



D7.4 Experiments and evaluation of testbed 2

Synthesis Report

Farid Cerbah (Dassault Aviation)

Abstract

This report presents the evaluation of TAO technology conducted in WP7 case study. The goal of the case study is to apply the transitioning support to semantic technologies developed in the project on three inter-related scenarios in the domain of aircraft maintenance. We provide a systematic assessment from several dimensions, including the transitioning help provided to acquire the needed semantic resources, the interoperability of tools and services, knowledge reuse and information access capabilities. The results of this evaluation are intended to provide a better understanding of the level of maturity reached with the TAO technology and to identify the relevant orientations with respect to the category of applications considered in case study 2.

EU-IST Specific targeted research project (STREP) IST-2004-026460 TAO

Deliverable D7.4 (WP 7)

Keyword list: Semantic Modelling, Ontology Acquisition, Data Mining, Semantic Annotation, Semantic Web Services, SOA, Aircraft Maintenance

WP7 Transitioning Business Process Applications

Nature: **Report**

Dissemination: **PU**

Contractual date of delivery: **28 Feb 2009**

Actual date of delivery: **12 Apr 2009**

Reviewer: **Internal review at Dassault**

TAO Consortium

This document is part of a research project partially funded by the IST Programme of the Commission of the European Communities as project number IST-2004-02646

University of Sheffield

Department of Computer Science
Regent Court, 211 Portobello St.
Sheffield S1 4DP
UK
Tel: +44 114 222 1930
Fax: +44 114 222 1810
Contact person: Kalina Bontcheva
E-mail: K.Bontcheva@dcs.shef.ac.uk

Mondeca

Mondeca
3, cité Nollez
75018 Paris
France
Tel: +33 (0) 1 44 92 35 03
Fax: +33 (0) 1 44 92 02 59
Contact person: Jean Delahousse
E-mail: jean.delahousse@mondeca.com

University of Southampton

Southampton SO17 1BJ
UK
Tel: +44 23 8059 8343
Fax: +44 23 8059 2865
Contact person: Terry Payne
E-mail: trp@ecs.soton.ac.uk

Sirma Group Corp., Ontotext Lab

Office Express IT Centre, 5th Floor
135 Tsarigradsko Shosse Blvd.
Sofia 1784
Bulgaria
Tel: +359 2 9768 303
Fax: +359 2 9768 311
Contact person: Atanas Kiryakov
E-mail: naso@sirma.bg

Atos Origin Sociedad Anonima Espanola

Dept Research and Innovation
Atos Origin Spain, C/Albarracin, 25, 28037
Madrid
Spain
Tel: +34 91 214 8835
Fax: +34 91 754 3252
Contact person: Tomás Pariente Lobo
E-mail: tomas.parientalobo@atosresearch.eu

Dassault Aviation SA

DGT/DPR
78, quai Marcel Dassault
92552 Saint-Cloud
Cedex 300
France
Tel: +33 1 47 11 53 00
Fax: +33 1 47 11 53 65
Contact person: Farid Cerbah
E-mail: Farid.Cerbah@dassault-aviation.com

Jozef Stefan Institute

Department of Knowledge Technologies
Jamova 39
1000 Ljubljana
Slovenia
Tel: +386 1 477 3778
Fax: +386 1 477 3131
Contact person: Marko Grobelnik
E-mail: Marko.Grobelnik@ijs.si

Executive Summary

This report presents the evaluation of TAO technology conducted in WP7 case study. The goal of the case study is to apply the transitioning support to semantic technologies developed in the project on three inter-related scenarios in the domain of aircraft maintenance.

Our evaluation concerns both the applied transitioning processes and the resulting semantic resources and applications. We provide a systematic assessment from several dimensions, including the transitioning help provided to acquire the needed semantic resources, the interoperability of tools and services, knowledge reuse and information access capabilities.

An analysis from the main evaluation dimensions

We identified several dimensions that allowed a precise evaluation of both the transitioning operation (and support) and the resulting semantic-based application. These dimensions are directly related with the business needs that semantic technologies are intended to address.

1 Transitioning help

The transitioning approach adopted in TAO involves the construction of knowledge rich applications. Without a significant support to ease the construction of the required semantic models and annotations the proposed approach can hardly convince adopters who are willing to migrate and improve their applications but often at low cost. This aspect of effort reduction has been a major concern in the project and active help has been provided at different steps of the transitioning processes.

Ontology Acquisition from Legacy Databases

The most prominent need is ontology acquisition. Given that the main data source in case study 2 is a relational database, one of the first initiatives taken was to look for existing software support for deriving ontologies from relational databases. Surprisingly, even though relational databases are valuable sources for ontology learning, we found that the existing methods and tools were far from being able to produce results at the level of the business expectations expressed in this case study. Those past few years, efforts from the semantic web community were more dedicated to mapping techniques between relational databases and predefined ontologies than on automated ways to derive ontologies from databases.

Since being able to construct ontologies from legacy databases was a mandatory requirement in our case study, the required efforts have been invested in the development of a comprehensive support for this task. Ontology learning from relational databases corresponds to a highly recurrent generic need and the development ended with the RDBToOnto + IDyllic software combination that has been already applied on a large pool of databases. Additionally, the tools have been successfully used in other projects. This RDBToOnto + IDyllic combination is undoubtedly the most comprehensive and efficient support available in this domain. The application of this software support on the maintenance database is a representative exploitation of this ontology learning help. A complex transitioning project has carefully elaborated to obtain precise maintenance ontologies.

Task: T7.4

Reuse of Existing Ontologies

Ontology acquisition is a multi-step process but the tools developed in TAO do not cover all the steps. More particularly, reuse of pre-existing ontologies is an important step in ontology acquisition which is not supported. Interoperability is often ensured through reuse of ontologies. In our specific context, the analysis of business data models in the aircraft maintenance domain revealed the importance of reuse in this case study.

In the defined acquisition process, ontology reuse is applied after a bootstrapping step where first ontologies are derived from legacy databases. Reuse amounts to replace parts of the automatically generated ontologies with references to equivalent widely shared ontologies. Automated processing for identifying relevant reusable ontologies could be envisaged through the exploitation of ontology mapping techniques. However, this issue was out of the scope of the project. It can be considered as a key extension to complement the existing support. In the case study, ontology reuse has been performed without software assistance.

Content Augmentation of Services

Finding ways to ease the semantic annotation of services was a key requirement. In this work package, the problem was focused on annotating maintenance services with the goal of exploiting as much as possible the ontologies that were semi-automatically derived from the legacy data. The maintenance semantic repository was actually reused and manually extended with additional concepts from the domain of maintenance management.

Automated annotation of services is a difficult task. References to concepts in service descriptions can take many forms in which few regularities can be found. An interactive approach to service annotation has been followed using the Radiant tool which is integrated in the TAO suite. Though useful, this support was partial. This category of tools seems to be the most adequate for this specific task of SAWSDL annotation and we can consider that the underlying interactive approach will gain maturity in the near future with consolidated tools.

Content Augmentation of Documents

On documents, more active help is expected, especially when dealing with technical documentation where the terminology employed is thoroughly controlled. The work on content augmentation of technical publications has been the opportunity to properly address with a well-suited technology a recurring problem in the domain of technical authoring: the annotation of documents with technical data for sake of efficient information access and update. Effort reduction is clearly demonstrated with the prototype based on ontology and NLP techniques.

Task: T7.4

Because of the large number of explicit references to concepts in these highly technical documents, the annotation task performed with the KCIT system achieves very good performance that makes it almost ready for exploitation in real applications. However, robustness needs to be strengthened to face the variability of the expressions which is lower in this context but still high for automated processes. A significant proportion of partial references (through abbreviations and various kinds of anaphoric expressions) can hardly be identified using simple matching techniques. Ongoing work at USFD on KCIT will help to also address the partial references.

2 Model quality

A key requirement stressed in this project is the need for systematic ways to improve the business data models. A better representation framework can contribute to this end. Ontology languages are more suitable for expressing conceptual relations, with built-in constructs to encode hierarchical relations and a powerful representational apparatus for properties that can be used to formalise a wide range of conceptual relations.

In the context of case study 2, several data models have been transitioned to ontologies. An important result is the better account of the underlying conceptual structure obtained with the target models.

Furthermore, thanks to the infrastructure provided by WP4, especially with OWLIM lightweight reasoner, model validation was addressed in a more systematic way. Consistency checking helps to assess the completeness of the elaborated semantic repository with respect to the model specification (e.g. relation cardinalities). In the specific domain of aircraft maintenance, the technical definition of the maintenance activities is an incremental and collaborative process. Completeness of work progress needs to be regularly assessed.

Data redundancy is often a major problem in relational databases. We clearly demonstrated how the transformation to ontologies of the maintenance models allows a drastic reduction of redundancy.

3 Interoperability

In our context, two aspects of interoperability should be distinguished:

- Interoperability between services with ontology-based exchanges
- Interoperability between the tools exploited to produce and manipulate the ontologies

Ensuring interoperability between services through ontologies is one of the primary motivations for the project. This issue has been properly addressed in case study 2

Task: T7.4

with an experimentation and evaluation of the proposed approach to semantic web services. Our reference scenario of maintenance management has been fully modelled with vanilla SOA technology augmented with SAWSDL.

Main finding is that we succeeded to set up for this application sound data mediation mechanisms based on SAWSDL. Data mediation based on ontologies allowed semantic interoperability between services relying on different yet compatible data models. However, mediation also means model duplication which is problematic in terms of model consistency. Additionally, communication between the enhanced services involves even more message transformations. This is discussed under the performance dimension section.

The ontologies produced in the development of this case study have been seamlessly manipulated with the tools. The models are not dependent on the tools, which is a highly valuable feature.

4 Performance

One of the first concerns is the performance of management and access to the semantic repositories. This is actually not supposed to be a burning issue in our application regarding the amount of data to be manipulated in our applications (500K to 5 million triples depending on the complexity of the aircraft to be maintained). Triple stores and scaled for much larger repositories. However, given that the legacy database in our case study is based on the relational model, it is natural to expect from the new technology the same level of performance.

Data loading, even on our "small" repositories, is a function which is not fully satisfactory. Users can be sceptical on the opportunity of adopting a new technology which is slower than the legacy one. However, as far as triple stores and light weight reasoners are concerned, the first impression is often quickly forgotten thanks to the good performance achieved by the browsing and querying services offered by the tools. Additionally, with OWLIM, loading includes the reasoning step that significantly enriches the data.

A limitation we had to manage is that manipulation of (relatively) large ontologies is not possible with general-purpose ontology editors.

The performance of ontology acquisition is good, even though the iterative process can be time consuming on large databases. However, RDBToOnto offers various options that allow iteratively refining the result without having to run the full generation process at each iteration (for example, the generation of the instances can be done at the end).

Task: T7.4

5 Browsing/querying capabilities

Many early adopters have been attracted by semantic technologies because of the availability of high quality editors like Protégé and Swoop which offer outstanding browsing and querying facilities. These tools have been extensively used in this case study.

The availability of the same (standard) query language (SPARQL) in the different access tools is a highly appreciated feature.

Reasoning significantly increases the browsing and querying capabilities (for example, being able to easily explore instances from intermediate classes in the hierarchies thanks to RDFS reasoning). However, such functionalities are not always available from editors with large or medium-size ontologies as reasoners which are integrated in the most popular editors can hardly handle large ontologies. Functionality integration is actually one of the limitations of the existing open source support in this matter.

To meet the information access needs in case study 2, we had to develop an ontology access tool that offers browsing and querying capabilities on large ontologies enriched through reasoning.

6 Coherence and coverage of the toolset

The tools employed to build the new application are all able to properly handle OWL DL ontologies (though "scalable" reasoning is only possible on a subset on DL constructs). Potential inconsistencies have been avoided by not involving several ontology formalisms with different semantics.

Major steps of the transitioning process applied in the case study are well supported. One important exception is ontology normalisation in ontology acquisition which is manually performed. However, the existing toolset can coherently be extended with a new component covering this step. It can be envisaged to benefit from efforts on ontology mapping from other projects.