

TAO – Transitioning Applications to Ontologies

Annual Report – November 2008



www.tao-project.eu

TAO¹, a strategic targeted research project within the EU's 6th Framework Programme, is developing the next generation of knowledge integration technologies: semantic web services and use of ontologies in software engineering. The goal is to provide an open-source infrastructure to support the transitioning of legacy software systems towards Semantic-based Service Oriented Applications.

To make this possible, TAO will tackle several major bottlenecks of knowledge technologies in the areas of semi-automatic creation of domain ontologies, inclusion of semantic content from legacy systems and augmentation and storage of distributed, heterogeneous content. The project will provide these services within a user-friendly integrated development environment for the description and development of knowledge services.

TAO began in March 2006 and runs until February 2009.

Summary of Activities

The demand for lowering the price of migrating existing systems and applications to a Semantic Web Service (SWS) platform is huge. The main innovation of TAO is semantic web-service bootstrapping, and it is an approach that will address this need by partial automation of the re-engineering process, supported by a unified framework.

The outcome is an infrastructure where the re-factoring and semantic augmentation of services can be informed by semantically rich, relevant and shared knowledge sources.

The main implementation outcome is the TAO Suite². It is a user-friendly infrastructure, implementing the TAO semantic web-service (SWS) bootstrapping methodology for the transitioning of existing application into semantic SOAs. It enables semantic interoperability between heterogeneous data resources and distributed applications. Components for ontology learning, content augmentation and heterogeneous knowledge stores are the major building blocks of the TAO suite.

The project has two case studies, which play two very different roles. One is on transitioning a comprehensive open source reference application, forming the basis of strong dissemination. The other is a data-intensive business process application that will form the basis for commercial exploitation.

¹ A project overview demonstration is available at <http://www.tao-project.eu/researchanddevelopment/demosanddownloads/tao-demo/index.html> . An online overview video presentation of TAO is here: http://videlectures.net/tao08_bontcheva_tao/

² An online preview of the TAO Suite is available at: http://www.tao-project.eu/researchanddevelopment/demosanddownloads/taosuite_gui_preprototype_video/taosuite_gui_preprototype_video.html

The commercially-oriented TAO case study focuses on providing semantic-based support to aircraft maintenance processes. The goal is to show how maintenance of legacy data repositories and business processes can be migrated into a semantic-powered SOA framework and to practically assess the added value of the resulting service-based applications in terms of interoperability, knowledge reuse and adaptability.

The second case study is in applying semantic technology to software engineering. It aims to provide a publicly available reference showcase of a re-factored ontology- and service-based open source legacy system.

In the first year of the project, which ended in March 2007, we focused our research on ontology learning and storage of heterogeneous content; development of case study scenarios; definition of the TAO methodology; and scientific dissemination. In the second project year (ended in March 2008), the emphasis was on further research and technological developments in the areas of ontology learning, content augmentation, TAO web service definition and design of the TAO Suite. Since April 2008 we have focused on the development of the TAO Suite and its components and the two case studies prototypes. Quantitative evaluation of the TAO technology is also under way. With respect to dissemination, 2008 saw more activities oriented non-scientific, industrial targets, but alongside this we have continued our scientific publications. TAO also presented its first tutorial at the European Semantic Web Conference (ESWC)³ and another informal one, focused around the TAO Suite is now scheduled for ISWC'08.

From an exploitation point of view, most TAO results will be made available via Open Source licenses. A number of software prototypes have been made available for download from the project web site:

- Ontology learning tools for software artefacts
- Content Augmentation
- Heterogeneous knowledge store
- Transitioning from relational databases to ontologies

In addition, Ontotext have created an open-source project for the ORDI heterogeneous knowledge store at SourceForge (<http://ordi.sourceforge.net/>).

Complementary to this, by the end of 2008 ATOS plans to upload beta versions of the TAO Suite to two different forges: the Spanish national forge MORFEO and VULCANO, the biggest IP funded by the European Commission focused on Open Source.

TAO has so far benefited from results from previous and ongoing European research projects, such as SEKT and NEON. We also pursue active cooperation and exchange of results with relevant ongoing projects. For instance, through collaboration with Knowledge Web WP2.4, especially with Tomas Vitvar, we had an exchange of ideas on Semantically Enabled SOA and WSMO-Lite. At ISWC'2008 TAO is also delivering an informal TAO Suite tutorial to the ServiceFinder project.

Important work areas

The Case Studies

The first, commercial TAO case study focuses on providing semantic-based support to aircraft maintenance processes. The goal is to show how maintenance legacy data repositories and business processes can be migrated into a semantic-powered SOA framework and to practically assess the added value of the resulting service-based applications in terms of interoperability, knowledge reuse and adaptability. In the

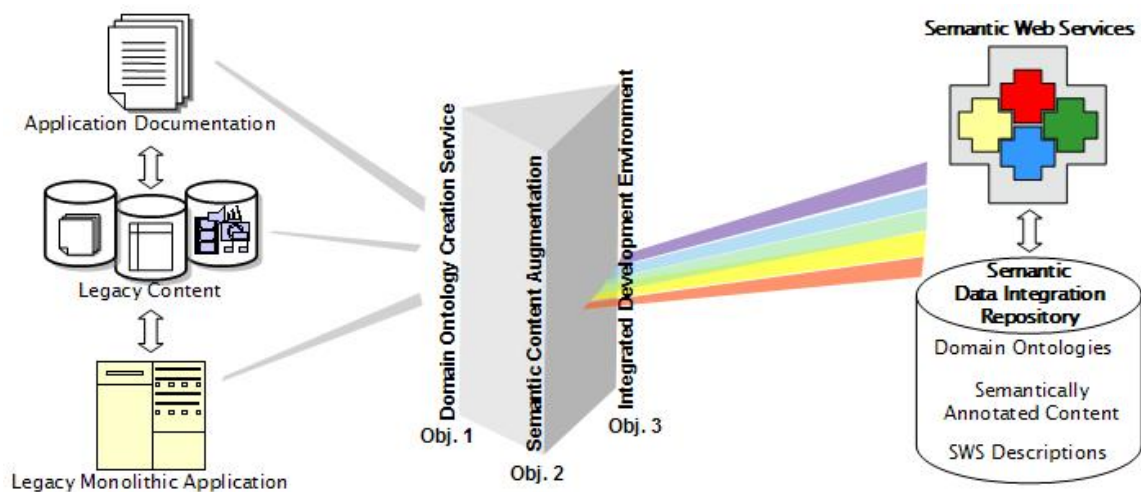
³ The tutorial is now available as online videos at: http://videlectures.net/eswc08_tutorials/

reengineered processes, ontology-based models of technical data and services are exploited to ensure semantic interoperability between maintenance management systems and to provide standardised knowledge inputs for user-mediated semantic annotation of technical publications.

One of the critical issues addressed here is assisted derivation of semantic repositories from existing databases and aeronautics standards. In this respect, experiments of ontology building from the logistic legacy databases and standards have been carried out. The results of these experiments provided valuable inputs to refine the needs related to ontology learning, performance of knowledge stores and querying facilities. Work has now moved on the end-user prototype which will make TAO technology available to the aircraft specialists in Dassault Aviation.

The second case study is in software engineering and provides a publicly available reference showcase of a re-factored ontology- and service-based open-source legacy system. The advantage of transitioning the system to ontologies (i.e., describing its resources with ontology based annotations) will be two-fold. Firstly, the system components and services are easier to discover and integrate within other applications. Secondly, users are able to find easily all information relevant to a given system module, using concept-based search across the different software artefacts. The effort in year one was on requirements analysis and design, following the TAO methodology. Initial experiments on automatic elicitation of a domain ontology from software artefacts and content augmentation of legacy data were carried out in year 2. As part of the public prototype for year 3, we have been working of semantic-based access methods, where research is focused on using natural language to formulate semantic queries and access knowledge.

Research Activities



The core research area in the first year was in defining the TAO methodology for transitioning legacy systems into reusable, semantically described services. In the second year, an investigation in the applicability of WSMO, OWL-S, and SA-WSDL has been conducted as part of the TAO methodology, including an examination of the underlying approaches and aims on which the different frameworks are based. This work has led to a number of papers that have been submitted and/or published. From a more practical side, the use of Methontology (Gomez Perez et al) has been adapted to the challenge of modelling legacy Amazon web services using OWL and SA-WSDL. Work is underway now on a step-by-step “Methodology Cookbook”, which will lower the barrier for early adopters of the TAO technology.

Another research area is automatic ontology learning, where in the first year we carried out a critical review of existing approaches and provided first implementations of ontology acquisition tools for software artefacts. In year 2, we implemented the LATINO web service which provides functionalities for data pre-processing (document network management), link analysis, matrix operations, feature vector computation, and I/O from/to 3rd party software applications such as OntoGen and Pajek. This work continued further in year 3, focusing now on quantitative evaluation in the two case studies.

The third important work area is on automatic content augmentation methods for legacy software artefacts and related content. In year 2 we produced the first versions of the concept identification and information consolidation modules and also experimented with semantic annotation of software screen shots, UML design diagrams, and video software tutorials. In year 3 we are addressing issues related to cross-media content augmentation and user tools for semi-automatic annotation and search.

The fourth active topic of research is on heterogeneous knowledge stores, where in year 1 we produced a design for efficient management of different types of knowledge: unstructured content, structured data, ontologies, and semantic annotations. In year 2, the core engine implementation has been extended to work with heterogeneous knowledge (ontologies, semantic annotations, etc.) and we also carried out an evaluation of the load and response times of the core engine, to assess its efficiency and scalability. In year 3 we worked on knowledge store scalability via distribution where we analysed approaches for distribution of ontology storage and have released the resulting software prototype.

Market Positioning

Over the last 10 years, Business Integration has become a key issue for many enterprises. There have been many drivers behind this interest – among others– the explosive growth of the Internet, competitive pressures, new business models and the need to streamline processes both internally and with partners. The concept of EAI (Enterprise Application Integration) refers to solutions that allow the unification and reuse of existing IT assets, and facilitate their integration into a cohesive corporate system framework. Keeping legacy applications on the mainframe is indeed a good financial decision. By exposing these legacy applications as Web services clients can deliver integration projects faster and cheaper. However, one of the reasons why many of EAI implementations fail is that the semantics of different systems has to be integrated somehow eventually.

Semantic Web Services (SWS) and Service-Oriented Architectures (SOA) make EAI less costly and more reliable. TAO aims to provide a low-cost migration path for legacy applications towards knowledge technologies and will be accessible to both SMEs (which are cost sensitive) and large enterprises (with huge investments in complex and critical IS). By integrating applications and devices at every level of the network, enterprises can transform their internal and external business processes from ponderous and faulty manual processes into real-time, accurate and highly automated processes that can deliver more competitive products and services.

This project will offer a low-cost migration path for legacy applications to knowledge technologies and will be accessible to both SMEs (which are cost sensitive) and large enterprises (with huge investments in complex and critical IS).

User Involvement, Promotion and Awareness

The aero-industry had been chosen as an explicit target market for TAO in order to deliver a specific business case. Dassault Aviation is an industrial group with numerous constituent parts, whose varied activities cover several areas of high technology. With 8500 people, the Dassault Aviation's primary vocation is the design, development, production, sale and maintenance of aircraft. Due to the large amount and diversity of corporate knowledge involved in the design of modern aircrafts, Dassault Aviation is well aware of the fact that significant increase in competitiveness could be achieved by taking advantage of innovative web technologies.

In order to promote take-up from the scientific and industrial communities, the project has released all software as open source, with accompanying demonstrations, all of which have been made available to download from the TAO web site, as follows:

- Ontology learning from software artefacts – demonstration and software
- Content augmentation of software artefacts – demonstrations and software
- Heterogeneous knowledge stores – software with forthcoming demonstration
- Natural language-based knowledge access to semantically enriched legacy content – demonstration
- Transitioning databases to ontologies (RDBToOnto) – software and demos

The TAO consortium is represented within relevant industry-led initiatives such as NESSI and INES. TAO also cooperates with relevant research projects: ServiceFinder, KnowledgeWeb, OntoGrid, SEKT, TripCom, Infrawebs, Super, Neon, LUISA.

Dissemination of research outputs has been carried out through publication of research papers and project posters. In 2008 alone, the project published 11 papers, gave 1 tutorial at a major international conference, published numerous online video lectures, had collaborative meetings with other relevant projects, issued several press releases and had presence at all important scientific conferences and workshops in the field (e.g., ISWC, ESWC, WWW). TAO also targeted dissemination at industry-oriented events such as the European Semantic Technology Conference (ESTC'08) and SemTech'08. A TAO wikipedia article was also created to increase outreach.

Future Work and Exploitation Prospects

The last 6 months of the project will see the completion of the two case studies prototypes, new releases of the TAO open source component technology (ontology learning, content augmentation, heterogeneous knowledge stores, RDBToOnto) and the official release in late October 2008 of the TAO Suite. The commercial case study will showcase a redesigned technical documentation process and demonstrate how knowledge reuse is enhanced through semantic services generated from technical documentation, enriched with ontology-based markup. The open-source software engineering case study is developing and evaluating a testbed which will act as a public reference showcase of an ontology-based software engineering application. The case studies will also feed into the dissemination activities and influence the final exploitation plans.

We will also have a number of industry-oriented dissemination activities. Firstly, we will complete the web site materials oriented towards industrial users, which address three target industry sectors - IT technology providers, developers of knowledge management applications, and materials for the airline industry, including presentations made to aviation standardisation bodies.

The project will also hold an industry-oriented workshop at the end of January 2009, where key early adopters and industrial contacts will be invited for demonstrations, tutorials, and open discussions on the project results.

Further Information

Project web site (includes software downloads, public deliverables and publications):
<http://www.tao-project.eu>

Online video lectures:

http://videlectures.net/eswc08_tutorials/
http://videlectures.net/tao08_bled/

TAO Wikipedia entry:

http://en.wikipedia.org/wiki/Transitioning_Applications_to_Ontologies