Towards a Successful Transfer of Knowledge-based Technology to European Industry

Alain Léger and Francois Paulus
France Telecom R
D
4 rue du clos courtel Cesson Sévigné Rennes, France

alain.leger, francois.paulus@

francetelecom.com

Lyndon J.B. Nixon
Freie Universität Berlin
Institute of Computer Science
14195 Takustr. 9
Berlin, Germany
nixon@inf.fu-berlin.de

Pavel Shvaiko
University of Trento
Sommarive 14, 38050, Povo, Trento, Italy
pavel@dit.unitn.it

ABSTRACT

Knowledge Web is an EU FP6 Network of Excellence whose mission is to strengthen European industry and service providers in one of the most important areas of current computer technology: Semantic Web enabled e-work and e-commerce. This paper presents the initial results from our activities in outreaching to industry, which has involved collecting and analysing business use cases to identify concrete needs in industry and their potential ontology-based solutions, and developing a typology of knowledge processing tasks and components. Our approach is based on providing European Semantic Web researchers with industrial requirements from concrete scenarios, in order to realise the needed task- and component-based solutions and to migrate them to European industry. We see much potential in the next years of the project to see a successful transfer of ontology-based technologies from academia to industry.

1. INTRODUCTION

The KnowledgeWeb Network of Excellence has as one of its key goals the transference of knowledge-based technologies (often referred to when deployed over the Web infrastructure as the Semantic Web) from the field of academia to industry. In order to achieve this we are collecting industrial requirements and communicating them to researchers so that Semantic Web research can better focus on producing the knowledge-based technological solutions that industry needs.

The 'Outreach to Industry' activities of KnowledgeWeb are the continuation of the OntoWeb project's SIG4 [1], which formed an industrial network for the stimulation of technological transfer to concrete business cases. Likewise, Knowledge Web forms an industrial board of firms interested and active in the application of Semantic Web technologies. By the end of the first year of the project as many as 50 companies from various different industry sectors have joined. A Web portal [2] has been set up to form a communication channel between the project and its industrial partners.

In order to carry out this approach, the Industry Area of KnowledgeWeb has been tasked with, among other things:

- Specifying business needs scenarios collected from the industry board
- Identifying problems in industry that can be successfully treated with the Semantic Web and technology locks blocking that treatment
- Identifying the knowledge components and processing mechanisms that Semantic Web applications will need
- Using co-operation between researchers and industry to migrate knowledge-based technologies to key business scenarios
- Showing the value of Semantic Web-based solutions to industrial partners

In the first year of the network, the 'Outreach to Industry' activity has published three deliverables, all of which are publicly available from the project website [3] :

D1.1.1v1 - The Industry Board members initial list, clustering and organi

, In deliverable D1.1.1 we describe our establishment of an industrial board composed of industrial organizations that can clearly benefit from ontology technologies. The board is considered representative of the current or prospective benefactors of Semantic Web technologies. We list the current

[1] http://sig4.ago.fr

[2] http://knowledgeweb.semanticweb.org/o2i

[3] http://knowledgeweb.semanticweb.org

board members, give the Memorandum of Understanding specifying the working rules agreed by all the members of the industrial board, the invitation letter sent to potential partners and a statement of interest filled by interested companies to join the board;

D1.1.2 [?]- Prototypical business use cases , In deliverable D1.1.2 we provide a set of business cases on how Semantic Web technologies have solved or could hypothetically solve concrete business issues which are relevant to some strategic industries. It also considers future developments in research and industry which are of relevance to the deployment of Semantic Web technologies in business cases. From this an analysis of the potential of and the needs for deployment of Semantic Web solutions in industry is carried out and a few key use cases are identified;

D1.1.3 [?]- Knowledge processing tasks , In deliverable D1.1.3 we have demonstrated a methodology for identifying knowledge processing tasks and corresponding high level components within the information systems by analyzing in detail some of the use cases of D1.1.2. We have developed a typology of knowledge processing tasks with respect to each use case and a general typology covering requirements of all the use cases together. The process of building a typology of knowledge processing tasks has shown that most of the knowledge processing tasks identified repeat with some variations/specificity from use case to use case. This observation suggests that the constructed typology is stable, i.e., it contains the core knowledge processing tasks stipulated by the current industry needs.

In this paper we summarize the results from the latter two deliverables. In Chapter 2 the collection of business use cases is described. Chapter 3 summarizes the analysis of those use cases, providing an informative overview of where European industry is seeking solutions from knowledge-based technologies and what is needed from research to realise those solutions. Chapter 4 describes how we produced a typology of knowledge processing tasks and components as a basis for identifying key research areas for meeting industrial needs. Finally, in Chapter 5 we conclude with how this work forms a solid basis for the transfer of knowledge-based technology to European industry.

2. USE CASE COLLECTION FROM EUROPEAN INDUSTRY

Once the Industry Board had been set up with a group of initial members, the next task of collecting use cases from those members began. The use case collection is considered a key activity in the Industry Area as it is recognized that a major barrier between industry and research is that the former speaks in terms of problems and solutions and the latter in terms of technologies and research issues. A business use case is basically a story that relates a business problem to a solution and a solution to a technology, which in turn may lead to a research issue. Therefore we see business use cases as effective tools for facilitating understanding between an organisation and a research group.

In the first phase of Knowledge Web we have collected a set of use cases from the Industry Board members to act as a representative collection of industrial requirements for which ontology-based technologies are a potential solution. Each member was contacted with a request to provide illustrative examples of actual or hypothetical deployment of Semantic Web technologies in concrete business cases using a simple two page questionnaire. Provided use cases were followed up with face-to-face meetings between Knowledge Web partners and Industry Board members to gain additional information and following a consultative write-up process the use cases have been collected into a project deliverable.

Figure 1: Breakdown of use cases by industry sector

The deliverable contains 16 use cases. It can be seen in Figure 1 below that these use cases are spread across 9 industry sectors.

It is important to note that the use case collection is also an ongoing activity. The use cases already collected will be published online on the Knowledge Web Industry Area portal alongside the use cases that will continue to be collected in the subsequent phases of the project. As a result, Knowledge Web will be continuing the dialogue process between industry and academia which has the aim of successfully transferring ontology-based technologies from academia to industry.

3. ANALYSIS OF BUSINESS USE CASES

A simple analysis of the use cases according to the solutions that are being sought and the problems encountered forms an initial basis for identifying what concrete business problems semantic technologies are being applied to solve and what technological issues exist that are preventing semantic technologies from realizing their potential to solve these business problems.

We have extracted from the use case text the key areas where solutions are sought and the key technology locks to their realization, categorizing them into sets of related issues.

Figure 2: Solutions sought in industry use cases

Figure 2 above shows the areas in which industry is seeking ontology-based solutions. Nearly half of the collected use cases seek solutions to *data integration* and *semantic search*. Other areas mentioned in a quarter of use cases were *data management* and *personalization*. However the issues of search and integration are clearly very significant application domains in industry for ontology-based technologies.

Figure 3: Technology locks to industry use cases

Figure 3 above shows the technology locks identified in the use cases, with only those mentioned at least twice shown in the diagram for legibility. We can see that three technology locks stand out, occurring in 4 to 6 of the collected use cases. These are *knowledge extraction*, *ontology mapping* and *ontology authoring*. While the relatively even spread of locks indicates that issues to be resolved in using ontology-based technologies in an industrial setting are broad, we can understand this as indicating that particular needs in industry are the extraction of knowledge from existing data, the creation of ontologies - whether from scratch and through re-use of existing ones - and the mapping between them as a result of unavoidable heterogeneity in ontology usage. This latter issue ties in well in the solution areas of search and integration, both of which requiring ontological mapping to operate successfully.

This analysis is able to offer an informative overview of European industry perspective of ontology-based technologies. This is valuable in terms of seeking to present these technologies to industry as a whole. However to be able to answer specific industrial requirements we need to conduct further, case specific analyses.

4. KNOWLEDGE PROCESSING TASKS

In this section we discuss the knowledge processing tasks required by the current industry needs. We first introduce a methodology used for the identification of knowledge processing tasks, then we discuss how it can be applied in practice by an example of a business case. Finally, we report a (general) typology of knowledge processing tasks and a library of corresponding high level components which summarizes the requirements of all the use cases we have analyzed in D1.1.3.

4.1 A methodology

The standard specification methodology used for the identification of knowledge processing tasks and corresponding high level components is based on Rational Unified Process (RUP) [**?**, **?**]. Out of its six standard steps (i.e., business modeling, service requirements, analysis, design, implementation, and validation), D1.1.3 focuses only on three of them, namely, service requirements, analysis, and design. Let us discuss them in some detail.

Service Requirements: These are a set of services available through a system in order to implement a business case. They are determined through analysis of functional needs, which in turn imply some technical constraints (e.g., time response, scalability, number of connected customers) of a system to be developed. Service requirements are expressed in terms of UML (technical) use cases [**?**].

Analysis. This step performs initial subsystem partitioning related to main processing tasks and then analyses in detail the use cases. In particular, use cases are refined with the help of sequence diagrams [**?**], which incorporate the modules for the architecture proposal and the information flow between these modules to fulfill the use case functionality.

Design. This step refines and homogenizes classes, and identifies the architecture design. It is partially specified in D1.1.3. In the Knowledge Web context, the aim of this step is only to identify knowledge processing tasks and components along with pointing out the technology locks they are stipulated by.

We structure knowledge processing tasks as primary and secondary tasks according to their influence on the architecture of the system. Primary tasks are the common parts for most of actions or parts of actions of the system. Secondary tasks are additional requirements, i.e., extensions of the common parts. Based on the primary and secondary knowledge processing tasks we build a typology of knowledge processing tasks and corresponding high level components for each use case under consideration and for all of them together as a final (general) typology.

4.2 The methodology by an example

Let us discuss with the help of the *B2C Marketplace for Tourism* business case how the above introduced methodology is used in order to determine knowledge processing tasks.

4.2.1 Use Case Summary

The main two uses of the B2C marketplace system for tourism are summarized in Figure 4. A detailed description of this business case can be found in [**?**].

Figure 4: UML use case diagram for *B2C marketplace for tourism*

The first use case, which is called *to plan a nice week-end*, constitutes the entry point inside the marketplace allowing customers to define their personal needs. The platform takes care of identifying potentially useful contents and services, accessing multiple providers and selecting only the relevant ones. The second use case, which is called *to package and purchase a nice week-end*, requires (i) a dynamic aggregation of relevant contents and services (e.g., transport, accommodation, leisure activities), (ii) an automated packaging of week-end proposals, and (iii) facilities for purchasing them on-line.

Now, based on the functional requirements, it is possible to specify also different technical use cases taking part in the platform. Those use cases will next allow for a detailed analysis of the technical needs.

4.2.2 Service requirements

Technical use cases diagram is presented in Figure 5. Let us discuss its actors.

Figure 5: UML technical use case diagram for *B2C market place for tourism*

Customer and Access Interface. A customer with the help of its access interface (e.g., mobile phone) accesses services available within the system through the authentication mechanism, personalization, and session management.

Contents and Services providers (C/S Ps). Contents and services providers manage their offers autonomously, i.e., the system does not impose any constraints. Each contents and services provider has its own rules for structuring information at the protocol, syntactical, and semantic levels.

Administrator performs (i) referencing of new contents and services providers, and (ii) internal knowledge representation and management.

4.2.3 Analysis

At this step, we analyze each technical use case of Figure 5 in detail. In particular, we consider *contents and services access*, *contents aggregation*, *contents association*, *knowledge and services management*, *content and services provider's integration*, *heterogeneity of contents and services provider's management*, and *knowledge and services management* technical use cases.

For lack of space we discuss here only the *contents aggregation* technical use case. First, we report the actors it involves, then we provide its summary, inputs and outputs, and finally we analyze with the help of sequence diagrams the flow of its events and possible technology locks.

Actors: Customer and access services, C/S P Service.

Summary: The use case contents aggregation is inherited from the use case combine contents. It performs the fusion of the information issued by different C/S Ps. The use case aims at providing a user with the result which has the following characteristics:

- No duplication and redundant information;
- A homogenous information structure;
- Avoid the user having to aggregate the contents issued from different C/S Ps.

Notice that a domain ontology, which is a model for the data of all the C/S Ps, captures the knowledge of the domain.

Preconditions and inputs:

- The use case contents and services access has been executed;
- The contents are represented in terms of the domain ontology.

Post-conditions and outputs:

- The aggregated contents are transferred to the access interface.

The flow of events for the contents aggregation technical use case is presented in Figure 6.

Figure 6: Flow of events: Contents aggregation technical use case

Let us discuss it in detail. Before operating the contents aggregation, the system (ManageContentAggregation component) needs to map the data (potentially expressed in different data models) among the C/S Ps involved in the processing of the request of a user. This step is essential in order to evaluate the contents of each C/S P, and hence, detect redundancies, complementary information, etc. The flow of events is as follows:

- Identification of the mappings between different data models (requestSchemas);
- Contents aggregation (manageContent): check for duplicated information, fusion of complementary information are operated by the ControlContent component;
- Transformation of the result of contents aggregation into XML formalism;
- The results encoded in XML formalism are transferred to the access service (loadXmlStream).

Technology locks identification: Technology locks are highlighted in Figure 6. For lack of space we discuss here only the *mappings discovery* technology lock. It is crucial to be able to dynamically discover semantic mappings between the contents of different C/S Ps. The current solution follows the data integration approach which is to create static correspondences between data models [**?**]. However, this solution does not satisfy the requirements of the current World Wide Web. In fact, C/S Ps may appear and disappear over the network, change their contents, schemas, and so on. Thus, the question is how to determine those correspondences dynamically. For example, given two XML schemas, suppose in the first schema the address element consists of the name, town, and postcode attributes, in the second schema the address element is split down into three sub elements: street name, post code and town. Then, a solution should be developed in order to determine correspondences between the semantically related entities, e.g., the address element in the first schema should be mapped to the address element in the second schema. A more complex solution is required to determine that attributes of the first schema are to be mapped (notice!) to the elements of the second schema.

4.2.4 Design

Having identified technology locks of the B2C tourism marketplace system, now we are able to state knowledge processing tasks required in order to develop plausible Semantic Web solutions to those technology locks. In particular, our example requires the *matching* knowledge processing task, whose implementation provides the *match manager* component:

Matching. This task aims at determining similarity relations between the contents of C/S Ps and the global schema. This task is necessary to ensure semantic homogeneity, and hence, it will appear as a primary task in the typology.

Match manager. This module takes two data/conceptual models as input and returns a similarity relation between the entities of those models that correspond semantically to each other. A similarity relation can be either in the form of a coefficient in the [0,1] range, rating match quality (i.e., the higher the coefficient, the higher the similarity between the entities, see [**?**, **?**] for some particular implementations) or in the form of a semantic relation (e.g., equivalence, more general, less general), see [**?**, **?**, **?**].

4.3 A typology of knowledge processing tasks and a library of high level components

In the above described manner we determine knowledge processing tasks and components for all the use cases under consideration in D1.1.3.

4.3.1 The typology

A general typology of knowledge processing tasks which summarizes the requirements of the use cases considered in D1.1.3 is presented in Tables 1 and 2. It includes 9 primary tasks and 4 secondary tasks. It is also worth noticing that some tasks are to be implemented within a single component. For example, such tasks as schema/ontology matching, ranking matching results, and producing explanations of mappings are the functionalities of the match manager component. Thus, the library of high level components contains less components than the number of knowledge processing tasks identified. In particular, it consists of 10 components.

Table 1: Typology of knowledge processing tasks & components. Part 1 - Primary tasks

#	Knowledge processing tasks	Components
1	Data Translation	Wrapper
2	Ontology Management	Ontology Manager
3	Matching	Match Manager
4	Matching Results Analysis	Match Manager
5	Content Annotation	Annotation Manager
6	Reasoning	Reasoner
7	Semantic Query Processing	Query Processor
8	Composition of Web Services	Planner
9	Results Reconciliation	Results Reconciler

4.3.2 Task and component descriptions

Below we provide short high level descriptions of knowledge processing tasks and components of Tables 1 and 2, while their details can be found in [**?**].

Table 2: Typology of knowledge processing tasks & components. Part 2 - Secondary tasks

#	Knowledge processing tasks	Components
1	Schema/Ontology Merging	Ontology Manager
2	Producing Explanations	Match Manager
3	Personalization	Profiler
4	Directory Management	Directory Manager

Data Translation and Wrapper. This task is in charge of translating/exchanging instances between heterogeneous information sources storing their data in different formats (e.g., RDF, SQL DDL).

Ontology Management, Schema/Ontology Matching, Merging and Ontology Manager. These tasks are in charge of ontology maintenance with respect to (evolving) business case requirements.

Matching, Matching Results Analysis, Producing Explanations and Match Manager. These tasks are in charge of determining mappings between the entities of multiple schemas, classifications, and ontologies. The mappings might be ordered according to some criteria. In addition, explanations of the mappings might be also produced.

Content Annotation and Annotation Manager. This task is in charge of automatic production of metadata for the contents.

Reasoning and Reasoner. This task is in charge of providing standard logical reasoning services (e.g., subsumption, concept satisfiability, instance checking tests).

Semantic Query Processing and Query Processor. This task is in charge of rewriting a query by using terms which are explicitly specified in the knowledge of the domain.

Composition of Web Services and Planner. This task and component is in charge of automated composition of web services into executable processes.

Results Reconciliation and Results Reconciler. This task is in charge of determining an optimal solution, in terms of contents (no information duplication, etc.), for returning results from the queried information sources.

Personalization and Profiler. This task is in charge of tailoring services available from the system to the specificity of each user (e.g., standard vs. professional profiles).

Directory Management and Directory Manager. This task is in charge of maintenance and interpretation of instance data distributed by contents and services providers.

4.3.3 Use of the typology

The case specific analysis of the business use cases allows us to move from the industrial perspective of the use case descriptions (looking at problems and solutions) to a more technical analysis from which we can derive key research challenges. We have modelled the use cases as business processes and identified where in the process we encounter technology locks. We have also linked those technology locks to knowledge processing tasks and components which should overcome those locks. A successful transfer of technology to industry can take place when components exist which, when implemented in the business process, remove the technology lock and solve the business problem. Where the state of the art of knowledge processing components does not meet industrial requirements, we can identify what those requirements are, focus researchers on developing components that meet them, and can point to concrete scenarios where the developed components could be tested and evaluated in an industry-strength setting.

Thus, for example in a Human Resources scenario, the sought-for solution is the *semantic* matching between job offers and job applications. By a technical use case analysis we located where in the business process the lock occurs and defined the requirements with respect to the *matching* task and the *match manager* component. Hence, we have already provided (i) a client industry with a clear identification of the place where the system requires knowledge-based solutions and (ii) researchers with a clear definition of the requirements that must be met by their prototypical implementations of knowledge processing components. In particular, in this scenario, some existing implementations of a match manager (e.g.,[**?**, **?**]) have been plugged into the business process at the identified location. A prototype has been tested by the client industrial partner, and it had demonstrated a better characteristics (e.g., precision, recall) with respect to the legacy solution. Thus, experience of this use case and some others (e.g., the MCAA scenario) gives us a preliminary vision that the proposed approach is able to facilitate the transfer of Semantic Web technology to industry. This will also prove to be a great benefit to European industry in that it will solve ídentified business problems.

We acknowledge too that with the emergence of new business cases it is likely that new knowledge processing tasks will appear. For example, web service discovery, orchestration, and so on. Therefore, the typology we have described here constitutes only a first step. Future work includes ongoing technical analysis of collected use cases till the saturation is reached.

5. CONCLUSIONS AND FUTURE WORK

In the first phase of Knowledge Web, the 'Outreach to Industry' activity has formed an Industry Board and collected business use cases from them. We have used this use case collection to make a general analysis of European industrial needs and the technology locks that exist in meeting those needs. Furthermore we have begun to carry out use case specific analyses and determined a typology of knowledge processing tasks and components that can meet industrial requirements.

As a Network of Excellence, Knowledge Web is made up of leading Semantic Web research institutions across Europe. This gives us the opportunity to feed the discovered industrial requirements into the European Semantic Web research effort as facilitated and coordinated by Knowledge Web.

Hence, the next document from the 'Outreach to Industry' activity, due June 2005, will be a collection of executive summaries of key use cases from our collection. These summaries will present the business problem described by each use case in terms of the knowledge processing tasks and components that are relevant to each technology lock and the requirements that implementations of those components need to meet. Ongoing work in Knowledge Web will be to co-ordinate research work in developing solutions to these challenges, evaluating them in the industrial context and promoting success stories to industry as a whole.

In particular, by concentrating initially on a number of key use cases that exemplify the application of Semantic Web technologies in the more mature application fields, we can realistically achieve within the next year of the project evidence of the benefits of knowledge-based technological solutions for industry and have an ideal summary of the application of this technology for further publication and dissemination.

6. ACKNOWLEDGEMENTS

The work described in this paper is supported by the EU Network of Excellence KnowledgeWeb (FP6-507482).