies in the design, evaluation and integration of ontologies and in the analysis of the potential of Web 2.0 technologies to build ontologies collaboratively.

2.5 Migration of the WebODE platform to open source

The development of WebODE OS has started. WebODE OS is an ontology development and management platform extensible, scalable and open source. The WebODE OS platform is being developed using J2SE 1.5 over JBoss, an application server that follows the J2EE 1.4 standard specification. The software is being developed platform-independent (Linux and Windows) and database independent (MySQL and Oracle). The knowledge model of WebODE OS has been developed and implemented in EJB 3.0; the ontology management API has also been implemented. During this process, unit tests have been performed.

Simultaneously, the FPI grant holder has started the benchmarking of the WebODE platform.
Taking into account the current State of the Art of software benchmarking, we have defined the process to follow for benchmarking WebODE, considering the different existing methodologies on software and software process improvement of the Benchmarking, Experimentation in Software Engineering and Measurement in Software Engineering areas. For each of these areas, we have identified the main definitions, classifications proposed in the literature, and the tasks of the methodologies.

From the State of the Art previously cited, we have defined the process and the criteria for benchmarking the WebODE platform and its services. One of the main factors taken into account when defining the process has been the continuous improvement of the WebODE platform.

This benchmarking has been planned with the goal of assessing and improving the interoperability of WebODE with the main ontology development tools and ontology repositories. The benchmarking will evaluate the degree of interoperability in terms of information addition or loss that occurs when two tools interchange an ontology by translating it to an interchange language. The interchange languages considered are the languages recommended by the W3C, namely RDF(S) and OWL.

Furthermore, different software has been developed to support the execution of the experimentation in the benchmarking and the analysis of the results obtained in the benchmarking.

3 Result indicators

3.1 Personnel being trained

Two grants have been assigned and charged to the project:

- A FPI grant (Beca Predoctoral de Formación de Personal Investigador) has been assigned to Raúl García Castro for 48 months (BES-2005-8024).
- A project grant has been assigned to José Ángel Ramos Gargantilla for 9 months.

A research stay related to the project:

- Mariano Fernández López had a research stay at UPM from 1st January 2006 to 1st February 2006 to work in the creation of a mapping service between heterogeneous knowledge models and in the improvement of the METHONTOLOGY ontology development methodology.

3.2 Dissemination

A summary of the dissemination activities performed in this project is the following: 2 papers in international journals, 14 papers in international conferences, 2 posters in international conferences, 3 papers in national conferences, 5 papers in international workshops, 2 papers in national workshops, 7 book chapters, and 1 PhD thesis. One of these posters was awarded as Best Poster in the WWW2006 international conference.

Furthermore, technology demonstrations have been carried out for the following companies or socio-economic agents: Intelligent Software Components, Atos Origin, TPI, Telefónica Sistemas, Gabinete de Investigación Militar Operativa (GIMO), Instituto Geográfico Nacional, and Parquesof (Colombia).

3.3 Technology transfer

The semantic portal technology is being used to develop semantic portals in the European projects such as Agentlink III, NeOn and OntoGrid, in the Network of Excellence Knowledge Web, and in the Spanish network Red Temática de Web Semántica. It has also been used to develop the
semantic portal of this project. The Knowledge Web, OntoGrid and NeOn portals include a functionality to help project management and to create progress reports. With this idea, we have proposed the creation of a spin-off company for managing EU projects based on this functionality built with ODESeW. The proposal won the “XIII Programa de Creación de Empresas de Base Tecnológica UPM”, a competition at UPM.

There also exist with the group on Sheffield University lead by Prof. Dr. Fabio Ciravegna a contract for using the ODESeW technology in the management of European projects. The R2O technology is being used in the SEEMP (Single European Employment Market-Place) European project and in the NeOn European project.

A collaboration contract has been signed with Spanish Instituto Geográfico Nacional (IGN) to develop an ontology by refining and extending BCN25, and to develop an algorithm for the automatic identification of mappings between the ontology, nomenclators and the IGN cartographic databases.

3.4 Participation in international projects

The research group participates in the following projects of the Sixth European Framework Programme:

- NeOn: Lifecycle Support for Networked Ontologies (FP6-027595).
  Type of participation: consortium member. From 01-03-2006 to 28-02-2010.
- Single European Market Place SEEMP (FP6-027347)
  Type of participation: consortium member. From 01-01-2006 to 30-06-2008.
- OntoGrid: Paving the way for Knowledgeable Grid Services and Systems (FP6-511513)
  Type of participation: project coordinator. From 01-09-2004 to 31-08-2007.
- Knowledge Web (FP6-507482) Network of Excellence
  Type of participation: consortium member. From 01-01-2004 to 31-12-2007.

Additionally, it must be highlighted the participation of the research group in the Spanish thematic network that follows below, whose leader is the project coordinator and in which 24 Spanish universities are participating with 180 researchers:

- Red Temática de Web Semántica (TSI2006-26928-E)
  Type of participation: project coordinator. From 01-01-2007 to 31-12-2007.

3.5 Collaboration with research groups

The Ontology Engineering Group collaborates with all the universities and companies that participate in the projects presented above. These partners are from Austria (1), Belgium (2), France (4), Germany (7), Greece (3), Ireland (1), Italy (8), The Netherlands (3), Poland (1), Slovenia (1), Spain (28), Switzerland (1), and United Kingdom (5).

Besides these collaborations, this project has provided the following collaborations with other non-academic entities:

- Several demonstrations to non-academic entities have been performed. Also, the group has applied for three CENIT projects with companies that have received technology demonstrations.
- The idea of networked ontologies has led to the creation of an IP European project called NeOn (FP6-027595).
The proposed architecture for semantic portal interoperability will be used to interchange content between the NeOn Project portals. These portals are now being developed by the Atos Origin company and by the UPM.

The WebODE ontology development platform is used by the Instituto Geográfico Nacional to develop ontologies in the domain of Geographic Information Systems.

This project has provided the following collaborations with other research groups:

- Collaboration with the Instituto de Lingüística Aplicada (IULA) from Universidad Pompeu Fabra.
- Contact with teachers from Bristol University in the BAAL Conference to present the work results.
- Meeting with Dr. Alejandro Curado from Universidad de Extremadura and a group of teachers from different universities (Valladolid, Granada, Valencia) to present the results of the linguistic section.
- Collaboration with the research center Centrum voor Vaktaal en Communicatie (CVC), led by Dra. Rita Temmerman, from Erasmushogeschool in Brussels.
- Collaboration with the research group coordinated by Dr. Ricardo Mairal, professor at the UNED and with Dr. Francisco Ruiz de Mendoza, professor at the University of La Rioja.
- Collaboration with Dr. Enrique Alcaraz Varó, professor and director of Instituto Universitario e Lenguas Modernas Aplicadas (IULMA) Alicante.
- Contact with Dr. Carlos Subirats, Professor at the Universidad Autónoma de Barcelona.
- The group coordinated by Dra. Alcina Caudet, Universitat Jaume I, Castellón.
- Antonio Pareja was invited to deliver a seminar at the Universitat Jaume I, in the summer course “Terminología e Ingeniería Lingüística”, from 10 to 13 July 2006.

During 2006 the NLP group has received two visiting scholars: Dra. Alcina who stayed three months, and Dña. Rosario Bautista, from the University of Malaga, who stayed two months.

4 References

Creation of semantic Web services

Interoperability and management of semantic portals

Mappings between heterogeneous knowledge models
TIN2004-02660


Extension of the automatic ontology learning service


[22] Aguado, G. & Alvarez de Mon "Corpus linguistics applied to knowledge extraction for ontological engineering: the phraseology of classification from English into Spanish", to be published by Equinox.

**Improvement of the METHONTOLOGY ontology development methodology**


**Benchmarking ontology development tools**