D 1.1.4 v2
System and knowledge technology components for prototypical applications and business cases

Lyndon Nixon (FU Berlin) & Alain Leger (FT)
with contributions from
WP1.1 partners: Malgorzata Mochol (FU Berlin), Stamatia Dasiopoulou (CERTH)
Research WP partners: Holger Wache (VUA, WP2.1), Jerome Euzenat (INRIA, WP2.2), Siegfried Handschuh and Tudor Groza (NUIG, WP2.3), Tomas Vitvar (NUIG, WP2.4), Rob Shearer (Uni Manchester, WP2.5)

Abstract.
EU-IST Network of Excellence (NoE) IST-2004-507482 KWEB
Deliverable D1.1.4 v2 (WP1.1)

This document extends the previous deliverable D1.1.4v1 with details regarding the Industry-Research co-operation activities. We describe the ongoing co-operation with our industry partners who have needs that can be met by research carried out in KnowledgeWeb. We identify results achieved to date and outline plans for technology transfer within the project. Furthermore we identify a further emerging scenario as well as take first steps towards ensuring the continued support for Semantic Web uptake in industry also beyond the timeframe of KnowledgeWeb.

Document Identifier:	KWEB/2005/D1.1.4v2/1.1
Class Deliverable:	KWEB EU-IST-2004-507482
Version:	V1.1
Date:	August 8, 2006
State:	Final
Distribution:	**Restricted**

Knowledge Web Consortium

This document is part of a research project funded by the IST Program of the Commission of the European Communities as project number IST-2004-507482.

University of Innsbruck (UIBK) – Coordinator
Institute of Computer Science,
Technikerstrasse 13
A-6020 Innsbruck
Austria
Contact person: Dieter Fensel
E-mail address: dieter.fensel@uibk.ac.at

École Polythechnique Fédérale de Lausanne (EPFL)
Computer Science Department
Swiss Federal Institute of Technology
IN (Ecublens), CH-1015 Lausanne.
Switzerland
Contact person: Boi Faltings
E-mail address: boi.faltings@epfl.ch

France Telecom (FT)
4 Rue du Clos Courtel
35512 Cesson Sévigné
France. PO Box 91226
Contact person : Alain Leger
E-mail address: alain.leger@rd.francetelecom.com

Freie Universität Berlin (FU Berlin)
Takustrasse, 9
14195 Berlin
Germany
Contact person: Robert Tolksdorf
E-mail address: tolk@inf.fu-berlin.de

Free University of Bozen-Bolzano (FUB)
Piazza Domenicani 3
39100 Bolzano
Italy
Contact person: Enrico Franconi
E-mail address: franconi@inf.unibz.it

Institut National de Recherche en Informatique et en Automatique (INRIA)
ZIRST - 655 avenue de l'Europe - Montbonnot
Saint Martin
38334 Saint-Ismier
France
Contact person: Jérôme Euzenat
E-mail address: Jerome.Euzenat@inrialpes.fr

Centre for Research and Technology Hellas / Informatics and Telematics Institute (ITI-CERTH)
1st km Thermi – Panorama road
57001 Thermi-Thessaloniki
Greece. Po Box 361
Contact person: Michael G. Strintzis
E-mail address: strintzi@iti.gr

Learning Lab Lower Saxony (L3S)
Expo Plaza 1
30539 Hannover
Germany
Contact person: Wolfgang Nejdl
E-mail address: nejdl@learninglab.de

National University of Ireland Galway (NUIG)
National University of Ireland
Science and Technology Building
University Road
Galway
Ireland
Contact person: Christoph Bussler
E-mail address: chris.bussler@deri.ie

The Open University (OU)
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA
United Kingdom.
Contact person: Enrico Motta
E-mail address: e.motta@open.ac.uk

Universidad Politécnica de Madrid (UPM)
Campus de Montegancedo sn
28660 Boadilla del Monte
Spain
Contact person: Asunción Gómez Pérez
E-mail address: asun@fi.upm.es

University of Karlsruhe (UKARL)
Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB
Universität Karlsruhe
D-76128 Karlsruhe
Germany

Contact person: Rudi Studer
E-mail address: studer@aifb.uni-karlsruhe.de

University of Liverpool (UniLiv)
Chadwick Building, Peach Street
L697ZF Liverpool
United Kingdom
Contact person: Michael Wooldridge
E-mail address: M.J.Wooldridge@csc.liv.ac.uk

University of Manchester (UoM)
Room 2.32. Kilburn Building, Department of Computer Science, University of Manchester,
Oxford Road
Manchester, M13 9PL
United Kingdom
Contact person: Carole Goble
E-mail address: carole@cs.man.ac.uk

University of Sheffield (USFD)
Regent Court, 211 Portobello street
S14DP Sheffield
United Kingdom
Contact person: Hamish Cunningham
E-mail address: hamish@dcs.shef.ac.uk

University of Trento (UniTn)
Via Sommarive 14
38050 Trento
Italy
Contact person: Fausto Giunchiglia
E-mail address: fausto@dit.unitn.it

Vrije Universiteit Amsterdam (VUA)
De Boelelaan 1081a
1081HV. Amsterdam
The Netherlands
Contact person: Frank van Harmelen
E-mail address: Frank.van.Harmelen@cs.vu.nl

Vrije Universiteit Brussel (VUB)
Pleinlaan 2, Building G10
1050 Brussels
Belgium
Contact person: Robert Meersman
E-mail address: robert.meersman@vub.ac.be

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed writing parts of this document:

Centre for Research and Technology Hellas
France Telecom
Freie Universität Berlin
University of Trento
Vrije Universiteit Brussel

Changes

Version	Date	Author	Changes
0.1	08-05-2006	Lyndon Nixon	Basic structure
0.2	26-05-2006	Lyndon Nixon	First drafts of co-operation in WPs 2.1, 2.2 and 2.5
0.3	13-06-2006	Lyndon Nixon, Holger Wache, Jerome Euzenat, Tomas Vitvar	Updates of WP2.1 and WP2.2, addition of WP2.4
0.4	15-06-2006	Lyndon Nixon, Tudor Groza	Added WP2.3
0.5	05-07-2006	Lyndon Nixon, Stamatia Dasiopoulou, Tomas Vitvar	Added emerging scenario (Multimedia), added updated version of WP2.4
0.75	07-07-2006	Lyndon Nixon, Siegfried Handschuh	Added Introduction, updated version of WP2.3, expanded Future Steps and Conclusions
0.8	10-07-2006	Lyndon Nixon	Final version prior to Quality Control
0.9	26-07-2006	Lyndon Nixon	First revision subsequent to Quality Control
1.0	31-07-2006	Lyndon Nixon	Final version following Quality Control
1.1	08-08-2006	Lyndon Nixon	Submission version following Quality Assurance

Executive Summary of Deliverable

This document extends the previous deliverable D1.1.4v1 with details regarding the Industry-Research co-operation activities. For each Research workpackage we describe the ongoing co-operation with our industry partners who have needs that can be met by research carried out in KnowledgeWeb. We identify results achieved to date and outline plans for technology transfer within the project. Furthermore we identify a further emerging scenario where we foresee a future need for Semantic Web technologies as well as take first steps towards ensuring the continued support for Semantic Web uptake in industry also beyond the timeframe of KnowledgeWeb.

Contents

1. Introduction

The Knowledge Web Industry Area, and in particular the Work Package 1.1, sees its role over the duration of the EU Network of Excellence in preparing for and achieving the transfer of semantic technologies from academia to industry. We have formed an Industry Board made up of industrial partners with an interest in the potential of semantic technologies for their business processes. Furthermore Knowledge Web contains a Research Area which brings together leading Semantic Web research institutes from across Europe.

The fundamental problem is that Semantic Web researchers face the danger of working in isolation and performing good research which, however, finds no practical application in the real world. On the other hand, European industry finds that its current business processes could still be significantly improved, and issues such as the heterogeneity of business data and processes, scalability of business IT systems and rapidly changing and evolving markets mean that enterprises are constantly in need of new solutions to remain competitive. Semantic technologies have the potential to solve enterprise data problems such as heterogeneity of data sources, difficulties in scaling knowledge systems, and issues arising from dynamically changing information (and hence these are key research topics of the Research Area work packages, together with Semantic Web Services as solution to business process problems). Yet to date these technologies have had little opportunity to demonstrate industrial viability. Enterprise spending on new technology will occur only when that technology can be concretely demonstrated as solving known business problems and is recognized as being industry mature.

In order to demonstrate the industrial value of Semantic Web technologies, Work Package 1.1 has performed a number of tasks:

- Formed an Industry Board made up of companies interested in the value of semantic technologies for their business activities,
- Established communication channels with Industry Board members as well as industry in a wider perspective through a Knowledge Web Industry Portal, mailing list, newsletter as well as Industry Day events at major conferences and industry-focused talks.
- Collected Use Cases from Industry Board members as indicative of typical industrial problems and current technology locks to using semantic solutions, as well as performed further analysis to derive industrial requirements for the Research Area.

We are now in the last phase of our efforts to achieve technology transfer from academia to industry. Building on the foundations of these earlier (and ongoing) tasks, we have recognized that it is necessary to bring *industry needs* and *research activities* together. We established an Industry-Research co-operation track and it is the activities and aims of this track that we introduce and describe in this deliverable.

1.1 The Industry-Research Co-operation

The aim of the Industry-Research co-operation is to focus aspects of Semantic Web research on real world business problems which may be solved through semantic technologies. The ultimate goal is transferring results of that research into enterprise scenarios to (prototypically) solve those business problems.

We take a number of concrete steps (being documented in D1.1.4 versions 1 to 3).

- In D1.1.4v1 we selected some of the use cases provided to us by the Industry Board members and provided *executive summaries* for each identifying industrial requirements for Semantic Web research.
- Each Research WP was invited to identify a use case and to prepare a research time plan for meeting the use case requirements through their research and (when possible) make a first transfer of technology to the enterprise for prototypical evaluation. The use case selection, progress to date and future plans leading to technology transfer are presented in this deliverable.
- D1.1.4v3 will report on what has been achieved, evaluate this achievement and outline how technology transfer may still be supported beyond the duration of KnowledgeWeb.

Communication has been enhanced in the previous year through dedicated Industry-Research slots in plenary meetings, direct contact between WP1.1 and the Research WP representatives and the production of an Industry newsletter to communicate results to our Industry Board members.

It is difficult to provide a measure for the success of this co-operation. However, we will document the Industry-Research co-operation measuring success as being the extent to which Semantic Web research has achieved results satisfactory for use in an industrial setting and the extent to which Semantic Web technologies have been actually (prototypically) applied by enterprises in solving business problems. This deliverable already offers a concrete vision of what may be achievable in the next 12 months. In D1.1.4v3 we will attempt to provide a measure of success by identifying for each Research WP the methodologies and tools which have been produced in response to industrial requirements and for each (use case providing) Industry Board member the advance that has been made in the understanding and use of Semantic Web technologies within their enterprise.

2. Progress Report on Business Use Cases

For each Research WP, we provide the following analysis:

- A brief description of the business case that was chosen (if it was an use case originally considered in D1.1.2 or D1.1.4v1 we also reference these texts)
- A justification of the applicability of the research being done in the workpackage to the chosen use case
- An outline of the technology transfer that is foreseen as going to take place as a result of the co-operation of this research workpackage with the industry partner in their use case
- A description of what progress has been achieved to date towards achieving that technology transfer
- A set of milestones for the remainder of the project which set concrete aims to be reached in order to ensure that the expected technology transfer will take place
- Concluding remarks on this particular Industry-Research co-operation activity

The following research workpackage co-operations are considered by this deliverable:

2.1 Research WP 2.1 : Scalability

Chosen Use Case:	Recruitment
KW Partners:	VUA, FU Berlin
IB Member:	WorldWideJobs GmbH

2.1.1 Chosen Business Use Case

The Scalability work package has chosen the Recruitment use case which was described in KnowledgeWeb deliverable D1.1.2 (section 2.1, p11).

The use case deals with the challenge of efficiently filling open job vacancies with qualified suitable candidates. A proposed solution is improved matching between job offers and job seekers making use of semantics to enable a richer matching between concepts which would not be possible using standard text analysis matching.

As candidate ontologies we selected some of the most relevant classifications in the area,

deployed by federal agencies or statistic organizations: German Profession Reference Number Classification (BKZ), Standard Occupational Classification (SOC), German Classification of Industrial Sector (WZ2003), North American Industry Classification System (NAISC), German version of the Human Resources XML (HR-BA-XML) and Skill Ontology developed by the KOWIEN.

2.1.2 Applicability of Research to the Use Case

In WP2.1, a number of approaches are being developed and evaluated which aim to improve the efficiency of search over semantic data such as RDF and OWL, which currently proves very resource heavy due to the added complexity of applying Description Logic based reasoning on the data. Currently, search across large knowledge bases can prove too taxing for an off-the-shelf reasoner, leading to long response times to semantic queries or even the failure of the query to return a results' set. In an enterprise setting, we can expect that large knowledge bases will be used which need to be searched quickly and efficiently, yet the current state of the art of Semantic Web querying can not provide this robustly.

Approaches being considered in WP2.1 to enable robust and efficient querying are:

- Approximation
- Modularisation / Distribution

Through query approximation, querying over large knowledge bases should become viable which is necessary for enterprise application. This is illustrated below. The approaches taken to approximate queries are:

- Language weakening
- Approximate deduction
- Knowledge compilation

In particular language weakening in approximated queries can also be used to ensure a result set when a query is too specific and would ordinarily return no answers. Hence there are two reasons why the Recruitment use case was chosen to test this research:

- As the number of job offers and job seekers increase to the scale of present syntactic job portals, we want to ensure robust and reasonably quick answers to user queries (we identified this requirement in D1.1.4v1 – section 2.1.3.2, p9);
- As queries for job offers or seekers could be too specific, we want to avoid that users simply find no responses – rather, we want to support a graded loosening of the query to allow the closest matches to still be found.

2.1.3 Planned Technology Transfer

Initial work in query approximation has focused on **rewriting rules** [DSW06]. Such kinds of rules are an expressive frame work for experimenting with different query relaxation techniques. Rewriting rules describes which parts of a query can be replaced

by which other parts; conditions in rewriting rules allow restricting and controlling the application of the rules.

The basic idea is a sort of middleware between the query engine and the reasoner which rewrites queries to relax/tighten the results. As an example, consider the query for "a Person who has Experience in at least 2 international projects". In DL:

```
Profile ∧ ∃experience.Experience ∧
≥2projekt.(Projekt ∧ ∃in.International)
```

So this can be relaxed in that, for example, we check if someone has experience in at least ONE international project, or relax the condition on projects that they can also be local. In other words, two typical ways to relax/tighten a query are:

- Relaxing or tightening a cardinality restriction
- Generalising or specialising a class membership restriction

Different combinations of rule rewrites produce smaller or larger result sets. In general the application of the rewriting rules span up a replacement tree (Fig. 1).

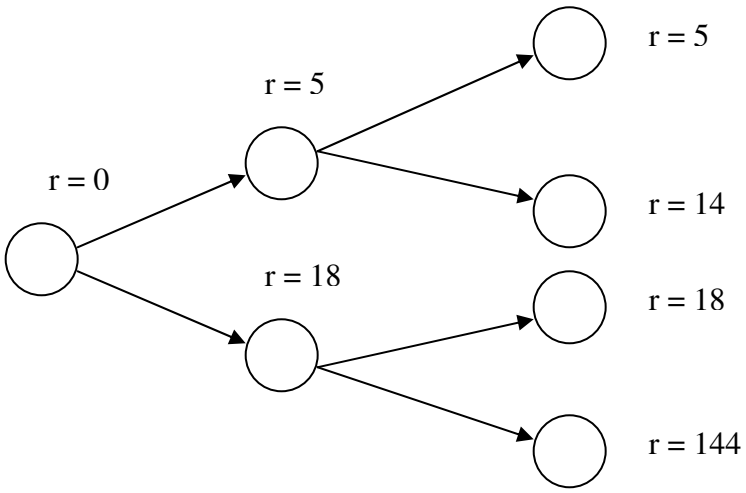


Figure 1. Search results tree.

The system will need to be able to determine which results it will select, e.g. let the user decide through a suitable user interface. This raises the possible need for a form of natural language feedback on rule rewrites (e.g. "Your query found **2** items. By querying on at least one international project, you will find **18** more items. Perform this query?")

For example, in the above diagram a combination of rule rewrites enables a query which has no results to find up to 144 results through relaxing that query. It would seem likely the best result set selection is from the 14 and 18 results (as 5 may be seen as too small, and 144 as too big).

Instead of asking the user which relaxation is chosen next another tactic would be "skylining", which is the policy of skimming the 'best' results from every direction of rule rewriting. Other search strategies are imaginable.

Possibly a core rule set for rule rewriting (meta-rules) could be developed, based on usual means to relax/tighten queries:

- Based on the ontology structure (e.g. OWL-DL): e.g. replace classes with subclasses or superclasses
- Based on mathematical laws: e.g. reduce or increase cardinality restrictions
- Based on the ontology content: e.g. identifying properties expressing preference and using them to select some results over others
- Based on logical rules: e.g. making an inconsistent rule (which obviously returns no results) consistent

All would be based on the knowledge in the ontology and possibly an explicit model of user preferences.

In the HR prototype currently, the single simple rule used to ensure query results is to repeatedly go up a level in the taxonomy tree derived from the ontology until answers are found. The further away a match is from the original query, the lower it is ranked in the results. Hence we are always taking more of the knowledge base into account which is clearly inefficient.

This is illustrated in Fig. 2. In the HR prototype, the match may be found first by returning to the root of the tree. However, by an appropriate rule rewrite, one aspect of the query is replaced by another and the new query can cover another part of the taxonomy tree in finding a potential match.

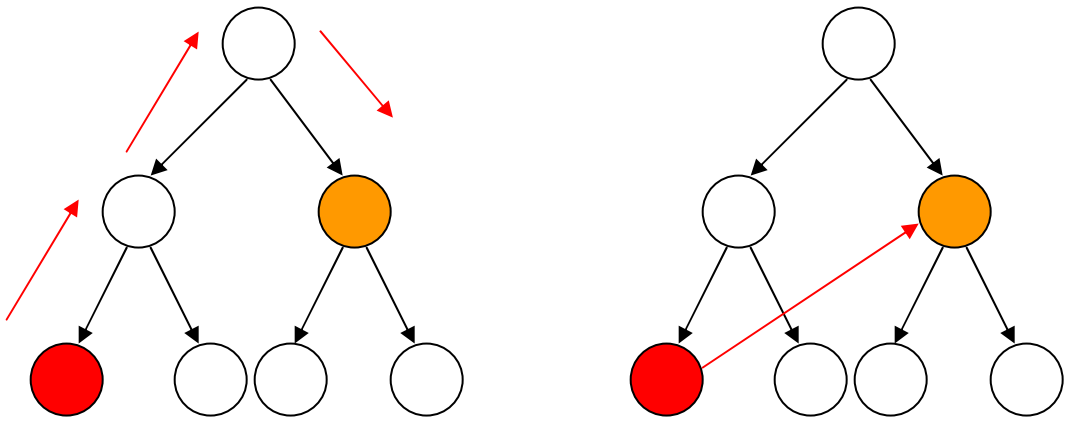


Figure 2. Efficient matching by rule rewriting.

Rule rewriting shows itself to be more expressive and more efficient. The open issue for the HR prototype is what rules for rewriting queries can be determined for this use case?

Hence we intend to produce a set of rewriting rules for the Recruitment use case and test them in the prototype against the current approach to query answering.

2.1.4 Progress to date

At the time of delivery of this deliverable, the following tasks have been achieved:

- Defined ca. 10-20 good examples of queries which will include tests on job applicant's experience against the requirements of the job vacancy
- Extended HR ontology to allow the expression of experience
- Added to the HR knowledge base ca. 250 job seeker & 250 job position instances which include experience.
- Specified the interface to the rule rewriting tool (more general than DIG in that it is not limited to DL)
- Defined concrete technical details of rule rewriting (e.g. abstract query syntax)
- First implementation of the rule rewriting tool

2.1.5 Milestones

By M36 (December 2006) it is planned to have a working extended version of the HR prototype which will support queries which include expressing the amount of experience required from a job seeker or expected in a job offer. A benchmarking set of queries which use the property of experience will be used to test the extended prototype against the original prototype which does not support query approximation.

2.1.6 Conclusion

The aim of query approximation is to allow more robust and efficient query response from knowledge bases which can scale to real world enterprise size. Furthermore, this approach is useful in areas such as Recruitment to loosen queries that are too specific in order to allow users to find best matches rather than simply receive no results at all. Extending the HR prototype provides us with a real life example to test the value of our approximation work.

2.2 Research WP 2.2 : Heterogeneity

Chosen Use Case:	Recruitment
KW Partners:	INRIA, FU Berlin
IB Member:	WorldWideJobs GmbH

2.2.1 Chosen Business Use Case

The Heterogeneity work package has also chosen the Recruitment use case which was described in KnowledgeWeb deliverable D1.1.2 (section 2.1, p11). A description of this use case and the ontologies used in it have been given in section 2.1.1.

2.2.2 Applicability of Research to the Use Case

The Heterogeneity WP concerns itself with the problems of ontology heterogeneity. There are many sources of heterogeneity: differences in terminology (the same concept can be given different names in different ontologies), differences in modeling (the same concept can be defined in different ways), or differences in knowledge representation languages (a concept cannot be defined as precisely in a language as in another). It seeks to find ways to reconcile heterogeneous resources by finding correspondences (between languages, terms or models) and applying it to knowledge, i.e.:

- Translating from one language to another
- Adding "bridge axioms" between ontologies
- Creating database-like views

In the Recruitment use case it is foreseen that job offers and job seekers will be matched by aligning the job offer ontology with the job seeker ontology. The requirement of semantic matching was identified in D1.1.4v1 (section 2.1.3.1, p8) – and is revisited in section 2.2.2.1 of this deliverable. The given requirement analysis upon the Semantic Matcher component reflects the importance of the algorithm for matching the two ontologies in the HR prototype. Given especially the constant changes in and different views upon the field of recruitment, we can consider differences of terminology and modeling to be a very typical aspect of any real world recruitment database. For example, different job titles are used to describe the same or very similar positions in a company, while different countries define their own qualification systems which are often equivalent or partially equivalent to other qualifications.

A number of alignment algorithms are being developed by the members of WP2.2 and the work package has launched the Ontology Alignment Evaluation Initiative (OAEI)[1]. The OAEI as an effort for evaluating alignment algorithms providing benchmark tests and results. The goals of the Ontology Alignment Evaluation Initiative are:

- assessing the strength and weakness of alignment/matching systems;
- comparing the performance of techniques;
- increase the communication among algorithm developers;
- improve evaluation techniques;
- most of all, helping improve the work on ontology alignment/matching.

These will be achieved through the controlled experimental evaluation of techniques' performance.

[1] http://oaei.ontologymatching.org/

While previous alignment tests have been carried out as part of this initiative, they have been criticized for not showing that they contribute to solving a real world problem (though this has never been claimed as a goal in the test). Hence the use cases collected by Knowledge Web were identified immediately as an important source for such real world problems. The Recruitment use case was chosen as the best case due to its matching problem and the dynamics of the recruitment domain.

2.2.2.1 Semantic Matching (from D1.1.4v1)

Problem Statement

Inside both job postings and job applications sub-ontological pieces of information can be grouped into "thematic clusters", e.g. information about competencies and skills, information regarding the industry sector of the job position, and job position details. The thematic cluster from a job posting is compared with the corresponding cluster from a job application. General similarities are calculated from an average of cluster similarities which result from the semantic similarity between concepts. We measure taxonomical similarity through the distance between the concepts, i.e. their respective positions in the concept hierarchy and the attributive similarity which is based on the comparison of attributes and their values. However some limitations to this approach become apparent.
For example, we noticed that, unlike "hard skills" (certifiable such as language or qualification), "soft skills" (uncertifiable such as creativity or teamwork skills) proof to be not as easy to quantify and hence are more difficult to automatically compare. Due to this limitation the last phase of recruitment process (the end decision) still has to be done by humans and cannot be automated even with the use of Semantic Web technologies.
Matchings may also be made more exact by adding weightings to concepts (i.e. describing or measuring the importance of concepts) or associating them with competency levels.

Knowledge Processing Task and Component

The identified knowledge processing tasks are **matching** and **ranking**. Matching refers to the task of discovering relationships between entities in ontologies and measuring the level of similarity between two entities. Ranking refers to ordering match results according to a desired criterion. In this use case, as exact matches between job requirements and applicants are unlikely to happen, the ranking mechanism is used to express the extent to which equivalence might be assumed. Both tasks are considered as occurring in the Semantic Matching Engine. The engine provides a similarity relation between two concepts in the form of a coefficient in the [0,1] range and uses this as the basis for results ranking.

Requirement Analysis

The requirements upon the Semantic Matching Engine, i.e. upon the matchmaking algorithm that it uses, are:

- Support the weighting of concepts as a means of tweaking matching results;
- Support a more precise matching between job position postings and seekers and better ranking of results (i.e. extending the criteria from solely using the similarity

coefficient), including the facility to provide natural language explanations of the ranking;
- Enable different rankings with respect to the thematic clusters defined in the ontology
- Support the consideration of measures in the matchings, such as competency levels (e.g. novice, intermediate, expert) or duration (e.g. number of years in a job).
- Support of parameterization of matching algorithm which allows the matching of other ontologies in other scenarios.

2.2.3 Planned Technology Transfer

Ontology alignment algorithms will be tested as part of the OAEI 2006 campaign using sample "real world" data from the Recruitment use case. The tests will be done by replacing the present matching engine of the HR prototype with submissions from ontology alignment researchers of ontology matching algorithms which are wrapped as matching engines (taking the input and output of the prototype), as illustrated below.

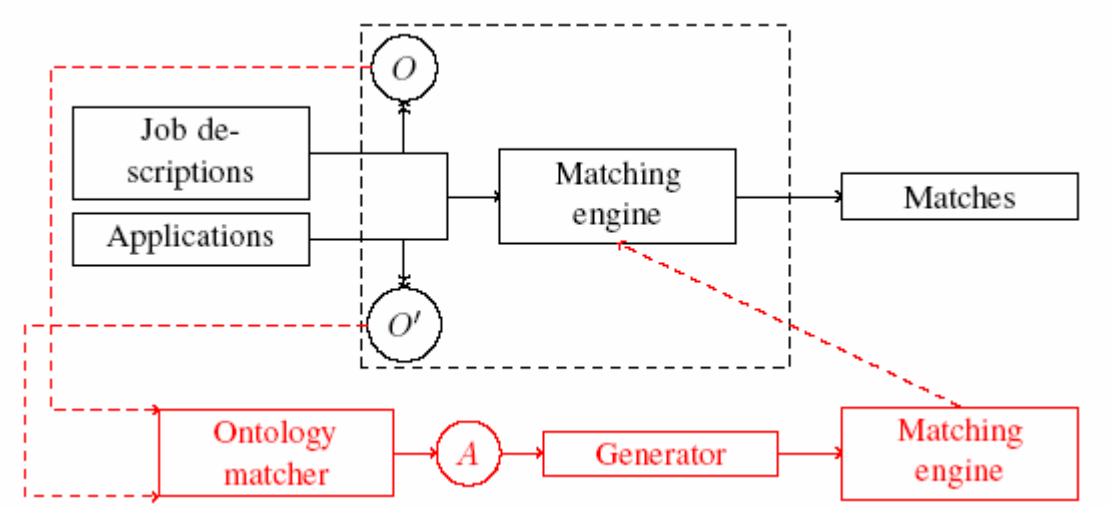


Figure 3. Testing ontology matching in a real world business case.

It is expected that the results of the benchmarking can help guide decision making in the development and choice of Semantic Matching Engines which would be best able to support real world requirements for ontology alignment, matching, merging etc.
It will help assessing the adequacy of state of the art prototypes with regard to a genuine real world need.

2.2.4 Progress to date

Results to date in WP2.2 are:
- An integrated toolbox and format for alignment representation

- Output to various executable languages
- Some matching algorithms

In terms of the co-operation between the Industry use case and WP2.2 a questionnaire was used to collect specific details about the chosen use case for preparing the alignment task. Section 2.2.4.1 gives the questionnaire used with the answers generated by WorldWideJobs GmbH. It is very representative of the challenges that arise when working with industry, which has its own (profit-driven) constraints.

Through this contact between the WP and WorldWideJobs GmbH, we learnt that the ontology used is partially protected by third party rights and could not be made public to the alignment initiative. Hence the following plan of action has been proposed:

1) WWJ identifies a test case (the pair of ontologies in OWL are available but protected, the data to be matched as well as the expected matches is available in large quantity);
2) WWJ identifies where is alignment involved in their process (this is the Semantic Matching Engine);
3) WWJ and INRIA instrument the SME so that the matchers provided by participants can be plugged in the process. This mostly involves transforming the Alignments of the Alignment API into suitable execution tools for the SME;
4) OAEI ask participants to provide their code (e.g., as a zip of java jar files) satisfying a particular API (namely, taking 2 OWL ontologies as input and providing an Alignment in the sense of the Alignment API as output);
5) Participants provide the required code;
6) WWJ run the code and input the obtained alignments into the SME, which evaluates on the given criteria the accuracy of the matching result.

2.2.4.1 Alignment Initiative Questionnaire (filled out by WorldWideJobs GmbH)

* What is the goal of the application and where is ontology matching necessary?

We developed Semantic Web job portal by allowing a uniform representation of job postings and job seeker profiles and semantic matching in job seeking and procurement tasks.

In a Semantic Web-based recruitment scenario the data exchange between employers, applicants and job portals is based on a set of vocabularies which provide shared terms to describe occupations, industrial sectors and job skills. This common vocabulary is defined in the human resources ontology (HR-ontology). Ontologies represent domain specific knowledge, e.g. job postings and applications, and might be used to determine semantic similarity between resources by using semantic matching technique that combines annotations using controlled vocabularies with background

knowledge about a certain application domain. This allows us to compare job descriptions and applicants' profiles based on their semantic similarity and not merely relying on the containment of keywords.

* Is the application actually:
- A prototype

* Is this application
- In-house, but we are open to discuss it

* Is the application for
- internet-based use

* The matching execution
- must be automatic

Questions regarding the dataset(s)
---

* How many ontologies are there to match
- 2 (with sub-ontologies)

* In what language are they expressed
- OWL
- RDFS

* Are these ontologies populated (with instance data)?
- Yes, but only generated data

* Are these instance data identifiable (i.e., it is easy to see that two instances are the same because they share the same social security number or because they have exactly the same representation in both ontologies)?
- Yes

* Can you quantify the size of ontologies in number of concepts, relations, individuals? (roughly)

- concepts: ca. 8650
- relations: 60
- individuals: ca 500

* Are these ontology widely available?
- No they come from other sources that did not made them public
- No they are our strategic assets

* In what natural language are the concepts expressed?
- German

Question about the evaluation
---

* How do you think it is possible to evaluate the performances
- From the final output of the system

* Will it be possible to "instrument" your application so that the
results can be directly feed to it and results measured?
- No, because we do not have resources for such implementation

* Do you have a clear criteria of the success (or better quality) in
the context of your application?
- No

* Would you be available to do it (with our technical help of course)
- Partially, but we have workforce constraints

2.2.5 Milestones

WP2.2 is continuing to research better matching algorithms and INRIA has developed an integrated framework for evaluating matching results.

The web site for the OAEI 2006 campaign has listed the jobs matching problem as a real world business case and there is a dedicated site for describing this problem at http://wissensnetze.ag-nbi.de/oaei/jobs.htm. Hence the details of the problem have already been communicated to the ontology alignment researchers.

Matching algorithms for the job matching problem will be collected by September. In October, the results of the alignments will be published for comments and the OAEI 2006 workshop, in which results will be presented and discussed, takes place in November. Hence we expect to complete all tests and derive conclusions by the end of the year (M36, December 2006).

2.2.6 Conclusion

The Heterogeneity WP is focused on overcoming ontology heterogeneity through supporting the alignment and mapping of ontologies. This support is enabled by algorithms which can analyze ontology structures and concepts and produce alignments between them. Existing evaluations of alignment algorithms have not had access to real world data. The Recruitment use case has been selected as the most appropriate from the Knowledge Web collection to perform first tests on “enterprise standard” data and is

expected to provide important insights into the performance of current alignment approaches. Results from this initial evaluation should help in the production of ontology alignment algorithms which can be used in industry mature Semantic Matching Engines.

2.3 Research WP 2.3: Dynamics

Chosen Use Case:	Application of Wiki Versioning to a Hospital Use Case
KW Partners:	DERI, NUI Galway
IB Member:	HP Galway

2.3.1 Chosen Business Use Case

This workpackage has chosen a new use case which deals with a multi-user environment for accessing and updating staff and patient-related information in a hospital setting. In this use case, a number of different user group requirements have been identified and using these requirements the exploitation of a derived wiki-based infrastructure for within this hospital setting is being investigated together with our partner Hewlett Packard.

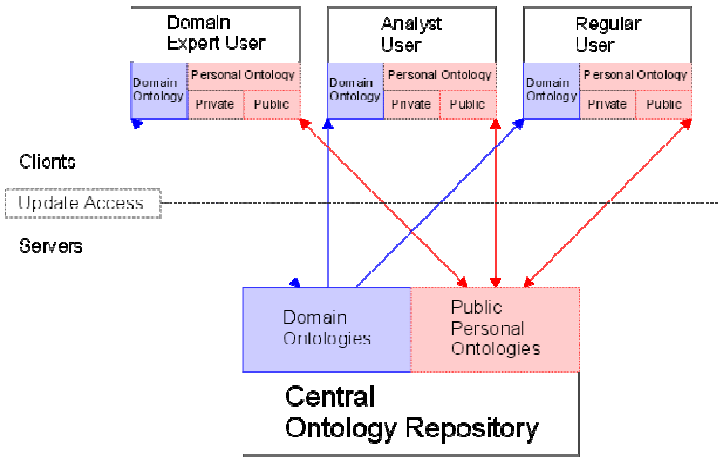


Fig. 4. The system's abstract architecture.

Versioning is required for both the underlying ontologies (which are created and updated by domain experts) and the associated metadata (e.g. when patient records are updated by other users of the system). A hospital is a highly collaborative environment with extreme requirements on the versioning of ontologies and metadata instances.

2.3.2 Applicability of Research to the Use Case

For many practical applications, ontologies can not be seen as static entities, they rather change over time. Support for change management is crucial to support uncontrolled, decentralized and distributed engineering of ontologies. In order to handle ontology modifications in time, a versioning system is needed, to keep track of changes and versions.

This workpackage introduces an RDF-based approach that provides versioning for RDF models and RDF-based ontology languages like RDFS. Our approach is inspired by the classical CVS system for version management of textual documents (e.g. Java source code). The core element of our approach, called *SemVersion* [VT06], is the separation of language-specific features (i.e. the diff) from general features (such as structural diff, branch and merge, management of projects and metadata).

The most elementary modeling primitive that is needed to model a shared conceptualization of some domain is a way to denote entities and to unambiguously reference them. For this purpose RDF uses URIs, identifiers for resources, which are supposed to be globally unique. Every ontology language needs to provide means to denote entities. For global systems the identifier should be globally unique. Having entities that can be referenced, the next step is to describe relations between them. As relations are semantic core elements, they should also be unambiguously addressable. Properties in RDF can be seen as binary relations. This is the very basic type of relations between two entities. More complex types of relations can be modeled by defining a special vocabulary for this purpose on top of RDF, like it has been done in OWL.

The various ontology languages differ in their vocabulary, their logical foundations and epistemological elements, but they have in common that they describe structures of entities and their relations. Therefore RDF is the largest common denominator of all ontology languages. RDF is not only a way to encode the ontology languages or just an arbitrary data model, but it is a structured data model that matches exactly the structure of ontology languages.

Our general idea is the re-use of data management functionality across ontology languages. The relations between different versions of an RDF model or ontology are the same, regardless of the semantics used. Data management deals with storage and retrieval of chunks of data. In our case, the smallest unit of data we store and retrieve is a model (also called 'triple set'). A model is a set of RDF triples. A versioned model consists of a triple set for the content plus an arbitrary number of statements about this model. We thus call this *model based versioning* in contrast to *statement based versioning*.

The versioning system's data model has a repository of projects. They can be created, listed and removed from that repository. A project can hold a number of versioned models. A versioned model is the container for a single RDF model or ontology under version control. It has a root version and also knows all other versions that are direct or indirect descendants of the root version. Versioned models are quite an important concept and give the user the ability to retrieve the right version by e. g. listing all branches or simple getting the most current "main" branch version.

A version is the most central concept. It is a model decorated with metadata. A version knows where it comes from (its parents), has a branch, a label and optionally even a comment and a provenance URI. This meta-information about versions can be managed independent of the versioned artifacts themselves. Thus this management layer can be designed to be very flexible and reusable. The user can commit a model as the successor of a version, create a new version by merging two existing models or commit a diff. Committing diffs is useful, if the models become really large and change only little.

Users can store arbitrary RDF encoded metadata objects for each project, versioned model and most important for each version. This data is stored in the RDF storage layer and linked by RDF statements to the versioning artifact it belongs to. Metadata models are also URI-addressable. This metadata strategy enables a good re-use of the versioning system, as e. g. the evolution log of an ontology engineering tool could be assigned to a version with this mechanism.

2.3.3 Planned Technology Transfer

The integration of the versioning system in the Wiki infrastructure can be made in two places. The first integration place is in the client, on top of the personal ontology, while the second one is in the server on top of the general domain ontology. Practically this respects the way in which the information management was split between the clients and server.

2.3.4 Progress to date

To date, work has been performed on the application of wiki-based technologies to the storage and versioning of ontologised information. Wikis can also be used as prototypical ontology editors, and the revision history in wikis make them an idea way of managing versioned pages (where the pages correspond to versions of the ontology). So far, the GeneOnt ontology has been used as an example of a medical ontology, but ontologies specific to the general patient-hospital domain will be investigated in the future.

2.3.5 Milestones

By the end of this year we plan to extend the current system and to run some tests in its "natural" (or a surrogate of it) deployment environment. The goal is to observe the versioning system's behavior and the users' understanding and adaptability to the new conditions.

Since the use case has as a target a hospital, there are some issues that need to be detailed in regards to the testing and usage of the system. Together with our industrial partner we are still in the process of negotiation with the hospital. And therefore, in order to emulate the target environment in the best possible way, we agreed to run the tests in the hospital's departments affiliated with the university. The users will have different responsibilities, as in an original setting. There will be domain users, having the power to control the general domain ontologies, and public users, with the possibility to control the public ontologies. As a conclusion, the experience gained from working with real users over significant time periods will provide insight into the above mentioned issues and the way in which actual deployments will be realized.

2.3.6 Conclusion

Knowledge domains and their formal representations via ontologies are typically subject to change in practical dynamic environments[2] and therefore the need for versioning support is present. The work in progress in this workpackage represents a methodology for RDF-based versioning that separates the management aspects from the versioning core functionality. As functionality, the versioning system provides structural diff as well as semantic diff. In terms of meeting the requirements, the structural diff will be able to provide an overview of the patients' history, while the semantic diff will offer meaningful information about the patients' clinical status in time.

In the future, the biggest challenge will be scalable reasoning, and we are looking forward to upcoming solutions. Until then, our versioning system represents a multi-language versioning system that will help research and industry to employ ontology based technologies in dynamic environments.

2.4 WP 2.4: Semantic Web Services

Chosen Use Case:	Dynamic Business to Business Integration
KW Partners:	DERI Galway, DERI Innsbruck
Industrial Partner:	Bell Labs Ireland

2.4.1 Chosen Business Use Case

The Semantic Web Services work package and its participants have chosen the use case called "Dynamic Business to Business (B2B) Integration". This use case has been added to the use case collection on the Industry Portal under the rubric 'Service Industry'.

In this use case depicted in the following figure, there is a buyer organization A, which manufactures electronic devices. For particular device, this organization needs specific component, in our case a *display unit X*. This display unit can be delivered by three different suppliers (further referred to as partners), namely B, C and D. The organization intends to build the B2B integration with all these partners and make preliminary

[2] S. Staab and R. Studer, editors. *Handbook on Ontologies in Information Systems*. Int Handbooks on Information Systems. Springer, 2004.

agreement on possible trades. As the partners are large companies, the organization cannot just dictate the way how the B2B integration should happen. Therefore, the organization has implemented separate B2B integration with each partner, namely RosettaNet using RosettaNet Implementation Framework (RNIF) over HTTP with partner B, EDI X12 using Value-Added Network operator over specific network communication with partner C and SAP Intermediate Documents (SAP IDOC) using Web Services standards with partner D.

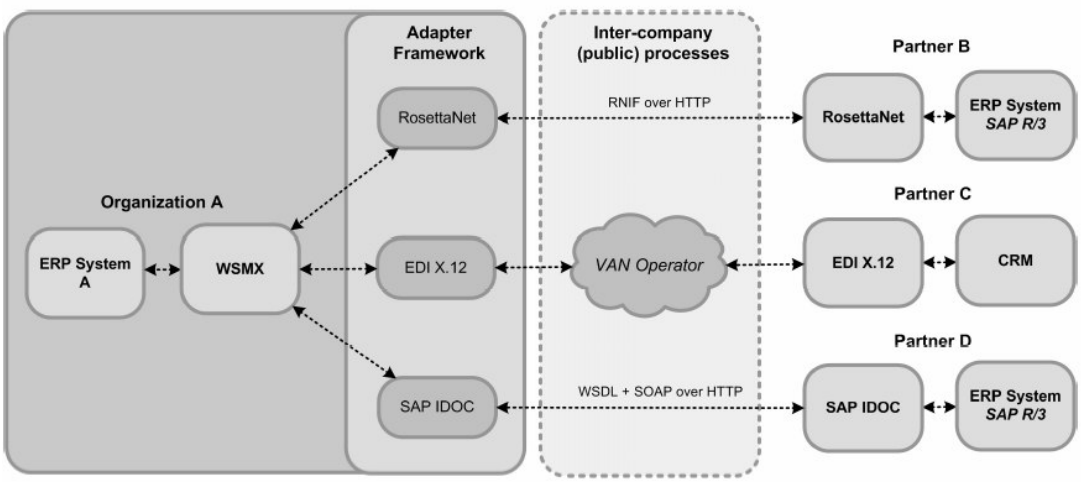


Fig. 5. Dynamic B2B integration over Semantic Web Services

2.4.2 Applicability of Research to the Use Case

The goal of the SWS research applied to this use case is to introduce dynamic B2B integration allowing seamless integration of business partners by building interfaces between the WSMX middleware platform and existing e-business frameworks, mapping e-business messages to a common representation model with the aim to overcome difficulties in data and process integration as well as with the aim to speed up the B2B integration process.

In general, the research in Semantic Web Services is building on paradigms of Service Oriented Architectures (SOA), drawing the way for the new cutting edge technology with the ultimate goal to enable total or partial automation of system integration processes. Today's computerized enterprise environment is facing difficulties in integrating diverse applications within inter- as well as intra-enterprise integration. Therefore, adoption of SOA principles is promising to solve integration problems with respect to ever changing business requirements in a time and cost effective fashion. Current SOA technology is mainly focused around Web Services standards, providing an underlying platform for interoperability of systems at a technical or technological level. With a growing number of services within and outside of enterprises, different languages and formats used by applications and services and different criteria or policies applied for their selection will make the adoption of emerging SOA technologies still difficult. Therefore, support for semi-automated discovery, selection, composition, mediation and invocation of services along with properly planned and methodology-driven SOA design will be essential, especially in large environments. However, in order to apply intelligent

techniques for service discovery, selection, composition, etc., the definition of a formal service model as well as a formalized information model using languages with proper semantic expressivity is essential. In addition, new technology around Semantic Web Services will be only successful if their integration with existing and already deployed integration platforms as well as existing integration frameworks (e.g. e-business frameworks like RosettaNet) will be solved. Research in the WP2.4 Semantic Web Services is therefore primarily focused around the definition of a formal service model (service ontology) known as WSMO, an ontology language for Semantic Web Services (i.e. rule language) known as WSML as well as the integration with existing e-business frameworks such as RosettaNet.

In order to transfer the new technology of the Semantic Web Services and its application to the B2B integration from research to industry, it is essential to build on strong concepts and technology of the SWS, provide clear concepts for integration of the SWS technology with e-business standards as well as existing integration technologies and systems, being aligned with standards and standardization efforts, and last but not least demonstrate the integration on a prototype and a realistic use case scenario including evaluation of added value of semantic integration as opposed to traditional approaches based on pre-defined or hard-wired integration of workflows.

Besides complete concepts and technology for SWS around WSMO, WSML and WSMX defined and implemented for general purpose, it is important to show how the new technology will co-exist with existing e-business standards as well as existing and deployed integration platforms and systems within the enterprise. The SOA nature of the WSMX middleware platform and its open architecture allows for building seamless interfaces with existing systems. User defined execution semantics composed of WSMX middleware services facilitates the overall process of business services in the seamless integration process. So that the uptake of the new technology by the industry is possible, it is important that such technology will be compliant with existing standards while at the same time will be standardized itself as much as possible. Standard based technologies are essential for their seamless integration with existing and future systems as well as allow for wide support by industry which in turn allows for better conditions for maintenance, etc. Regarding standardization efforts in SWS, WP2.4 with the leadership of DERI Innsbruck and DERI Galway are working on the standardization of a SWS architecture and its interfaces as part of the OASIS Semantic Execution Environment Technical Committee. In addition, the W3C Semantic Annotations for WSDL WG and WSDL to RDF mapping (led by DERI Innsbruck) are important steps towards semantically enabled integration based on Web Services.

2.4.3 Planned Technology Transfer

Taking into account all of the above mentioned aspects, the technology transfer is formulated as a plan for how new technology of the Semantic Web Services for B2B integration can be achieved:

1. *Definition of SWS model, language and architecture including discovery, selection, composition, mediation, invocation of services.* This represents the overall research on the SWS concepts and technology around WSMO, WSML

and WSMX and is currently a subject of research in a number of EU funded projects including KnowledgeWeb and work in the WP2.4.

2. *Integration of SWS technology with existing technologies and e-business standards.* This task is subject of research as part of WP2.4 where in the deliverable D2.4.7 Interoperation and Invocation of Services we focus on building the interface between SWS technology and RosettaNet e-business framework. This includes the mapping of selected RosettaNet messages to WSML (lifting/lowering) and definition of general guidelines for such integration.
3. *Standardization of the new technology.* Contributing to the standardization efforts around Semantic Web Services (OASIS SEE TC, SAWSDL WG, WSDL mapping to RDF).
4. *Demonstration of SWS technology for a particular B2B use case scenario and its evaluation.* This task is currently being done as part of SWS challenge. The evaluation is only related to the semantic layer of the integration platform and its added value in the integration process as opposed to traditional approaches.
5. *Commercialization Plan.* This tasks is related to organizational and business issues regarding technology transfer. This includes (1) *Product and Market Analysis* (uniqueness of the solutions and market benefits, collection of potential licensees, mentors, industry groups, commercial intent of the technology transfer (start up/licence), (2) *Industry Analysis* (overview of the industry where the product/service is positioned, focused market leads – CEO, Venture Capitalists), (3) *Market Engagement* (demonstrator, business plan). This task is currently not tackled in KW WP2.4.

2.4.4 Progress to date

Regarding demonstration and evaluation, we (DERI Stanford, DERI Innsbruck and DERI Galway) decided to organize the Semantic Web Services Challenge, which will address requirements of the B2B use case by means of various technologies (please see more information on SWS Challenge initiative at the SWS challenge web site[3]). In the SWS Challenge, the WP2.4 contribution in cooperation with Bell Labs lies in the WSMX technology, its concepts and architecture as well as its integration with existing standards (RosettaNet) and interoperation, mediation and invocation of services. WSMX and its application to B2B integration showcasing data and process mediation was successfully evaluated at the workshop in contrast with other contributions.

Following is the summary of technology transfer to date according to the list from the previous subsection. In detail, regarding our collaboration with Bell Labs Ireland we worked on a supply chain scenario aligned with a B2B integration scenario of the SWS challenge. In addition, Bell Labs Ireland provides requirements for supply chain management regarding policy management and integration of rules within the service discovery and selection process.

[3] http://www.sws-challenge.org

- **Point 1** – We are currently in the stage where about 40% of SWS concepts and technology are defined and implemented (point 1 in the list) (taking into account all research and development efforts across SWS projects – all EU and national funded).
- **Point 2 –** We have created initial guidelines for integration of WSMX with an e-business standard, namely RosettaNet and the mapping of RosettaNet messages to WSML in the deliverable D2.4.7 Invocation and Interoperation of Services. It is expected that such work will provide basic input for technology transfer showing how our semantic technology can co-exist with existing e-business standards triggering additional effort in industry as well as research. It is important to note that such work is essential for evaluation, however it will not provide a complete solution. The complete solution should be provided by the industrial partner after successful evaluation and uptake of the technology. So far, we haven't tackled integration of our technology with other industry integration technologies nor existing systems in the enterprise. We do not plan this work to be done in WP2.4 however this will be subject of work in other research and development projects.
- **Point 3 –** SAWSDL WG has been launched in March 2006 and is chattered for one year. OASIS SEE TC has been launched in November 2005 (there is already working draft of the architecture available). It is expected that new working groups or technical committees regarding Semantic Web Services will be created within W3C and OASIS after additional exploitation of SWS research and development results.
- **Point 4 –** The evaluation in the context of the SWS challenge showed that our solution based on WSMX successfully fulfilled requirements for data and process mediation and that there were required only changes in the descriptions of services rather than changes in the code. The challenge is now in the second phase where our solution regarding data and process interoperation will be demonstrated and evaluated according to the defined criteria. Other phases will follow with focus on simple discovery as well as discovery including contracting and negotiation.
- **Point 5 –** We haven't done any work on commercialization within KW WP2.4. However such activities will be subject of work within commercialization efforts of DERI.
-

2.4.5 Milestones

Following is the list of milestones regarding the plan of technology transfer listed in section 2.4.4.

- **Point 1** – Improved data mediation will be available by M30 (deliverable *D2.4.12 Data Mediation in Semantic Web Services*) and improved architecture for the Semantic Web Services will be available by M42 in deliverable *D2.4.10 Architecture for the Semantic Web Services*. There is however other work going

on in other projects which is essential for the complete SWS model, language and architecture, which is a prerequisite for industry uptake.

- **Point 2** – The final version of the deliverable *D2.4.7 Interoperation and Invocation of Services* (RosettaNet and WSMX integration and guidelines for e-business standard integration with WSMX) will be available by M30 of the project. There is additional ongoing work in other projects (e.g. DIP, SUPER, SemanticGov). There is no planned activity in KW WP2.4 dealing with integration of WSMX with existing systems in the enterprise. In other projects we however plan with Bell Labs to integrate policy management systems (Vortex) within the discovery and selection process of services. Also, we plan to integrate WSMX with Software AG CrossVision registry and repository (CentraSite) and to engage with IBM to work on extension of IBM WebSphere registry with semantically annotated WSDL descriptions. In the future we also plan to target other scenarios from the area of discovery of services with increasing complexity.
- **Point 3** – The first working draft of the W3C SAWSDL WG will be provided in June 2006. By March 2007, the final working draft will be available triggering the standardization process (e.g. call of comments, call for implementations, etc.). In OASIS, there is already a working draft for the architecture available. Both standardization efforts are supported by WP2.4. Standardization is vital to reassure industry of the stability of their investments in this new technology.
- **Point 4** – The first evaluation results for our technology will be available by M30 and documented in deliverable *D2.4.13 SWS challenge*. Milestones for other phases of the SWS challenge will be available subsequently. Such results are important to demonstrate the industrial viability of the technology.
- **Point 5** – There is no planned activity regarding commercialization within the WP2.4.

2.4.6 Conclusion

In order to successfully transfer the new technology to industry, several tasks should be completed including a complete SWS model and architecture, integration of SWS technology with existing e-business standards, integration of the new technology with existing integration platforms and systems within the enterprise, standardization of the technology, demonstration and evaluation of the new technology as opposed to traditional/existing integration approaches and a plan for commercialization of the new technology. Not all tasks are however a subject of work in WP2.4. Work in WP2.4 considerably contributes to this process mainly along the lines of general SWS concepts and architecture, e-business standards integration, standardization efforts and evaluation of the technology. With this respect, it is important to note that development of the SWS technology is a long term effort which can't be fully satisfied within one project but requires more partners to be involved working on that technology from different perspectives, in different application domains and on different levels (conceptual/theoretical, development of core platform, implementation of platform to particular application domains). With this respect, the plan for the technology transfer is a long process too. Our objective is to complete all aspects of the plan by the end of the year of 2008. In the meantime it is important to focus on demonstration and evaluation of

intermediate results. This will trigger interest for commercialization followed by the essential software engineering approach driven by a software company towards development of the new technology with industry strength.

2.5 Research WP 2.5 : Language Extensions

Chosen Use Case: Health Care Kidney Case
KW Partners: VUM, France Telecom
IB Member: Institute de Biomedicine/Uni Rennes

2.5.1 Chosen Business Use Case

The overall goal is to improve dialysis and organ transplantation decisions within the French health care system. To this end, data from multiple databases maintained by hospitals as well as dialysis and kidney transplant centers are combined with newly-solicited data to form a centralized data repository. A unified query interface is provided over the aggregated knowledge base. A semantic approach is important in order to mediate between different schemas and terminologies used by different databases, as well to offer a richer, more expressive language in which medical knowledge can be expressed in addition to raw patient data. A full use case description is available to Industry Board members on the Knowledge Web Industry Portal.

2.5.2 Applicability of Research to the Use Case

Work package 2.5 focuses on extensions to existing semantic web languages as well as related new languages, including query and rule languages.

2.5.2.1 Expressive ontology language

Problem Statement

This use case makes use of complex medical, biological, and anatomical knowledge, some of which cannot be captured using OWL-DL version 1.0. For example, the knowledge that a person can have at most two kidneys should be expressible, as should the knowledge that a disorder of one part of a kidney affects the entire organ. A more expressive core ontology language is necessary, and the syntax and semantics of the extended language must be standardized in order to provide interoperability between authoring, reasoning, and querying tools.

Knowledge Processing Task sand Components

Data Translation, Ontology Management, Content Annotation, and Reasoning, as well as interfaces between all components

Requirements Analysis

Specific user needs include support for qualified number restrictions and partonomy semantics. A well-defined ontology language with this increased expressiveness is required.

2.5.2.2 Standardized rules language

Problem Statement

Some aspects of the use case, particularly with respect to mapping between properties in different ontologies, are described by users not in terms of ontological modeling, but by "rules" capturing additional knowledge. For example, the relationship between the `connectedTo` property in one ontology and the `boundedBy` property in another is given by a rule involving both properties and class memberships of instances which participate in property relations:

```
connectedTo(?x1,?x2) ← boundedBy(?x1,?x3)
        ∋ boundedBy(?x2,?x3) ∋ MAE(?x1) ∋ MAE(?x2)
        ∋ GyriConnection(?x3)
```

Some form of such rules must be supported in a standardized way to allow sharing between institutions.

Knowledge Processing Task sand Components

Data Translation, Ontology Management, Content Annotation, and Reasoning, as well as interfaces between all components

Requirements Analysis

A rules language standard must be defined.

2.5.2.3 Expressive query language

Problem Statement

The goal of the use case is not merely to provide knowledge models of kidney function for human exploration, but to actually answer queries about the implications of the

defined knowledge. For example, users wish to use the system to determine whether a particular course of action is appropriate to a specific patient based on the rich semantics encoded in the ontology. A standard query interface to the unified knowledge store is necessary.

Knowledge Processing Task sand Components

Matching, Matching Results Analysis, Reasoning, Semantic Query Processing, Results Reconciliation

Requirements Analysis

The unified knowledge store is expected to use expressive ontology and rules languages; a query interface is needed which can access this data in the context of its rich semantics. Again, standardization would increase interoperability and reduce application development costs.

2.5.3 Planned Technology Transfer

The basic research underpinning solutions for the cited requirements is expected to be transferred to industry largely through standardization of languages based on the developed formalisms as well as implemented systems derived from algorithms and implementation techniques resulting from this research.

Further, technology transfer will continue to operate in both directions, as industry needs continue to drive research toward solutions with wide applicability and power within realistic usage scenarios. This is illustrated by the activity of both the research and industry partners in the W3C Rules Interchange Format WG where they have presented requirements relevant to language extensions[4] arising from this use case.

2.5.4 Progress to date

The basic research underpinning expressiveness extensions to OWL and query languages has already been completed as part of the KnowledgeWeb project, with rule language integration investigated in D2.5.1, query language formalisms described in D2.5.2, and a query optimization and implementation techniques, as well as a developed prototype query platform, presented in D2.5.3.

Industry use cases have been analyzed to determine where further language extension is needed and specific solutions proposed in D2.5.4. Research within the work package has focused on developing rich extensions to the core OWL-DL language to improve its

[4] See http://www.w3.org/2005/rules/wg/wiki/SW_rules_for_Health_Care_and_Life_Sciences and http://www.w3.org/2005/rules/wg/wiki/Decision_making_in_Health_Care

expressiveness. A comprehensive and rigorously-specified proposal for an ontology language which incorporates many of these extensions, including qualified number restrictions and powerful role-composition operators which can be used to implement “partonomy” semantics, while retaining important properties of OWL is presented in D2.5.5. A great deal of work has already been devoted to developing and standardizing the SPARQL query language for RDF, which has entered "last call" status. A formal setting for rules language standardization (The W3C Rules Interchange Format Working Group) has been established.

2.5.5 Milestones

The extended ontology language presented in D2.5.5 will be presented as a formal member submission to the W3C by M36, with 'de facto' standardization of those extensions and wide tool support to follow within six to twelve months. SPARQL will proceed through W3C standardization, again with tool support expected to become common before M42. Formalization of ontology systems which work with imprecise knowledge, along with practical experience, will be described in D2.5.6 (M36) which will inform the applicability of such techniques to real-world use cases. Integration between different semantic layers and user interaction paradigms will be investigated in deliverables expected in M42 and M48. Members of the work package will continue to assure that this use case is among those considered during design of the Rules Interchange Format specification, which will continue for some time. Given the availability of prototypical tools and systems supporting the language extensions next year, the industry member is prepared to perform tests in coordination with Knowledge Web partners.

2.5.6 Conclusion

Work Package 2.5 expects that continued work which balances basic research in language extensions and additional semantic layers along with formal standardization of well-understood constructs will continue to the benefit of industry use cases, including the one described.

3. Outlook

Even with the ongoing Industry-Research co-operations detailed in the last chapter, the world of business and of enterprise IT solutions is moving rapidly and it is important that we remain aware of emerging challenges and their potential (semantic) solutions. For this reason we introduce an emerging multimedia scenario, being aware of the increasing importance that multimedia data will have in future enterprise networks.

Additionally Knowledge Web is a funded network only until the end of 2007. However, the Semantic Web will not cease to exist at this time. Rather, as a long term technology, we expect the Semantic Web and Semantic Web Services to be gradually taken up into industrial systems and by the end of Knowledge Web we expect to be at the beginning of this process. Hence it is vital that Knowledge Web establishes the basis within its funded duration for making it possible to continue to support Semantic Web research for industry requirements, making Semantic Web technology industry-ready and transferring Semantic Web results into the industry partners.

3.1 Emerging scenario

In order to be in a position to bring Semantic Web technologies to industry, it is also important to be aware of emerging scenarios and trends which are both of increasing interest to industry and are requiring the use of semantic technologies. While the Industry-Research co-operation outlined in this document is already covering a number of key business problem areas with semantic solutions, we consider multimedia as a key emerging scenario which is also deserving of our consideration from an Industry-Research perspective. In our scenarios semantics are being applied largely to textual data, or describing a business process in the case of Semantic Web Services. However, it is clear that the data being handled in industry is increasingly also non-textual and that this data must also be made available to semantic-based systems (whether for retrieval, adaptation, organization, integration or presentation purposes) through high level annotation and specific knowledge components developed for using such annotations.

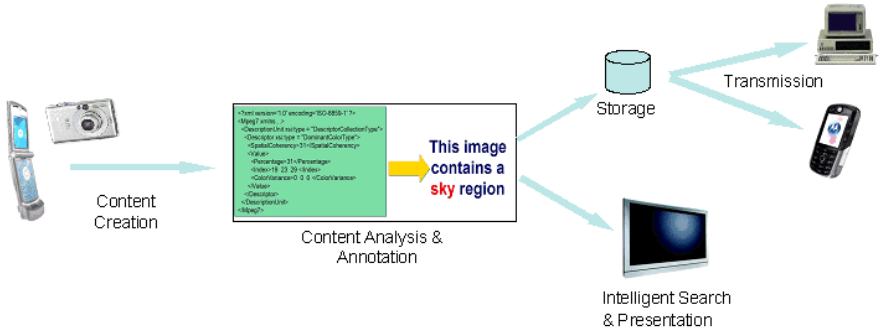
Hence, we introduce in this deliverable an emerging scenario focused on multimedia and semantics, based on a further analysis of the second use case of D1.1.2 which was provided by the industrial partner Motorola in co-operation with the EU project AceMedia.

3.1.1 Multimedia Content Analysis and Annotation

Topic: Automated Semantic Multimedia Annotation
KW Partner: CERTH
IB Member: Motorola

General Description of Business Use Case

This business use case describes the aceMedia "Integrating knowledge, semantics and content for user-centred intelligent media services[5]" 6th FP Integrated Project that focuses on generating value and benefits to end users, content providers, network operators, and multimedia equipment manufacturers, by introducing and developing a system based on an innovative concept of knowledge assisted, adaptive multimedia content management, addressing user needs. The main technological objectives are to discover and exploit knowledge inherent to the content in order to make content more relevant to the user; to automate annotation at all levels; and to add functionality to ease content creation, transmission, search, access, consumption and re-use. In addition, available user and terminal profiles, the extracted semantic content descriptions and advanced mining methods will be used to provide user and network adaptive transmission and terminal-optimized rendering.



aceMedia in the content value chain

Current approaches operate either on the assumption of manually annotated content that has to be organized and managed by the end user, or on query-by-example, typically PC-based, search and retrieval approaches that build on the premise of emulating humans' perception of visual similarity. As a result, access to multimedia content is hindered due to the high cost of manual annotations, the incompleteness of the textual annotations and the lack of an underlying conceptual framework, thus requiring still significant effort from the end user in sorting and selecting relevant content.

Proposed Semantic Web-based Solution

The aceMedia project (http://www.aceMedia.org/aceMedia) utilizes domain, multimedia and general knowledge and provides a conceptualization for modelling and performing audiovisual content annotation, self-organization, and intelligent search and retrieval at a semantic level. The general overview of the proposed solution is illustrated in the following, highlighting a number of open research issues with potential SW-based solutions:

- ✓ Multimedia descriptor and structural modelling and representation

- ✓ Ontology-based framework and tools for linking multimedia representations to domain ontologies
- ✓ NLP (for user-entered annotations & queries)
- ✓ Personalized retrieval
- ✓ Privacy policies
- ✓ Semantic reasoning in multimedia annotation
- ✓ Context ontologies
- ✓ Content self-organization at semantic level

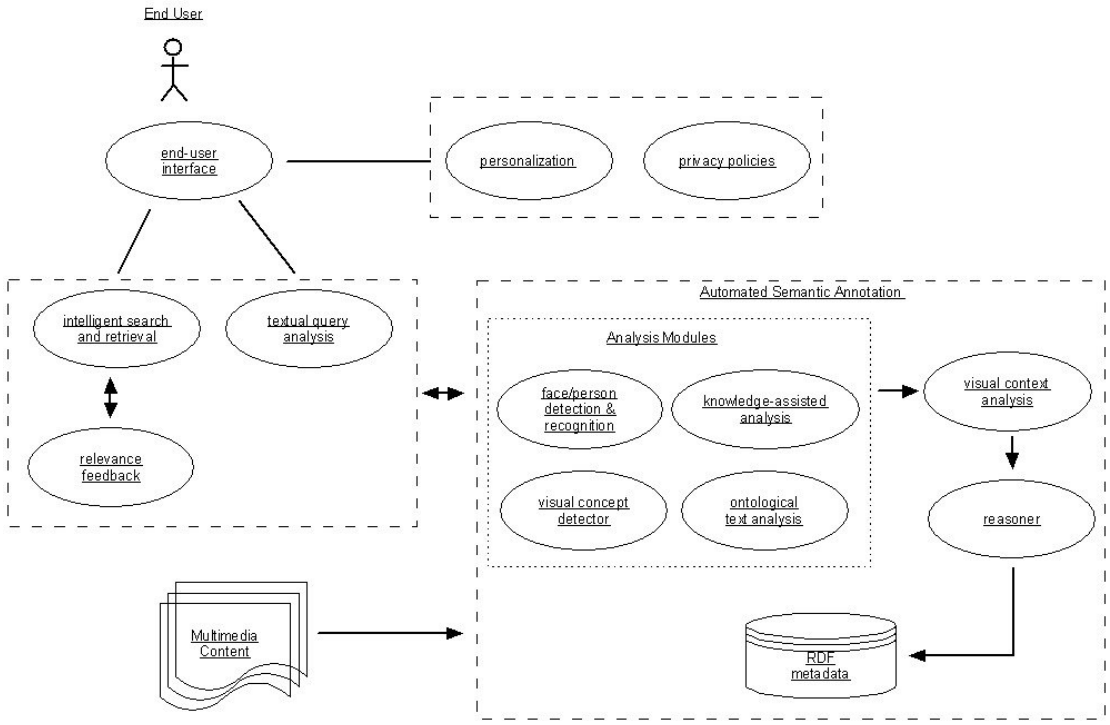


Fig. 6. Proposed semantic multimedia solution

Identified Research Challenges

1. Automated Content Annotation

Problem Statement

There is an urge to move from the numerical low-level representations that can be automatically extracted to high-level ones that meet a human level of cognition, i.e., semantic annotations.

Knowledge Processing Task

The **Annotation Manager** is the component responsible for performing knowledge-assisted analysis by utilizing the provided prior knowledge for the content under consideration.

Requirements Analysis

Multimedia come in two intertwined layers, namely the content one that refers to the actual meaning conveyed and the media one that relates to the structural and descriptor aspects. Consequently, in order to achieve automated multimedia content annotation, the interrelations between the actual content and the information carried need to be sufficiently understood to enable effective analysis.

2. Ontology Development/Management.

Problem Statement

Appropriate ontologies need to be developed to represent media-related aspects, i.e., the media structure and multimedia descriptors used to describe such content, and to link them with the domain and general knowledge definitions. Additionally, modeling solutions are required for representing contextual information with respect to the considered domain conceptualization.

Knowledge Processing Task

The development and maintenance of the domain and multimedia ontologies is handled by the **Ontology Manager** component.

Requirements Analysis

As ambiguity is inherent in multimedia, coming with a solution that allows incorporating uncertainty into the low-level descriptor representations associated to domain concepts and the context definitions is of crucial importance. Adherence to standards such as MPEG-7 is desired for enhancing interoperability, however careful modeling is required to preserve the intended semantics and to provide support for the definition of data-types common in the multimedia domain. The introduction of context relations on top of domain ontologies is required for modeling solutions of sufficient expressivity while preserving well-defined inference services. Finally, a significant issue refers to the fact that the adopted modeling solutions and languages should not only serve as means for "elegant" knowledge representation of the multimedia-related aspects, but they should allow inference on top of them.

3. Support for Reasoning in Multimedia Annotation.

Problem Statement

Automated extraction of semantic annotations from multimedia content is a complicated task that involves the fusion of the different and incomplete (on their own) analysis modules' outputs. The objective of reasoning is to invoke inference on the metadata produced by the individual knowledge-assisted analysis modules in order to derive high-level semantic representations of the content. The reasoning process thus, has to ensure consistency of the final annotations as well as their further enhancement, for instance by events (by exploiting spatiotemporal knowledge) or by deriving more abstract descriptions based on the provided domain conceptualization.

Knowledge Processing Task

The **Reasoner** is the component in charge of reasoning on top of the metadata produced by the individual analysis and context modules, and generating the final annotation.

Requirements Analysis

A trade-off between expressivity and computability is necessary. As the different analysis modules, as well as the context analysis one, provide hypotheses on the analyzed content semantics, i.e., graded annotations, it is essential for the reasoner to be able to handle such kind of uncertainty. This translates in providing the means to support fuzziness both in knowledge representation and the subsequent inference. Although initiatives for fuzzy OWL and fuzzy SWRL have been reported, the lack of corresponding reasoners results in *ad hoc* solutions tailored to the application at hand. Combining rules and ontologies forms another issue for multimedia automated annotation and retrieval: ontologies provide the well-defined semantics required for ensuring knowledge sharing and inference, while rules are more flexible in terms of expressivity and closer to human cognition but lack a well-defined conceptualization.

4. ***Personalization and Media Adaptation***

Problem Statement

Content delivery and management needs to adapt to user preferences and corresponding profiles to ensure an enhanced user experience, as needs and requirements change according to the context within a user is acting and the corresponding device.

Knowledge Processing Task

The **Profiler** component is responsible for acquiring and updating user preferences by monitoring user behavior and learning how the current context of use/work/environment etc. associates with her search and retrieval needs.

Requirements Analysis

To allow for such functionalities, appropriate models need to be developed to model the semantics of user preferences and device capabilities and context of usage in accordance with the knowledge structures that drive the generation of the content annotations.

5. ***Intelligent Search and Retrieval***

Problem Statement

To enhance the user experience and interaction with the system, different approaches to retrieval need to be considered. The user could type a textual query describing the desired content, she could perform concept-based navigation and search, provide an example image/video, or even combine a visual similarity query with a concept or textual one (hybrid query) to allow each to benefit from the other.

Knowledge Processing Task

The **Query Processor** is the component that handles the free-text and visual example queries posed by the end users.

Requirements Analysis

In all cases the semantics of the user query need to be considered. In the first case the textual query needs to be translated to an ontological representation, while in the query-by-example case, the semantics of the visually similar content item need to be obtained. To improve accuracy and reliability in the case of query by visual similarity, a relevance feedback mechanism has been introduced to ensure that the intended user semantics are correctly captured. The hybrid query is a special case where the semantic and the audiovisual metadata both need to be taken into consideration in order to allow for retrieval based on the fusion of their corresponding semantics.

Given the size of multimedia collections, scalability raises an important issue, especially in the case of content providers and commercial content delivery services.

6. ***Security and Trust***

Problem Statement

An important aspect of the approach followed within the aceMedia project is to facilitate the sharing of content among users. Naturally, issues of security and trust emerge that require for different privacy policies according to the content item and corresponding user in consideration.

Knowledge Processing Task

The principles of the personal content ownership policies are based on digital rights standards from the Open Mobile Alliance (OMA) with particular focus on the rights

expression language, which captures some core concepts for commercial content that is applicable to personal content. The design of the policy model uses these core concepts and will be integrated with the personalisation system.

Requirements Analysis

A trust architecture is required to allow for deciding which content should be shared with which of the corresponding user contacts and which content is considered trustworthy.

3.2 Basis for ongoing Industry-research co-operation

It is important to note how the Industry-Research co-operation will also be supported and utilized across Knowledge Web during the lifetime of the network as well as how a basis will be established for supporting ongoing technology transfer also beyond the timeframe of the network.

3.2.1 Industry: success stories and dissemination

In the Industry Area, WP1.4 and WP1.5 focus on dissemination of results to industry. In WP1.4 a number of reports on success stories are foreseen as well as the organization of a technology showcase. WP1.5 focuses on cross-network co-operations. It is clear that results from the Industry-Research co-operations should be disseminated using these channels.

Given that first demonstrable results are given by the Research WPs at milestone M36 (December 2006), we have focused on industry events in 2007 where we could present concrete results of Semantic Web technologies solving business problems to an industrial audience. The form of event is not fixed, and could be a workshop, tutorial or panel event. In particular, a demonstration event was suggested as industry wishes to see real concrete uses of semantic technologies addressing business problems, which is precisely what the Industry Research co-operation aims to show. Discussions have mentioned both ESWC and WWW 2007 conferences, both of which have had in 2006 a significant industrial presence and had held industry-specific events in the area of Web and semantic technologies. Furthermore, in the interests of cross-network strengthening of our demonstration of the value of semantic technologies, both the AceMedia and REWERSE projects were discussed as a potential co-operation partner in the event. The Industry Area has already been involved in a successful event for industry in 2005 (Semantic Web Days) and if a similar event is organized by Knowledge Web and REWERSE again in 2007, it would be a strong platform for presenting the results of this co-operation to a broad industrial audience. AceMedia, which deals with another important area of semantic technology application which is gaining importance in an industrial setting – i.e., multimedia data organization – would make an ideal companion in a demonstration event, as semantic solutions from this co-operation which deal primarily with annotation

and use of textual data would be complemented by semantic solutions dealing with annotation and use of multimedia data.

3.2.2 Education: events and materials for professionals

In the Education Area, focus is made not only on educational resources for students and researchers but also since the beginning of the network it was acknowledged that providing suitable educational resources for professionals would also be very important for wider Semantic Web uptake. In this sense, the activity of the Education Area complements well the Industry-Research co-operation.

A dedicated task has been established for Industry-Education co-operation and the first report from its activities was released as D3.2.9. From our experiences with a first set of learning materials based on the use cases collected in the Industry Area and a tutorial which used those learning materials, we were able not only to establish that this approach is of definite interest to business professionals but also that what is needed is concrete results of the application of semantic technologies to the business problems which we had described.

As this is the aim of the Industry Research co-operation and we hope by early 2007 to have some initial results we see the Industry-Education co-operation task as fundamentally consisting of transferring those results into educational resources, both learning materials which can be uploaded and made available on the educational Web repository REASE as well as tutorials based on those learning materials which should be co-located with an industry event in order to maximize its reach to interested business professionals. These materials shall also be promoted to the Industry Board members through the Industry Portal and newsletter so that they can benefit from these results, and not only the few members whose use cases have been used.

We will measure the success of our work by the amount of material for business professionals that can be produced from the experiences gained in the Industry-Research co-operation. That business want to learn about the Semantic Web through concrete use cases has been established in prior work; the early adopter experiences of applying Semantic Web technologies in enterprises will prove an important measuring stick for the potential and limitations of the technology and can be developed into presentations and tutorials.

4. Conclusions

The purpose of this deliverable has been to provide an update of progress in the Industry Research co-operation.

The overall target of this co-operation by the end of Knowledge Web would be:

(1) successful transfer of Semantic Web technologies (prototypically) into enterprise environments (where research is mature enough)

(2) orientation of the Semantic Web research to meeting industrial requirements/producing industry ready tools (where research is not yet mature)

(3) dissemination of results in industry events ("showcases" in WP1.4)

(4) production of educational materials based on concrete business cases for use in Semantic Web education for industry

D1.1.4v3 in M42 (June 2007) will report on the further progress and success in this activity, as well as outline what would still be to be achieved before the end of the network and how we will ensure that the benefits of this task can continue beyond to take effect after the network's lifetime.

Bibliography

[DSW06] Peter Dolog and Heiner Stuckenschmidt and Holger Wache: *Robust Query Processing for Personalized Information Access on the Semantic Web***.** In Proceedings of Flexible Query Answering Systems, 7th International Conference, FQAS 2006, Milan, Italy, June 7-10, 2006. Lecture Notes in Computer Science 4027, pp343-355.

[VT06] Max Voelkel and Tudor Groza. *SemVersion: An RDF-based ontology versioning system.* In Proceedings of IADIS International Conference on WWW/Internet, Murcia, Spain, Oct. 2006.