T-D5
REWERSE industrial training modules: a preliminary list.

Project number:	IST-2004-506779
Project title:	Reasoning on the Web with Rules and Semantics
Project acronym:	REWERSE
Document type:	D (deliverable)
Nature of document:	R (report)
Dissemination level:	PU (public)
Document number:	IST50679/Telefonica/T-D5/D/PU
Responsible editor:	Sara Carro
Contributing participants:	webXcerpt, Viena, Göttingen
Contributing workpackages:	TTA
Contractual date of delivery:	31 August 2005
Actual date of delivery:	30 December 2005

Abstract

The objective of this deliverable is to develop a description of a set of modules on Semantic Web issues, prepared in T-D3 and adapted to industry audience needs. These modules aim at the creation of a infrastructure for industrial education on Semantic Web issues so as to spread REWERSE's research results. To achieve this goal, these modules are envisaged to be presented at several institutions, among others the Semantic Web School in Austria. Apart from the main features of this set of modules to be developed by REWERSE participants, which will be listed in detail in T-D7, this deliverable will include a work-plan for their development.

Keyword List
Semantic Web education, Semantic Web modules, industrial training

REWERSE industrial training modules: a preliminary list.

Sara Carro[1] and Almudena Vicente[1] and Robert Baumgartner[3] and Wolfgang May[4]

[1]Telefónica Investigación y Desarrollo S.A.U., Spain
Email: scm@tid.es

[3]Technische Universität Wien
Email: baumgart@dbai.tuwien.ac.at

[4]Universität Göttingen
Email: may@informatik.uni-goettingen.de

30 December 2005

Abstract

The objective of this deliverable is to develop a description of a set of modules on Semantic Web issues, prepared in T-D3 and adapted to industry audience needs. These modules aim at the creation of a infrastructure for industrial education on Semantic Web issues so as to spread REWERSE's research results. To achieve this goal, these modules will be presented at several institutions, among others the Semantic Web School in Austria. Apart from the main features of this set of modules to be developed by REWERSE participants, which will be listed in detail in T-D7, this deliverable will include a work-plan for their development.

Keyword List
Semantic Web education, Semantic Web modules, industrial training

Contents

1 Introduction

This deliverable aims at giving a preliminary list about modules regarding Semantic Web issues which are interesting for industry needs and a work-plan for their development.

Technology Transfer Awareness Group (TTA), together with Education and Training Group (ET), have the long range objective of providing durable infrastructure for academic and industrial education on Semantic Web issues, especially those related to the use of rules in semantic modelling. Thus, ET deliverables will contribute mostly to the technology transfer tasks of TTA. For this reason, the choice of modules has been made on the basis of T-D3 and take into account developments in ET work-packages E-D1 and E-D5.

Apart from those modules selected in T-D3 because they are suited for industrial education, new modules have been identified from ET developments which are also addressed to industrial needs.

The structure of this document is as follows:

- Section 2 describes the work-plan for the development of the industrial training modules.
- Section 3 gives brief guidelines about which issues have to be taken into account when adapting the selected material.
- Section 4 enumerates the modules which take part in the preliminary list of the industrial curriculum and gives a brief description about their contents.
- Section 5 depicts the future work to be fulfilled.

2 Work-plan for the development of industrial training modules

The first step towards the creation of a durable infrastructure for industrial education on Semantic Web issues was the identification of those modules which are suitable to industry needs. After that, brief guidelines were given in order to adapt the academic material to an industry audience.

By month 24, a description of the development of the modules as defined in T-D5 including information on the approach taken for the development and the support given by TTA will be included as input for T-D7.

By month 36, deliverable T-D9 will be completed including a report on validation and use of the industrial training infrastructure: software and existing course materials. T-D9 will also review the work-plan for integration of the new industrial modules.

By month 42, deliverable T-D11 will include the updated list of modules for industrial training on the Semantic Web and REWERSE specific topics in particular. This will most likely be prepared in co-operation with Knowledge Web and other partners. The development of the modules including information on which modules will be included in the industrial training infrastructure will be described. This deliverable is an extension and revision of T-D7 and builds on E-D11.

Finally, deliverable T-D13 will include the final version of T-D11. It will also include several reports: one on its use and another one on the co-operation between REWERSE nodes involved in the creation of industrial training modules with the perspective covering the period after the end of REWERSE funding by the Commission. These activities will be fulfilled by month 48.

The development of these modules will be done in parallel with the development of a durable educational infrastructure for industrial modules defined in E/T-D4.

It is important to take into account that both modules and work-plan can be modified if needed during the evolution of the project according to new results from research.

3 Guidelines to adapt material addressing industrial needs

Before focusing on how to adapt the material of the modules selected in former deliverables, there is the need to identify the industrial needs which can be addressed with Semantic Web technologies.

The main problems which can be identified are: Knowledge Management, Enterprise Integration and Electronic Commerce.

At present, in the Knowledge Management area, companies make use of intranet solutions so as to store all relevant information and have this information at their employees disposal. As a result, there are large information sources not well organised and provide poor services. That is to say, current intranets which do not make use of Semantic Web technologies can be useless for users since they do not provide a hierarchical organization of information nor semantic search tools. Intranets are simply containers of information, where this information does not have any sense. It is a human task to look up and select useful information which can be a complete waste of time. As a result, employees complain about the "chaos" and lack of structure in a free-form intranet. For this reason, it can become a hard issue to find the right information at the right moment.

The use of Semantic Web and Ontologies can give the following advantages in the Knowledge Management area:

- Turning information into useful knowledge.
- Harness corporate knowledge assets (maintaining and finding the right information for companies at the right time and right place).
- Secure company competition (efficiently release and reuse of knowledge).
- Improve the understanding and smoothing the changing contextual knowledge.
- Foster collaboration by capturing, representing and interpreting the knowledge resources of their organisations.

Regarding e-Commerce, several features must be taken into account:

- Competitiveness of companies depends on the products and services they offer.
- The ability to find, query and exchange knowledge is fundamental for B2B and B2C e-Commerce.
- Supplier and providers want to reach different customers, while customers want a large range of products at lower price

So as to fulfil these challenges, Semantic Web and Ontologies:

- Enable content-based access, interoperability, integration information and communication.
- Facilitate finding, interrogating and exchanging knowledge for B2B an B2C e-Commerce.
- Smooth the B2B or B2C marketplace friction (content standard initiatives).
- Transform and innovate traditional business intelligence.

The main areas where companies are seeking solutions are that of data integration and semantic search, data management and personalization.

In order to successfully incorporate Semantic Web technologies in industry areas, these technologies should be applied where they really offer benefits and conversely not applying them where they do not: early adoption of Semantic Web technologies for the wrong tasks will lead to industry-wide negative opinions of these new approaches and discourage interest in correct solutions where industry can take advantage of Semantic Web technologies.

For this reason, it is very important to disseminate Semantic Web knowledge to industry areas and provide a real vision on how Semantic Web technologies can contribute to enhance industry issues.

So as to fulfil this goal, several guidelines are provided to prepare material for industrial modules:

- Focus on solving industrial needs.
  It means that courses must be adapted in such a way so that companies can notice what are the benefits of adopting Semantic Web technologies. In this sense, courses should be practically oriented avoiding too many theoretical explanations.

- Be aware of the kind of the target audience.
  Each company is hierarchical and well structured. This means that industrial courses should be tailored according to the needs of each target audience: managers, programmers … So, each audience can notice the advantages which provide the use of Semantic Web technologies for their particular tasks.

- Be aware of time restriction of industries
  In industry, time is money. So, at the time of developing or adapting industrial courses time must be taken into account. This means that an overview about each module must be provided to companies. The overview helps them to decide about the topics they are more interested in and then to focus on the selected topics.

- Do not include details about the research.
  This means that industry audience is interested in the applicability of Semantic Web technologies and not in the theoretical differences among technologies.

- Forget about technical details.
  As it was formerly explained, companies are interested in practical solutions of Semantic Web technologies, especially in the benefits they can obtain if they use them in their industry areas. So, technical details should be offered only on demand.

- Be brief and concise
  As it was formerly introduced, time is vital to industry. So concise and brief overviews focusing on the advantages and solutions of Semantic Web issues will lead to a positive opinion of the company about the technologies exposed.

- Provide a business context and perspective
  In order to fulfil this point, uses cases, statistics, experience in former projects must be included in the courses. So, industry audience, especially managers, can envisage if investments in Semantic Web technologies will provide good results to particular needs.

- Provide state-of-the-art of each technology.
  Industrial modules must be up-to-date on latest Semantic Web technologies and provide a concise overview about features of these technologies in order to help companies to choose the technology which best fits their needs. In this sense, it will be also interesting to introduce comparative chapters about different Semantic Web technologies.

- Give further information and assistance on demand.
  From the previous guidelines, we can extract the idea that industrial courses must give a brief and concise overview about Semantic Web technologies. The main objective is to expose features, advantages and solutions of Semantic Web issues. Therefore, people interested in studying in depth special topics will be assisted on demand.

- Focus on applications and results.
  Including experience in previous projects and solutions on Semantic Web technologies will help companies to really be aware about the benefits of applying Semantic Web issues in real use applications so as they can draw their conclusions about the advantages and disadvantages of adopting these technologies.

- Give examples (preferably out of the context of the customer)
  It will be also interesting to give some examples (preferably out of the context of the customer) to show the applicability of Semantic Web technologies. Even, including small pieces of applications will be useful so that companies will have at their disposal an application example which will be tailored to their own needs and taken as basis for complex developments.

- Give a tiered structure of the course material, so each industry could pick out the most suitable for it.
  This aspect will avoid, for instance, to give a course about foundation topics to an IT company which will have knowledge in these topics, but another industry company will maybe need an introductory course on these issues. In this sense, it is needed to build a multi-layer structure of learning material so that each company can start technical courses from the point they consider appropriate.

4 Industrial training modules description

This section contains a preliminary list of the industrial training modules to be developed in subsequent deliverables so as to include them in the industrial technology transfer infrastructure.

4.1 Information Extraction for the Semantic Web

- Author: Robert Baumgartner
- REWERSE Working Group: I4

- Abstract:

Extraction technologies for the web as of today help unfold the structure of the desired pieces of information from HTML documents and translate it into XML (and subsequently, into a semantic representation if desired) in a very cost-effective way. This bridges the gap between unstructured Web data and structured databases. High-level languages have been developed for Web extraction, and machine learning approaches are used for automatic generation of so-called wrappers; they rely on learning from examples and counter-examples. Interactive approaches allow for semi-automatic extraction generation and offer convenient visual dialogues to generate a wrapper based on a few examples and user interaction. Deep Web navigation and expressiveness are important aspects of extraction technologies. A further step is integration of extracted Web data into enterprise processes, a task e.g. crucial in the communication of automotive and automotive supplier companies. Integration brokers provide infrastructure for sharing information between applications (EAI, B2B), many of them featuring Web connectors. Business Activity Monitoring vendors strive to deliver instant awareness of business critical events to customers. Heterogeneous environments such as integration and mediation systems require a conceptual information flow model. For data integration schema matching concepts are researched, and methods for treating incomplete and inconsistent data. Web Data Extraction and Integration technologies open the way to use today's web as the largest database to provide input for business intelligence applications (including cooperative and non-cooperative sites) and to reuse data for portals and personalisation and for portal-based cross-company communication. The application areas are manifold - every vertical business domain requires to analyse, monitor and interact with Web data. The course puts special focus to B2B and BI application scenarios that strongly rely on Web data. Moreover, the Semantic Web can be populated with data from today's web: In REWERSE project, e.g. the award-winning Lixto technology (www.lixto.com, basic research carried out by the Vienna University of Technology, www.tuwien.ac.at) was successfully applied to extract scientific publication data for the "Personal Reader" (www.personal-reader.de). Semantic Web technologies can furthermore help to reason about extracted data and hence pave the way to a personalized view on integrated and analyzed Web data.

- Industrial justification:

Information extraction systems analyse unrestricted text in order to extract information about pre-specified types of events, entities or relationships. It pulls facts and structured information from the content of large text collections. These techniques resemble what we do as we scan an article to assess its relevance to our goals. As a result, information extraction can help enterprises to turn information into useful knowledge , improve searches and save time and effort.

4.2 Evolution and Reactivity for the Web

- Author: Wolfgang May
- REWERSE Working Group: I5

- Abstract:

The Web and the Semantic Web is presented as a "living organism" combining autonomously evolving data sources, each of them possibly reacting to events it perceives. Rather than a Web of data sources, we envisage a Web of Information Systems, where each such system, besides being capable of gathering information (querying persistent data, as well as "listening" to volatile data such as occurring events), is capable of updating persistent data, communicating the changes, requesting changes of persistent data in other systems, and being able to react to requests from other systems. The dynamic character of such a Web requires declarative languages and mechanisms for specifying the evolution of the data.

This course includes a first talk about foundations of evolution and reactive languages in general, and then concentrates on some specific issues posed by evolution and reactivity in the Web and in the Semantic Web.

- Industrial justification:

Reactivity on the Web, the ability to detect simple and composite events that occur on the Web and respond in a timely manner, has recently emerged as an issue of concern in Web and Semantic Web circles. Although a common perception of the Web is that of a distributed information system giving rise to access data in a read only manner, many Web-based systems need to have the capability to update data found at (local or remote) Web resources, to exchange information about events (such as executed updates), and to detect and react not only to simple events but also to complex, real-life situations. The issue of updating data plays an important role, for example, in e-commerce systems receiving and processing buying or reservation orders, and e-learning systems selecting and delivering teaching materials depending on the students' performances on tests. Businesses that are able to react to events quickly and take appropriate decisions are likely to have a competitive advantage.

4.3 Fundamentals of Semantic Web Ontology Languages

- Author: Enrico Franconi
- REWERSE Working Group: Non - REWERSE Contributors

- Abstract:

This course is a tutorial on Semantic Web ontology languages, designed to provide such formal descriptions. The focus is on the languages RDF Schema and (different variants of) OWL, adopted as standards by the W3C. An introduction to the languages illustrated by examples, will clarify the importance of ontology engineering. The course will also explain the logical foundations of OWL by providing an introduction to the underlying Description Logic and its use in ontology reasoning on the web.

- Industrial justification:

This module is included in order to give to the industry audience a whole vision about what Semantic Web is as well as the languages which support this issue specially focusing on ontology definition. The use of Semantic Web technologies in industry and business workflow can help to improve the efficiency and productivity of their tasks. In other words, using a combination of web pointers, web mark-up and ontology languages, service descriptions can be enriched by including a machine-readable description of how the service runs and some explicit logic describing the

consequences of using the service. As a result, Web Services can extend programs to more efficiently perform tasks for users with less human intervention. Since Ontologies describe the semantics of a domain in a human-understandable and computer-processable way, they play a crucial role to enable content-based access, interoperability and communication across the Web. For these reasons, it is believed that a module which explains the basis of ontology engineering meets the industrial needs.

4.4 Personalization for the Semantic Web

- Author: Matteo Baldón, Cristina Baroglio and Nicola Henze
- REWERSE Working Group: A3

- Abstract:

An adaptive web site observes the visitor access patterns of its pages, and automatically improves its structure and presentation. Personalization is a process by which it is possible to give the user optimal support in accessing, retrieving, and storing information, where solutions are built so as to fit the preferences, the characteristics and the taste of the individual. This result can be achieved only by exploiting machine-interpretable semantic information, e.g. about the possible resources, about the user him/herself, about the context, about the goal of the interaction. Personalization is realized by an inference process applied to the semantic information, which can be carried out in many different ways depending on the specific task.

The core problem in finding the information the user wants, is *describing* what the user wants. Results from search engines are often disappointing because most search requests are too short and unspecific to yield good results. Once a Web site with interesting information is found, it is often difficult to navigate to interesting pages only, because the site can only be navigated using its predefined link structure, independently of the search request that brought the user to that site.
The community of *user modeling* and *adaptive hypermedia* offers solutions for this problem: using information gathered about the user during the browsing process to change the information *content* and *link structure* on-the-fly. User modeling captures the mental state of the user, and thus allows that knowledge to be combined with the explicit queries (or links) in order to determine precisely what the user is looking for.

The huge amount of information on the web raises the need of more sophisticate tools to access it and of more sophisticate tools to organize it. Semantic Web suggests a way to achieve these goals. In other words, Semantic Web allows the users to access information on the net in a personalized way, so as to satisfy their needs, goals, and preferences.

The objective of this course is to provide a coherent introduction into issues and methods for realizing personalization in the Semantic Web.

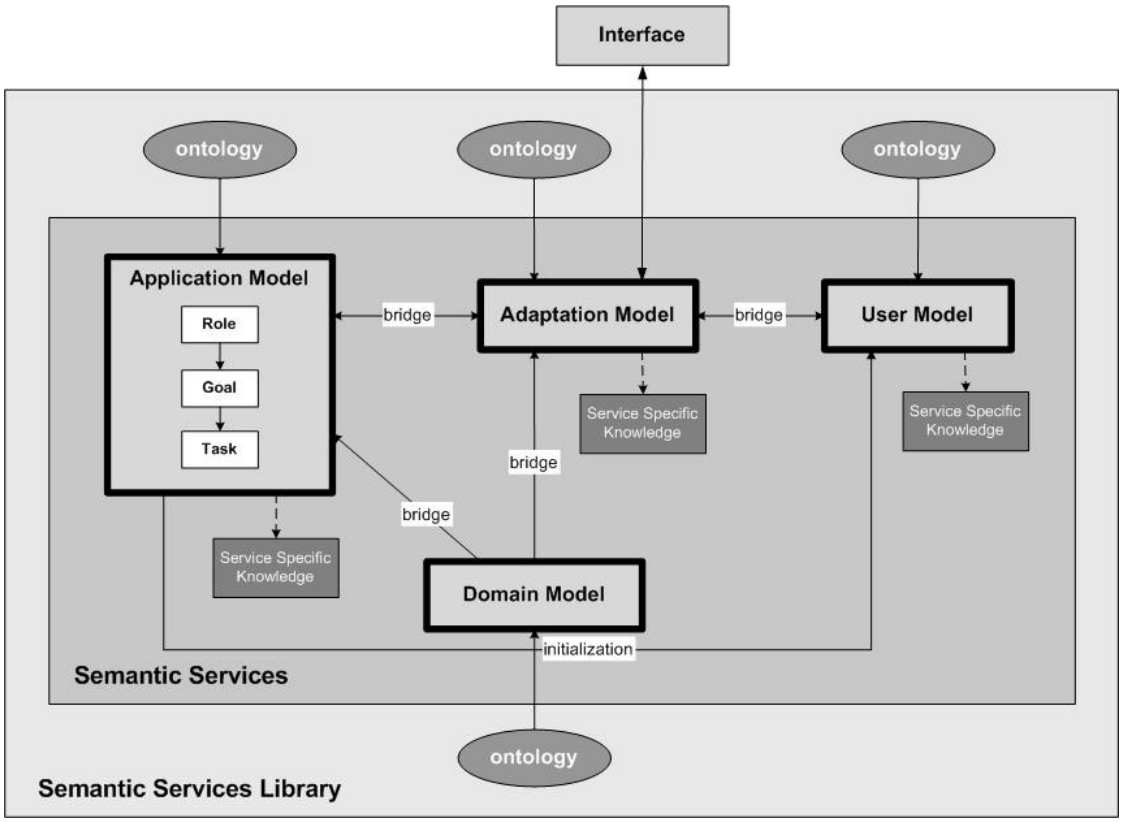

Figure 1. Architecture for adaptive Web-based systems (adapted from Motta et al. 2003)

- Industrial justification:

Success in business depends on forging strong relationships. Personalized information systems aim at giving the individual user optimal support in accessing, retrieving, and storing information. From industry point of view, it's a means of meeting the customer's needs more effectively and efficiently, making interactions faster and easier and, consequently, increasing customer satisfaction and customer service.

5 REWERSE activities and industrial training modules

REWERSE working groups are supposed to work out a set of concrete deliverables, in terms of languages and tools for the Semantic Web and advanced treatment of information.

The research activities of REWERSE are organised in 8 Working Groups:

- Working Groups on Web issues
  - WG I1 Rule Markup Languages
  - WG I2 Policy Specification, Composition, and Conformance
  - WG I3 Composition and Typing
  - WG I4 Reasoning-Aware Querying
  - WG I5 Evolution and Reactivity

- Working Groups on advanced Web applications
  - WG A1 Web-based Decision Support for Event, Temporal, and Geographical Data
  - WG A2 Towards a Bioinformatics Semantic Web
  - WG A3 Personalised Information Systems

In addition, REWERSE has the following workpackages aiming at the dissemination of knowledge:

- ET “University Education and Training”
- TTA “Technology Transfer and Awareness”
- PRA “Presentation, Reviewing and Assessment”

This deliverable T-D5 is part of the set of deliverables belonging to Technology Transfer and Awareness workpackage. Its objective is to provide a first approach about the industrial modules which can be included in the infrastructure for industrial education on the Semantic Web issues.

Once we have reviewed the working groups of the REWERSE activities and we have given an overview about the courses suitable for industrial needs, we are going to expose the relationship between this set of modules and REWERSE activities:

- **Information Extraction for the Semantic Web**

This course is related to WG I4 “Reasoning-Aware Querying” which aims at integrating reasoning capabilities within Web query and transformation languages.

Retrieving data on the (conventional as well as Semantic) Web calls for a database-like querying and for “transforming” (i.e. re-structuring) the collected data. For this, “Web query and transformation languages” inspired from database query languages give rise to an easier programming and software maintenance. Data transformation is a first essential aspect of Semantic Web applications such as context-aware Web systems. As reasoning is a second essential aspect of Semantic Web applications, a Web query and transformation language offering reasoning capabilities would considerably ease the implementation Semantic Web applications. The objective of the WG is to develop, implement, and test (on selected Semantic Web applications) such a (provisionally rule-based) language at a pre-standard level.

It was considered to include this course in the infrastructure training module since these two aspects, data transformation and reasoning, are also pretty important for industry and business issues.

This module aims at presenting the main features of Information Extraction and the importance of the languages and tools needed to extract and integrate information from various different Web sources, or in general, various heterogeneous sources. Information extraction issue can take advantage of tasks and researches which are being developed by this working group, such as query languages and reasoning capabilities researches.

As WG4 is working in the development of a new query and transformation language offering both access to Web data and reasoning capabilities, this new language can be included in following versions of the module. The researches and surveys carried out by the members of this working group will help to the development of the content of the module.

- **Evolution and Reactivity for the Web**

This course is related to WG I5 “Evolution and Reactivity” which addresses specifying the updates of data in Web information systems. Besides the realization (i.e. technologies and languages) and the use (i.e. querying) of the Semantic Web, its maintenance and evolution are important issues. The dynamic character of the Semantic Web requires (declarative) languages and mechanism for specifying its maintenance and evolution. Semantic Web data sources may be reactive, not only due to the incorporation of updates, but also in that they perceive events and incoming messages, communicate with other components, and execute actions. In general, reactive behaviour is useful to mediate between data sources and to implement workflows integrating various data sources and spreading information through a network of data sources.

In this working group of Rewerse, participants investigate on all these aspect of evolution and reactivity in the Web. So, this module summarizes the activities carried out by WG I5 and shows the benefits of which industry can take advantage.

- **Fundamentals of Semantic Web Ontology Languages**

This course is not related to any of the working groups or workpackages of the REWERSE’s activities, but it is included in order to give a theoretical overview about Semantic Web issues and technologies.

- **Personalization for the Semantic Web**

The working group A3 on "personalized information systems" aims at advancing the state of the art of customized information delivery applications in the Semantic Web. Applications like adaptive information systems and personalized e-commerce systems use backbones build upon reasoning. Results from working groups I1-I5 will bring decisive input for bringing personalization to the Semantic Web and thus advance the design of the Adaptive Web.

To provide users *optimised* access to information, with appropriate *quality*, with required *information depth*, according to *personal preferences* of this particular user,

according to the user's *actual situation* and *context* will be one of the key technologies for usability in the Semantic Web: the vision of a "Semantic Web" includes the vision of an "adaptive Web" which knows like a personal agent the specific requirements of a user, takes goals, preferences or the actual context into account in order to optimise the access to electronic information.

This module tries to highlight important issues of the achievements carried out by this working group and also shows the industrial benefits of personalized and adaptive systems and how Semantic Web technologies can help to fulfil these goals.

6 Conclusions and future work

The aim of this report has been fulfilled: a preliminary list of the industrial modules to be developed is included as well as brief guidelines about how to adapt the academic material to industrial needs. The modules exposed above have been chosen because they can contribute to solve and improve industrial needs.

These modules will be developed in detail in the next deliverables:

- T-D7. Report on the development of industrial training modules.
- T-D11. Revised list of industrial training modules.
- T-D13. The final list of industrial training modules and its use.