

Semantic Web applications: Fields and Business cases. The Industry challenges the research.

Alain Léger¹, Lyndon J.B. Nixon², Pavel Shvaiko³, Jean Charlet⁴,

¹ France Telecom R&D - Rennes, 4 rue du clos courtel, Boîte postale 91226,
35512 Cesson-Sévigné, France
alain.leger@rd.francetelecom.com

² Freie Universität Berlin, Takustrasse, 9
14195 Berlin, Germany

nixon@mi.fu-berlin.de

³ University of Trento (UniTn)

Via Sommarive 14

38050 Trento, Italy

pavel@dit.unitn.it

⁴ Jean Charlet, STIM, DPA/AP-Hopitaux Paris & Université Paris 6,
75006 Paris, France

charlet@biomath.jussieu.fr

Abstract. Semantic web technology is more and more often applied to a large spectrum of applications where domain knowledge is conceptualized and formalized (Ontology) as a support for diversified processing (Reasoning) operated by machines. Moreover through a subtle joining of human reasoning (cognitive) and mechanical reasoning (logic-based), it is possible for humans and machines to share complementary tasks. To name few of those applications areas: Corporate Portals and Knowledge Management, E-Commerce, E-Work, Healthcare, E-Government, Natural Language understanding and Automated Translation, Information search, Data and Services Integration, Social networks and collaborative filtering, Knowledge Mining, etc. From a social and economic perspective, this emerging technology should contribute to growth in economic wealth, but it must also show clear cut value in our everyday activities in being technology transparent and efficient. The uptake of Semantic Web technology by industry is progressing slowly. One of the problems is that academia is not always aware of the concrete problems that arise in industry. Conversely, industry is not often well informed about the academic developments that can potentially meet its needs. In this paper we present an ongoing work in the cross-fertilization between industry and academia. In particular, we present a collection of applications fields and use cases from enterprises which are interested in the promises of Semantic Web technology. We explain our approach in the analysis of the industry needs. We summarize industrial knowledge processing requirements in the form of a typology of knowledge processing tasks. These results are intended to focus academia on the development of plausible knowledge-based solutions for concrete industrial problems, and therefore, facilitate the uptake of Semantic Web technology within industry.

1 Introduction

Through the invading, pervasive and user-friendly digital technology within the information society, the fully open web content emerges as multiform, inconsistent and very dynamic. This situation leads to abstracting (via Ontology) this complexity and to offer new and enriched services able to reason on those abstractions (Reasoning) via automata – e.g. Web services. This abstraction layer is the subject of a very dynamic activity in research, industry and standardization in what is usually called worldwide "Semantic Web" [e.g. DARPA, European IST Research Framework Program, W3C]. The very first application of the semantic web technology has focused on Information Retrieval (IR) where access by to semantic content instead of the classical (even sophisticated) statistical analysis was sought to give far better results (Precision and Recall). The next natural extension was on IR applied to enterprise legacy databases integration for leveraging the company information silos. The present large field of applications is now focusing on the seamless integration of applications or services through a full usage of semantic web services for expected fast ROI and efficiency in E-Work and E-Business.

This new technology takes its roots in the cognitive sciences, machine learning, natural language processing, multi-agents systems, knowledge acquisition, mechanical reasoning, logics and decision theory. It can be separated in two distinct – but cooperating fields - one adopting a formal and algorithmic approach for common sense automated reasoning (automated Web) and the other one "keeping the human being in the loop" for socio-cognitive semantic web (automated social Web).

On a large scale, industry awareness of the knowledge-based technology has started only recently, e.g., at the EC level with the IST-FP5 thematic network Ontoweb¹ which had brought together around 50 motivated companies worldwide.

Based on this experience, within the IST-FP6 network of excellence Knowledge-Web², an in-depth analysis of the concrete industry needs in the key economic sectors has been identified as one of the next steps towards stimulating the industrial uptake of Semantic Web technology.

The paper is organized as follows. Three prototypical application fields are presented in Section 2: KM, E-Commerce and Healthcare. Use cases collection methodology from industry and their preliminary analysis leading to the identification of key knowledge processing components are presented in Section 3. Finally, Section 4 reports some conclusions and discusses future effort.

¹ <http://www.ontoweb.org>

² <http://knowledgeweb.semanticweb.org>

2 Some prototypical application fields

2.1 Knowledge Management

Nowadays, knowledge is one of the key success factors for today and tomorrow's enterprises. Therefore, company Knowledge Management (KM) has been identified as a strategic tool for enterprises. However, if Information Technology is one foundation element of KM, KM is also interdisciplinary by nature, and includes human resource management, enterprise organization and culture³.

So KM is the management of the activities and the process aiming at leveraging the use and the creation of knowledge in organizations for two main objectives: capitalization of the corporate knowledge and durable innovation, and fully aligned with the strategic objectives of the organization:

1. Access, sharing, reuse of knowledge (explicit or implicit, private or collective) ;
2. Creation of new knowledge.

A recent CEN/ISSS⁴ project (KM Workshop 2002-2003) has issued a finalized proposal on good practices in KM (September 2003). The project began in October 2002 on KnowledgeBoard⁵, which is the European Commission public KM portal, and is supposed to close with a final set of CEN recommendations in fall 2003 entitled "European guide to Good Practice in Knowledge Management".

The European KM Framework is designed to support a common European understanding of KM, to show the value of this emerging approach and help organizations towards its successful implementation. The Framework is based on empirical research and practical experience in this field from all over Europe and the rest of the world. The European KM Framework addresses all relevant elements of a KM solution and serves as a reference basis for all types of organizations, which aim to improve their performance by handling knowledge in a better way.

³ Some definitions:

" Knowledge management is the systematic, explicit, and deliberate building, renewal and application of knowledge to maximize an enterprise's knowledge related effectiveness and returns from its knowledge assets" (Wiig 1997) [1]

"Knowledge management is the process of capturing a company's collective expertise wherever it resides in databases, on paper, or in people's heads and distributing it to wherever it can help produce the biggest payoff" (Hibbard 1997) [2]

"KM is getting the right knowledge to the right people at the right time so they can make the best decision" (Pettrash 1996) [3]

⁴ <http://www.cenorm.be/cenorm/index.htm>

⁵ <http://www.knowledgeboard.com>

2.1.1 Where should Knowledge-based KM benefit?

In the past, Information Technology for knowledge management has focused on the management of knowledge containers using text documents as the main repository and source of knowledge. In the future, Semantic Web technology, especially ontologies and machine-interpretable metadata will pave the way to KM solutions that are based on semantically related pieces of knowledge. The knowledge backbone is made of ontologies that define a shared conceptualization of the application domain at hand and provide the basis for defining metadata, that have precisely defined semantics, and that are therefore machine-interpretable. Although, the first KM approaches and solutions have shown the benefits of ontologies and related methods, a large number of open research issues still exist that have to be addressed in order to make Semantic Web technologies a complete success for KM solutions:

- Industrial KM applications have to avoid any kind of overheads as far as possible. Therefore, a **seamless integration** of knowledge creation, e.g. content and metadata specification, and knowledge access, e.g. querying or browsing, into the working environment is required. Strategies and methods are needed to support the creation of knowledge, as side effects of activities that are carried out anyway. These requirements mean **emergent semantics**, e.g. through **ontology learning**, are needed, which reduces the current time consuming task of building-up and maintaining ontologies.
- Access as well as presentation of knowledge has to be **context-dependent**. Since the context is set-up by the current business task, and thus, by the business process being handled, a tight integration of business process management and knowledge management is required. KM approaches can manage knowledge and provide a promising starting point for **smart push services** that will proactively **deliver relevant knowledge** for carrying out the task at hand more effectively.
- **Conceptualization** has to be supplemented by **personalization**. On one hand, taking into account the experience of the user and his/her personal needs is a prerequisite in order to avoid information overload, and on the other hand **to deliver knowledge on the right level of granularity**.

The development of knowledge portals serving the needs of companies or communities is still more or less a manual process. Ontologies and related metadata provide a promising conceptual basis for generating parts of such knowledge portals. Obviously, among others, conceptual models of the domain, of the users and of the tasks are needed. The **Generation of knowledge portals** has to be supplemented with the (semi-) automated evolution of portals. As business environments and strategies change rather rapidly, **KM portals have to be kept up-to-date in this fast changing environment**. Evolution of portals should also include some mechanism **to ‘forget’ outdated knowledge**.

KM solutions will be based on a combination of intranet-based functionalities and mobile functionalities in the very near future. Semantic Web technologies are a promising approach to meet the needs of mobile environments, like e.g. location-aware personalization and adaptation of the presentation to the specific needs of mobile devices, i.e. the presentation of the required information at an appropriate level of granularity. In essence, employees should have access to the KM application **anywhere and anytime**.

Peer-to-Peer computing (P2P), combined with Semantic Web technology, will be an interesting of getting rid of the more centralized KM solutions that are currently used in ontology-based solutions. P2P scenarios open up the way to derive consensual conceptualizations among employees within an enterprise in a bottom-up manner.

Virtual organizations are becoming more and more important in business scenarios, mainly due to decentralization and globalization. Obviously, semantic interoperability between different knowledge sources, as well as trust, is necessary in inter-organizational KM applications.

The integration of KM applications (e.g. skill management) with **E-Learning** is an important field that enables a lot of synergy between these two areas. KM solutions and E-Learning must be integrated from both an organizational and an IT point of view. Clearly, interoperability and integration of (metadata) standards are needed to realize such integration.

Knowledge Management is obviously a very promising area for exploiting Semantic Web technology. Document-based KM solutions have already reached their limits, whereas semantic technologies open the way to meet the KM requirements in the future.

2.1.2 Knowledge-based KM applications⁶

In the context of geographical team dispersion, multilingualism and Business Units autonomy, usually the company wants a solution allowing the identification of strategic information, the secured distribution of this information and the creation of transverse working groups. Some applicative solutions allowed the deployment of an Intranet intended for all the marketing departments of the company worldwide, allowing a better division and a greater accessibility to information, but also capitalisation

⁶ <http://www.arisem.com>
<http://www.mondeca.com>
<http://www.ontoknowledge.com>
<http://www.distributedthinking.com>
<http://www.ontoknowledge.com>
<http://www.si.fr.atosorigin.com/sophia/comma/Htm/HomePage.htm>

on the total knowledge of the company group. There are three crucial points that aim to ease the work of the various marketing teams of the company group: automatic competitive intelligence of the Web, skill management and document management.

Thus, the system connects the "strategic ontologies" of the company group (brands, competitors, geographical areas, etc...) with the users, via the automation of related processes (research, classification, distribution, representation of knowledge). The result is a dynamic "Semantic Web" system of navigation (research, classification) and collaborative features.

From a functional point of view, KM server organises skill and knowledge management within the company, in order to improve interactivity, collaboration and information sharing. This constitutes a virtual workspace which facilitates work between employees that speak different languages; automates the creation of work groups; organises and capitalises structured and unstructured, explicit or tacit data of the company organisation, and offers advanced features of capitalisation. Furthermore, the semantic backbone also makes possible to cross a qualitative gap by providing cross-lingual data. Indeed, the semantic approach allows ontologies to overcome language barriers (Culture and Language differences).

Some lessons learnt⁷:

- **Main strong benefits** for the enterprise are high productivity gains and operational valorisation of knowledge legacy
- **Productivity:** Automation of knowledge base maintenance, Automation of content indexing, Augmented productivity in publication cycle (commercial proposals, reports ...), Search efficiency (a reduction factor on research time of the order (1000 to 1) is claimed possible by the use of ontologies)
- **Quality and operational valorisation of knowledge legacy:** Unified management of heterogeneous resources, Information relevancy, Capacity to represent complex knowledge, Gains in development and maintenance of knowledge and content management solution, Generic and evolvable solution
- **Human factors** are key difficulties in full groupware functionalities of the KM solution towards the employees of the company, so adopt a step-by-step approach
- **Access to information portal** must be well designed and must be supported by a group of people dedicated to information filtering and qualifying (P2P is possible)

⁷ Le Monde Informatique 11 July 2003 and <http://www.mondeca.com>

2.2 E-Commerce

Electronic Commerce is mainly based on the exchange of information between involved stakeholders using a telecommunication infrastructure. There are two main scenarios: Business-to-Customer (B2C) and Business-to-Business (B2B).

B2C applications enable service providers to promote their offers, and for customers to find offers, which match their demands. By providing a single access to a large collection of frequently updated offers and customers, an electronic marketplace can match the demand and supply processes within a commercial mediation environment.

B2B applications have a long history of using electronic messaging to exchange information related to services previously agreed among two or more businesses. Early plain-text telex communication systems were followed by electronic data interchange (EDI) systems based on terse, highly codified, well structured, messages. A new generation of B2B systems is being developed under the ebXML (electronic business in XML) label. These will use classification schemes to identify the context in which messages have been, or should be, exchanged. They will also introduce new techniques for the formal recording of business processes, and for the linking of business processes through the exchange of well-structured business messages. ebXML will also develop techniques that will allow businesses to identify new suppliers through the use of registries that allow users to identify which services a supplier can offer. ebXML needs to include well managed multilingual ontologies that can be used to help users to match needs expressed in their own language with those expressed in the service providers language(s).

2.2.1 Where is the value of Knowledge-based E-Commerce?

At the present time, ontology and more generally ontology-based systems, appear as a **central issue** for the development of **efficient and profitable** Internet commerce solutions. However, because of an actual lack of standardization for business models, processes, and knowledge architectures, it is currently difficult for companies to achieve the promised ROI from Knowledge-based E-Commerce.

Moreover, a technical barrier exists that delays the emergence of E-Commerce, laying in the need for applications to **meaningfully share information**, taking into account the lack of reliability and security of the Internet. This fact may be explained by the variety of enterprise and e-commerce systems employed by businesses and the various ways these systems are configured and used. As an important remark, such **interoperability problems** become particularly acute when a large number of trading partners attempt to agree and define the standards for interoperation, which is precisely a main condition for maximizing the ROI.

Although it is useful to strive for the adoption of a single common domain-specific standard for content and transactions, such a task is often still difficult to achieve,

particularly in cross-industry initiatives, where companies co-operate and compete with one another. Some examples of the difficulties are:

- **Commercial practices** may vary in a wide range and consequently, cannot always be aligned for a variety of technical, practical, organizational and political reasons.
- **The complexity of the global description** of the organizations themselves: their products and services (independently or in combination), and the **interactions** between them remain a formidable task.
- It is usually very difficult to establish, a priori rules (technical or procedural) governing participation in an electronic marketplace.
- Adoption of a **single common standard may limit business models**, which could be adopted by trading partners, and then, potentially reduce their ability to fully participate in Internet commerce.

An ontology based approach has the potential to significantly accelerate the penetration of electronic commerce within vertical industry sectors, by **enabling interoperability at the business level**, reducing the need for standardisation at the technical level. This will enable services to adapt to the rapidly changing online environment.

The following uses for ontologies and classification schemes that could be defined using ontologies, have been noted within electronic commerce applications:

- Categorization of products within catalogues
- Categorization of services (including web services)
- Production of yellow page classifications of companies providing services
- Identification of countries, regions and currencies
- Identification of organizations, persons and legal entities
- Identification of unique products and saleable packages of products
- Identification of transport containers, their type, location, routes and contents
- Classification of industrial output statistics.

2.2.2 Knowledge-based E-Commerce applications

According to (Zyl et al.) [4], applications of this kind use one or more shared ontology to integrate heterogeneous information systems and allow common access for humans or computers. This enforces the shared ontology as the standard ontology for all participating systems, which removes the heterogeneity from the information system. The heterogeneity is a problem because the systems to be integrated are already operational and it is too costly to redevelop them. A linguistic ontology is sometimes used to assist in the generation of the shared ontology, or is used as a top-level ontology, describing very general concepts like space, time, matter, object, event, action, etc, for the shared ontologies to inherit from it. Benefits are the integra-

tion of heterogeneous information sources, which can improve interoperability, and more effective use and reuse of knowledge resources⁸.

Yellow Pages and products catalogue are direct benefactors of a well structured representation which coupled to multilingual ontology enhances clearly the precision / recall of products or services search engine. The ONTOSEEK system (1996-1998) is the first system being prototyped associating domain ontology (in KR conceptual graph CG with very limited expressiveness) to a large multilingual linguistic ontology (SENSUS – WORDNET) for natural language search of products (Guarino et al., 1998) [5]. ONTOSEEK search products by mapping natural human language human requests to domain ontology. Unlike traditional eCommerce portal search functions the user is not supposed to know the vocabulary used for describing the products and thanks to the SENSUS ontology he is able to express himself in its own vocabulary.

The main functional architectural choice of ONTOSEEK:

- Use of a general linguistic ontology to describe products;
- Great flexibility in expressing the request thanks to the semantic mapping offered between the request and the offers;
- Interactive guided request formulation through generalisation and specialisation links

A Conceptual Graph KR is used internally to represent Request and Products. The semantic matching algorithm is based on a simple subsumption on the ontology graph and does not make use of a complex graph endomorphism.

ONTOSEEK has not been deployed commercially but at its trial period has fully demonstrated the potential benefits making use of preliminary semantic web tools.

The MKBEEM [6] prototype and technology (Multilingual Knowledge Based European Electronic Marketplace - IST-1999-10589, 2000 – 2003) concentrate on written language technologies and its use in the key sector of worldwide commerce. Within the global and multilingual Internet trading environment, there is an increasing pressure on e-content publishers of all types to adapt content for international markets. Localization – translation and cultural adaptation for local markets – is proving to be a key driver of the expansion of business on the web. In particular MKBEEM is focusing on adding multilingualism to all stages of the information cycle, including multilingual content generation and maintenance, automated translation and interpretation, and enhancing the natural interactivity and usability of the service with unconstrained language input. On the Knowledge technology side, the MKBEEM Ontologies provide a consensual representation of the electronic commerce field in two typical Domains (B2C Tourism, B2C Mail order) allowing the commercial exchanges to be transparent in the language of the end user, of the service, or of the product provider. Ontologies are used for classifying and indexing

⁸ <http://www.chemdex.com>
<http://kmi.open.ac.uk/projects/alice/>
<http://www.telecom.ntua.gr/smartec/>
<http://www.mkbeem.com/>

catalogues, for filtering user's query, for selecting relevant products and providers, for facilitating multilingual man-machine dialogues, and for inferring information that is relevant to the user's request and eventually trading needs. The Key Innovative approach is based on a combined use of human language processing and ontologies based reasoning, for:

The effectiveness of the developed generic solutions has been tested in Finnish, French, Spanish and English in the domains of travel booking (SNCF French Rail services) and mail order sales (La Redoute - Ellos).

2.3 Biosciences and Medical applications

The Medical domain is a favourite target for semantic web applications just as the expert system was for Artificial Intelligence applications 20 years ago. The medical domain is effectively very complex: medical knowledge being difficult to represent in a computer, which makes the sharing of information difficult. Semantic web solutions become very promising in this context.

Thus one of the main mechanisms of the semantic web, resource description using annotation principles, is of major importance in the medical informatics (or "bioinformatics") domain, especially as regards the sharing of these resources (e.g. medical knowledge in the Web or genomic data-base). Through the years, the information retrieval domain has been developed by medicine: the medical thesaurus is enormous (1,000,000 terms for UMLS) and is principally used for bibliographic indexation. Nevertheless, the MeSh thesaurus (Medical Subject Heading) or UMLS⁹ (Unified Medical Language System) is used in the web semantic paradigm with varying degrees of difficulty. Finally, the web services technology allows us to imagine some solutions to the interoperability problematic, which is substantial in medical informatics. We will describe current research, results and expected perspectives in these biomedical informatics topics in the context of the semantic web.

2.3.1 Biosciences resources sharing

In the functional genomics domain, it is necessary to have access to several data bases and knowledge bases which are accessible via the web but are heterogeneous in their structure as well as in their terminology. Among such resources, we can cite SWISSPROT¹⁰ where the gene products are annotated by Gene Ontology¹¹, GenBank¹², etc. In comparing the resources, it is easy to see that they propose the same information in different formats. The XML language, described as the unique common language of these bases proposes as much Document Type Definition (DTD) as resources and does not resolve the interoperability problem.

⁹ <http://www.nlm.nih.gov/research/umls/umlsmain.html>

¹⁰ <http://us.expasy.org/prot/>

¹¹ <http://obo.sourceforge.net/main.html>

¹² <http://www.ncbi.nlm.nih.gov/Genbank/index.html>

The solution comes from the semantic web with the mediator approach (Wiederhold, 1992) [7] which allows the accessing of different resources with an ontology used as Interlingua pivot. For example, and in another domain than that of genomics, the mediator mechanisms, the NEUROBASE project (Barillot et al., 2003) [8] attempts to federate different neuroimagery information bases situated in different clinical or research areas. The proposal consists of defining an IT architecture that allows the access to and the sharing of experimental results or data treatment methodologies. It would be possible to search in the various data bases for similar results or for images with peculiarities or to perform data mining analysis between several data bases. The mediator of NEUROBASE is tested on decision support systems in epilepsy surgery.

2.3.2 Web services for interoperability

The web services technology can propose some solutions to the interoperability problematic. We describe now a new approach based on “patient envelope” and we conclude on the implementation of this envelope with the web services technologies.

The patient envelope is a proposition of the Electronic Data Interchange for Healthcare group (EDI-Santé¹³) with an active contribution from the ETIAM society¹⁴. The objective of the work has been to focus on filling the gap between “free” communication, using standard and generic Internet tools, and “totally structured” communication as promoted by CEN¹⁵ or HL7¹⁶. After a worldwide analysis of existing standards, the proposal consists of an “intermediate” structure of information, related to one patient, and storing the minimum amount of data (i.e. exclusively useful data) to facilitate the interoperability between communicating peers. The “free” or the “structured” information is grouped into a folder and transmitted in a secure way over the existing communication networks (Cordonnier et al., 2003) [9]. This proposal has reached widespread celebrity with the distribution by Cegetel.rss of a new medical messaging service, called “Sentinelle”, fully supporting the patient envelope protocol and adapted tools.

After this milestone, EDI-Santé is promoting further developments based on ebXML and SOAP (Simple Object Access Protocol) in specifying exchange (1,2) and medical (3, 4) properties:

1. Separate what is mandatory to the transport and the good management of the message (patient identification, ...) from what constitute the “job” part of the message
2. Provide a “container”, collecting the different elements, texts, pictures, videos, etc.
3. The patient as unique object of the transaction. Such an exchange cannot be anonymous. It concerns a sender and an addressee who are involved in the ex-

¹³ <http://www.edisante.org/>

¹⁴ <http://www.etiam.com/>

¹⁵ <http://www.centc251.org/>

¹⁶ <http://www.hl7.org/>

change and responsible. The only way to perform this exchange between practitioner about a patient who can demand to know the content of the exchange imply to retain a structure which is unique, a triplet {sender, addressee, patient}.

4. The conservation of the exchange semantics. The information about a patient is multiple. It comes from multiple sources and has multiple forms and supports (data base, free textual document, semi-structured textual document, pictures ...). It can be fundamental to maintain the existing links between elements, to transmit them together, e.g. a scanner and the associated report, and to prove it.

The interest of such an approach is that it prepares the evolution of the transmitted document, from free document (from proprietary ones to normalize as XML) to elements respecting HL7v3 or EHRCOM data types.

2.3.3 And next?

These different projects and applications highlight the main consequence of the semantic web, expected by the medical communities, the sharing and integration of heterogeneous information or knowledge. The answers to the different issues are the mediators, the knowledge-based system, and the ontologies, all based on normalized languages as RDF, OWL or others. The work of the semantic web community must take into account these expectations - see FP6 projects¹⁷⁻¹⁸⁻¹⁹. Finally, it is interesting to note that the semantic web is an integrated vision of the medical community's problems (thesaurus, ontology, indexation, inference) and provides a real opportunity to synthesize and reactivate some research (Charlet et al., 2002) [10].

3 Use Case collection and Analysis

We have formed a group of companies interested in Semantic Web technology. By the end of 2004, this group consisted of 34 members (e.g., France Telecom, IFP, Illy Caffe, Trenitalia, Daimler Chrysler ...) from across 12 economic sectors (e.g., telecoms, energy, food, logistics, automotive).

The companies were requested to provide illustrative examples of actual or hypothetical deployment of Semantic Web technology in their business settings. This was followed up with face-to-face meetings between researchers and industry experts from the companies to gain additional information about the provided use cases. Thus, in 2004, we collected a total of 16 use cases from 12 companies.

¹⁷ <http://www.cocoon-health.com>

¹⁸ <http://www.srdc.metu.edu.tr/webpage/projects/artemis/index.html>

¹⁹ <http://www.simdat.org>

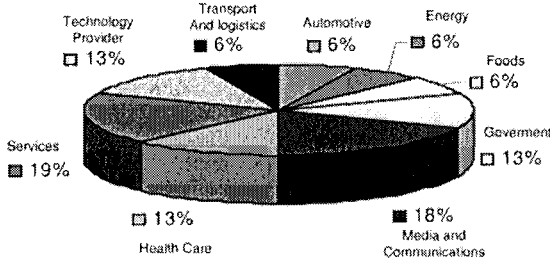


Figure 1: Breakdown of use cases by industry sectors

In particular, it represents (the most active) 9 sectors, with the highest number of the use cases coming from the service industry (19%) and media & communications (18%) respectively. The entire collection of use cases can be found in [11], or on the Outreach to Industry portal²⁰.

3.1 Preliminary Analysis of Use Cases

A preliminary analysis of the use cases has been carried out in order to obtain a first vision of the current industrial needs and to estimate the expectations from knowledge based technology with respect to those needs. The industry experts were asked to indicate the existing legacy solutions in their use cases, technological locks they encountered, and how they expected that Semantic Web technology could resolve those locks. As a result, we have gained an overview of:

- Types of business problems where the knowledge-based technology is considered to bring a plausible solution;
- Types of technological issues (and corresponding research challenges) which knowledge based technology is expected to overcome.

Let us discuss some concrete types of business problems/technological issues we have identified with the help of experts (see Figure 2 and Figure 3 for a summary).

²⁰ <http://knowledgeweb.semanticweb.org/o2i/>

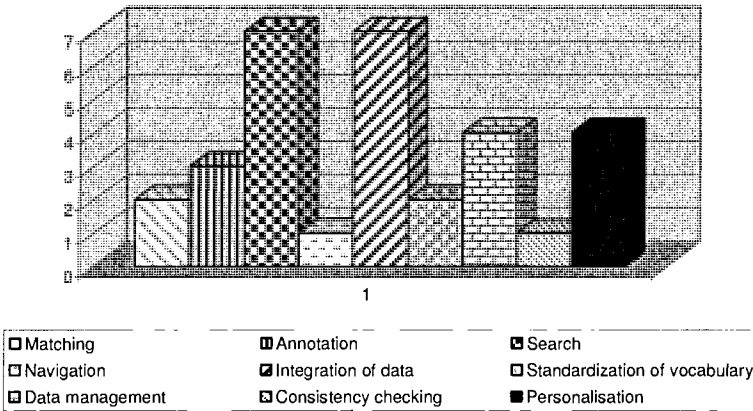


Figure 2: Preliminary vision for solutions sought in use cases

Figure 2 shows a breakdown of the areas in which the industry experts thought Semantic Web technology could provide a solution. For example, for nearly half of the collected use cases data integration and semantic search were areas where industry was looking for knowledge-based solutions. Other areas mentioned, in a quarter of use cases, were solutions to data management and personalization.

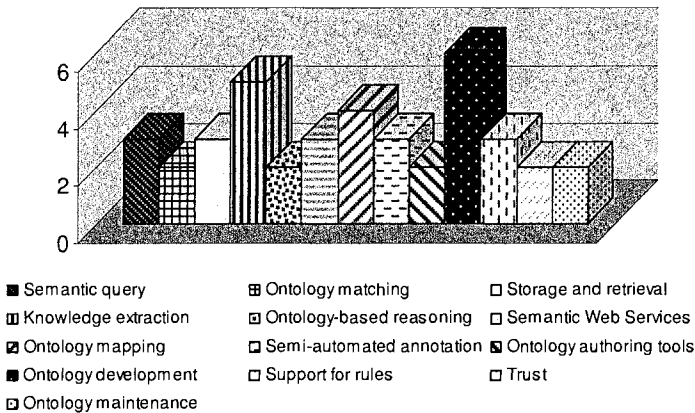


Figure 3: Preliminary vision of technology locks in use cases

Figure 3 shows a breakdown of the technology locks identified in the use cases. There are three technology locks which occur the most often in the collected use cases (namely, from 4 up to 6 use cases). These are: *ontology development*, i.e., modeling of a business domain, authoring, reusing existing ontologies; *knowledge extraction*, i.e., populating ontologies by extracting data from legacy systems; and *ontology matching*, i.e., resolving semantic heterogeneity among multiple ontologies.

Below, we illustrate, with the help of another use case from our collection, how a concrete business problem can also be used to indicate the technology locks for which knowledge-based solutions potentially might be useful. This use case addresses the problem of an intelligent search of documents in a corporate data of a coffee company.

The company generates large amount of internal data and its employees encounter difficulties in finding the data they need for the research and development of new solutions. The aim is to improve the quality of the documents retrieval and to enable the personalization services of individual users when searching or viewing the corporate data. As technology locks, the expert mentioned here the corporate domain *ontology development* and *maintenance*, and *semantic querying*.

The above three examples illustrate some concrete business scenarios in which an "abstract" research issues such as matching, data integration, etc., are viewed to be of great value to industry. This analysis (by experts estimations) provides us with a preliminary understanding of scope of the current industrial needs and concrete technology locks where knowledge-based technology is expected to provide a plausible solution. However, to be able to answer specific industrial requirements, we need to conduct further a detailed technical analysis of the use cases, thereby associating to each technology lock a concrete knowledge processing task and a component realizing its functionalities.

3.2 Knowledge processing tasks and components

Based on the knowledge processing needs identified during the technical use cases analysis [12], we built a typology of knowledge processing tasks and a library of high level components for realizing those tasks, see Table 1.

N°	Knowledge processing tasks	Components
1	Ontology Management	Ontology Manager
2	Matching	Match Manager
3	Matching results Analysis	Match Manager
4	Data Translation	Wrapper
5	Results Reconciliation	Results Reconciler
6	Composition of Web Services	Planner
7	Content Annotation	Annotation manager
8	Reasoning	Reasoner
9	Semantic Query Processing	Query Processor
10	Schema/Ontology Merging	Ontology Manager
11	Producing explanations	Match Manager
12	Personalization	Profiler

Table1. Typology of knowledge processing tasks & components

Our first tentative typology includes 12 knowledge processing tasks. Let us discuss knowledge processing tasks and components of Table 1 in more detail.

Ontology Management, Schema/Ontology Merging and Ontology Manager. These tasks and component are in charge of ontology maintenance (e.g., reorganizing taxonomies, resolving name conflicts, browsing ontologies, editing concepts) and merging multiple ontologies (e.g., by taking the union of the axioms) with respect to evolving business case requirements, see [13, 14, 15].

Matching, Matching Results Analysis, Producing Explanations and Match Manager. These tasks and component are in charge of (on-the-fly and semi-automatic) determining semantic mappings between the entities of multiple schemas, classifications, and ontologies, see [16, 17]. Mappings are typically specified with the help of a similarity relation which can be either in the form of a coefficient rating match quality in the [0,1] range (i.e., the higher the coefficient, the higher the similarity between the entities, see [18,19,20,21,22]) or in the form of a logical relation (e.g., equivalence, subsumption), see [23, 24]. The mappings might need to be ordered according to some criteria, see [25, 21].

Finally, explanations of the mappings might be also required, see [26, 27]. Matching systems may produce mappings that may not be intuitively obvious to human users. In order for users to trust the mappings (and thus use them), they need information about them. They need access to the sources that were used to determine semantic correspondences between terms and potentially they need to understand how deductions/ manipulations are performed. The issue here is to present explanations in a simple and clear way to the user.

Data Translation and Wrapper. This task and component is in charge of automatic manipulation (e.g., translation, exchange) of instances between heterogeneous information sources storing their data in different formats (e.g., RDF, SQL DDL, XML), see [28, 29]. Here, mappings are taken as input (for example, from the match manager component) and are analyzed in order to generate query expressions that perform the required manipulations with data instances.

Results Reconciliation and Results Reconciler. This task and component is in charge of determining an optimal solution, in terms of contents (no information duplication, etc.) and routing performance, for returning results from the queried information sources, see [30].

Composition of Web Services and Planner. This task and component is in charge of automated composition of web services into executable processes, see [31]. Composed web services perform new functionalities by interacting with pre-existing services that are published on the Web.

Content Annotation and Annotation Manager. This task and component is in charge of automatic production of metadata for the contents, see [32]. Annotation manager takes as input the (pre-processed) contents and domain knowledge and produces as output a database of content annotations. In addition to the automatic production of content metadata, prompt mechanisms should enable the user with a possibility to enrich the content annotation by adding some extra information (e.g., title, name of a location, title of an event, names of people) that could not be automatically detected.

Reasoning and Reasoner. This task and component is in charge of providing logical reasoning services (e.g., subsumption, concept satisfiability, instance checking tests), see [33]. For example, when dealing with multimedia annotations, logical reasoning can be exploited in order to check consistency of the annotations against the set of spatial (e.g., left, right, above, adjacent, overlaps) and temporal (e.g., before, after, during, co-start, co-end) constraints. Thus, ensuring that the objects detected in the multimedia content correspond semantically to the concepts defined in a domain ontology. For example, in the racing domain, it should be checked whether a car is located above a road or whether the grass and sand are adjacent to the road.

Semantic Query Processing and Query Processor. This task and component is in charge of rewriting a query by using terms which are explicitly specified in the model of a domain knowledge in order to provide a semantics- preserving query answering, see [32, 34]. Examples of queries are “Give me all the games played on grass” or “Give me all the games of double players”, in the tennis domain. Finally, users should be able to query by a sample image. In this case, the system should perform an intelligent search of images and videos (e.g., by using semantic annotations) where, for example, the same event or type of activity takes place.

Personalization and Profiler. This task and component is in charge of tailoring services available from the system to the specificity of each user, see [35]. For example, generation and updating of user profiles, recommendation generation, inferring user preferences, and so on. For example users might want to share annotations within trusted user networks, thus having services of personal metadata management and contact’s recommender. Also, a particular form of personalization, which is media adaptation, requires knowledge-based technology for a suitable delivery of the contents to the user’ terminal (e.g., palm, mobile phone, portable PC).

4. Conclusions and future work

The most relevant initiative to our efforts is IST-FP5 Ontoweb (2001-2004). It formed a special interest group (SIG) on Industrial Applications²¹ which collected over 50 use cases. However, the majority of those use cases dealt with technology producers rather than potential adopters of the technology. Ontoweb achieved a good overview of the main roadblocks on the way towards a successful transfer of knowledge-based technology to industry. Based on those foundations, the subsequent IST-FP6 Network of Excellence KnowledgeWeb (2004-2007), has continued the Ontoweb initiative by going into the detail of each particular business case, targeting at (i) collecting industry needs from potential client industry with a specific focus on a few most promising sectors; (ii) identifying the key processing components emerging from the concrete needs analysis; (iii) evaluating research and technology for answering industry needs; (iv) making recommendations through best-of-class guidelines; (v) providing education for practitioners via competence centers, thereby enabling the transfer of a technology know-how.

²¹ <http://ago.sig4.fr>

In this paper we have reported some results on the first two topics as addressed by Knowledge Web. By a preliminary analysis of the collected use cases we categorized the types of solutions being sought for, and the types of technological locks which arise when realizing those solutions. By a detailed technical analysis of the selected use cases we identified precisely where in the business processes the technology locks occur, described the requirements for technological solutions that overcome those locks, and argued for the appropriateness of knowledge-based solutions. Moreover, a quick analysis of the other business cases of [11] have shown that most of the knowledge processing tasks of Table 1 repeat with some variations/specificity from use case to use case. This observation suggests that the constructed typology is stable, i.e., it contains (most of) the core knowledge processing tasks stipulated by the current industry needs. By identifying concrete industry needs through tasks and components, we link them to specific research challenges which we expect the Semantic Web researchers to focus on. As such components are made available from the research, it is possible to evaluate them in different industry-strength settings, and therefore, estimate their practical impact and a contribution to the industrial uptake of Semantic Web technology.

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. Thus, future work includes continuing to collect business cases and to carry out their technical analysis until the saturation is reached.

6. Acknowledgments

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

References

1. K.Wiig, Knowledge management: where did it come from and where will it go? *Journal of Expert Systems with Applications*, 13(1), 1–14, 1997.
2. J. Hibbard, Knowledge management—knowing what we know. *Information Week*, 20 October 1997.
3. G. Petrash, Managing knowledge assets for value. In *Proceedings of the Knowledge-Based Leadership Conference*, Boston, MA, October 1996. Boston, MA: Linkage.
4. Zyl J, Corbett D. (2000), A framework for Comparing the use of a Linguistic Ontology in an Application, *Workshop Applications of Ontologies and Problem-solving Methods, ECAI'2000*, Berlin Germany, August, 2000
5. Guarino N., Masolo C., Vetere G., *OntoSeek: (1999) Content-Based Access to the Web*, *IEEE Intelligent System*.
6. MKBEEM (2002) *Multilingual Knowledge-Based E-Commerce* <http://www.mkbeem.com>
7. Wiederhold G. (1992). *Mediators in the architecture of future information systems*, *Computer*, Vol. 25(3). p.38-49
8. Barillot C., Amsaleg L., Aubry F., Bazin J-P., Benali H., Cointepas Y., Corouge I., Dameron O., Dojat M., Garbay C., Gibaud B., Gros P., Inkingehun S., Malandain G., Matsumoto J.,

- Papadopoulos D., Péligrini M., Richard N., Simon E., Neurobase: Management of distributed knowledge and data bases in neuroimaging . In *Human Brain Mapping*, Volume 19, Pages 726-726, New-York, NY, 2003.
9. Cordonnier E., Croci S., Laurent J.-F., Gibaud B. (2003) Interoperability and Medical Communication Using “Patient Envelope”-Based Secure Messaging Proceedings of the Medical Informatics Europe Congress.
 10. Charlet J., Cordonnier E., Gibaud B. (2002) Interopérabilité en médecine: quand le contenu interroge le contenant et l’organisation. *Revue Information, interaction, intelligence* 2(2).
 11. L. Nixon, M. Mochol, A. Léger, F. Paulus, L. Rocuet, M. Bonifacio, R. Cuel, M. Jarrar, P. Verheyden, Y. Kompatsiaris, V. Papastathis, S. Dasiopoulou, and A. Gomez Pérez. D1.1.2 Prototypical Business Use Cases. Technical report, Knowledge Web NoE, 2004.
 12. P. Shvaiko, A. Léger, F. Paulus, L. Rocuet, L. Nixon, M. Mochol, Y. Kompatsiaris, V. Papastathis, and S. Dasiopoulou. D1.1.3 Knowledge Processing Requirements Analysis. Technical report, Knowledge Web NoE, 2004.
 13. D. Dou, D. McDermott, and P. Qi. Ontology translation on the Semantic Web. *Journal on Data Semantics*, pages 35–57, 2005.
 14. Stanford Medical Informatics. Protégé ontology editor and knowledge acquisition system. <http://protege.stanford.edu/index.html>.
 15. D. L. McGuinness, R. Fikes, J. Rice, and S. Wilder. An environment for merging and testing large ontologies. In *Proceedings of KR*, pages 483–493, 2000.
 16. E. Rahm and P. Bernstein. A survey of approaches to automatic schema matching. *VLDB Journal*, (10(4)):334–350, 2001.
 17. P. Shvaiko and J. Euzenat. A survey of schema-based matching approaches. Submitted to the *Journal on Data Semantics*, 2004.
 18. A. Billig and K. Sandkuhl. Match-making based on semantic nets: The xml-based approach of baseweb. In *Proceedings of the 1st workshop on XML-Technologien für das Semantic Web*, pages 39–51, 2002.
 19. M. Ehrig and S. Staab. QOM: Quick ontology mapping. In *Proceedings of ISWC*, pages 683–697, 2004.
 20. J. Euzenat and P. Valtchev. Similarity-based ontology alignment in OWL-lite. In *Proceedings of ECAI*, pages 333–337, 2004.
 21. H.H. Do and E. Rahm. COMA - a system for flexible combination of schema matching approaches. In *Proceedings of VLDB*, pages 610–621, 2001.
 22. J. Zhong, H. Zhu, J. Li, and Y. Yu. Conceptual graph matching for semantic search. In *Proceedings of the 2002 International Conference on Computational Science*, 2002.
 23. F. Giunchiglia and P. Shvaiko. Semantic matching. *Knowledge Engineering Review Journal*, (18(3)):265–280, 2003.
 24. F. Giunchiglia, P. Shvaiko, and M. Yatskevich. S-Match: an algorithm and an implementation of semantic matching. In *Proceedings of ESWS*, pages 61–75, 2004.
 25. T. Di Noia, E. Di Sciascio, F. M. Donini, and M. Mongiello. A system for principled matchmaking in an electronic marketplace. In *Proceedings of WWW*, pages 321–330, 2003.
 26. R. Dhamankar, Y. Lee, A. Doan, A. Halevy, and P. Domingos. iMAP: Discovering complex semantic matches between database schemas. In *Proceedings of SIGMOD*, pages 383 – 394, 2004.
 27. P. Shvaiko, F. Giunchiglia, P. Pinheiro da Silva, and D. L. McGuinness. Web explanations for semantic heterogeneity discovery. In *Proceedings of ESWC*, 2005.
 28. J. Petrini and T. Risch. Processing queries over rdf views of wrapped relational databases. In *Proceedings of the 1st International workshop on Wrapper Techniques for Legacy Systems*, Delft, Holland, 2004.
 29. Y. Velegrakis, R. J. Miller, and J. Mylopoulos. Representing and querying data transformations. In *Proceedings of ICDE*, 2005.

30. N. Pregoica, M. Shapiro, and C. Matheson. Semantics-based reconciliation for collaborative and mobile environments. In *Proceedings of CoopIS*, 2003.
31. P. Traverso and M. Pistore. Automated composition of semantic web services into executable processes. In *Proceedings of ISWC*, pages 380–394, 2004.
32. aceMedia project. Integrating knowledge, semantics and content for user centred intelligent media services. <http://www.acemedia.org>
33. V. Haarslev, R. Møller, and M. Wessel. RACER: Semantic middleware for industrial projects based on RDF/OWL, a W3C Standard. <http://www.sts.tu-harburg.de/~r.f.moeller/racer/>
34. E. Mena, V. Kashyap, A. Sheth, and A. Illarramendi. Observer: An approach for query processing in global information systems based on interoperability between pre-existing ontologies. In *Proceedings of CoopIS*, pages 14–25, 1996.
35. G. Antoniou, M. Baldoni, C. Baroglio, R. Baumgartner, F. Bry, T. Eiter, N. Henze, M. Herzog, W. May, V. Patti, R. Schindlauer, H. Tompits, and S. Schaffert. Reasoning Methods for Personalization on the Semantic Web. *Annals of Mathematics, Computing & Teleinformatics*,