Semantic Web take-off in a European Industry Perspective

Alain Léger[1], Johannes Heinecke[1], Lyndon J.B. Nixon[2], Pavel Shvaiko[3],

Jean Charlet[4], Paola Hobson[5], François Goasdoué[6]

[1] France Telecom R&D - Rennes,
4 rue du clos courtel, 35512 Cesson-Sévigné, France
a-t.leger@wanadoo.fr; johannes.heinecke@orange-ftgroup.com

[2] Freie Universität Berlin
Takustrasse 9, 14195 Berlin, Germany
nixon@inf.fu-berlin.de

[3] University of Trento
Via Sommarive 14, 38050 Trento, Italy
pavel@dit.unitn.it

[4] STIM, DPA/AP-Hopitaux Paris & Université Paris 6
75006 Paris, France
charlet@biomath.jussieu.fr

[5] Motorola Labs, Centre for Applications, Content and Services
Hants, UK
Paola.hobson@motorola.com

[6] LRI, CNRS et Université Paris Sud XI
Bâtiment 490, 91405 Orsay Cedex, France
fg@lri.fr

Abstract

Semantic Web technology is being increasingly applied in a large spectrum of applications in which domain knowledge is conceptualized and formalized (e.g., by means of an ontology) in order to support diversified and automated knowledge processing (e.g., reasoning) performed by a machine. Moreover, through an optimal combination of (cognitive) human reasoning and (automated) machine processing (mimicking reasoning); it becomes possible for humans and machines to share more and more complementary tasks. The spectrum of applications is extremely large and to name a few: corporate portals and knowledge management, e-commerce, e-work, e-business, healthcare, e-government, natural language understanding and automated translation, information search, data and services integration, social networks and collaborative filtering, knowledge mining, business intelligence and so on. From a social and economic perspective, this emerging technology should contribute to growth in economic wealth, but it must also show clear cut value for everyday activities through technological transparency and efficiency. The penetration of Semantic Web technology in industry and in services is progressing slowly but accelerating as new success stories are reported. In this chapter we present ongoing work in the cross-fertilization between industry and academia. In particular, we present a collection of application fields and use cases from enterprises which are interested in the promises of Semantic Web technology. The use cases are focused on the key knowledge processing components that will unlock the deployment of the technology in industry. The chapter ends with the presentation of the current state of the technology and future trends as seen by prominent actors in the field.

1 Industry perspective

As a result of the pervasive and user-friendly digital technologies emerging within our information society, web content availability is increasing at an incredible rate but at the cost of being extremely multiform, inconsistent and very dynamic. Such content is totally unsuitable for machine processing, and so necessitates too much human interpretation and its respective costs in time and effort for both individuals and companies. To remedy this, approaches aim at abstracting from this complexity (i.e., by using ontologies) and offering new and enriched services able to process those abstractions (i.e., by mechanized reasoning) in a fully – and trusted - automated way. This abstraction layer is the subject of a very dynamic activity in research, industry and standardization which is usually called "Semantic Web" (see for example, DARPA, European IST Research Framework Program, W3C initiative, OASIS). The initial application of Semantic Web technology has focused on Information Retrieval (IR) where access through semantically annotated content, instead of classical (even sophisticated) statistical analysis, aimed to give far better results (in terms of precision and recall indicators). The next natural extension was to apply IR in the integration of enterprise legacy databases in order to leverage existing company information in new ways. Present research has turned to focusing on the seamless integration of heterogeneous and distributed applications and services (both intra- and inter-enterprise) through Semantic Web Services, and respectful of the legacy systems already in place, with the expectation of a fast return on investment (ROI) and improved 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, automated reasoning, logics and decision theory. It can be separated into 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 a socio-cognitive Semantic Web (automated social Web) which is gaining momentum today with the Web 2.0 paradigm[1].

On a large scale, industry awareness of Semantic Web technology has started at the EC level with the IST-FP5 thematic network Ontoweb[2] [2001-2004] which brought together around 50 motivated companies worldwide. Based on this experience, within IST-FP6, the Network of Excellence Knowledge Web[3] [2004-2008] made an in-depth analysis of the concrete industry needs in key economic sectors, and in a complementary way the IST-FP6 Network of Excellence Rewerse[4] was tasked with providing Europe with leadership in reasoning languages, also in view of a successful technology transfer and awareness activities targeted at the European industry for advanced Web systems and applications. This impetus will continue and grow up in the EU IST-FP7 [2007-20013][5].

The rest of the chapter is organized as follows. Four prototypical application fields are presented in Section 2, namely (i) healthcare and biotechnologies, (ii) knowledge management (KM), (iii) e-commerce and e-business, and finally (iv) multimedia and audiovisual services. Finally, Section 3 reports on a current vision of the achievements and some perspectives are given.

1.1 Overall business needs and key knowledge processing requirements

1.1.1 Use case collection and analysis

In order to support a large spectrum of application fields, two EU FP6 Networks of Excellence NoE-Knowledge Web and NoE-REWERSE are tasked with promoting transfer of best-of-the-art knowledge-based technology from academia to industry. The networks are made up of leading European Semantic Web research institutions that co-ordinate their research efforts while parallel efforts are made in Semantic Web education to increase the availability of skilled young researchers and practitioners and last but not the least, in pushing the take-up in Business and Industry.

In order to accelerate the transfer from research to industry, the objective of an Industry-Research co-operation is to establish a working relationship between Semantic Web researchers and an industry partner, in which research results being produced in an area of Semantic Web research will be prototypically applied to the industry partner’s selected business case. The co-operation not only seeks to achieve an individual success story in terms of some specific research and a given business case, but also to establish the value of Semantic Web technologies to industrial application in a more general sense. It achieves this by demonstrating the use of Semantic Web technology in a business setting, exploring their usefulness in solving business problems and ensuring future applicability by directing researchers towards meeting industrial requirements in their work.

In NoE-Knowledge Web, an Industry Board was formed at the beginning of the network to bring together potential early adopters of Semantic Web technologies from across a wide spread of industry

[1] http://www.web2con.com

[2] http://www.ontoweb.org

[3] http://Knowledge Web.semanticweb.org

[4] http://rewerse.net

[5] http://cordis.europa.eu/fp7/home_en.html

sectors. Industry Board members have been involved in many initiatives of the Knowledge Web Industry Area, including the collection of business use cases and their evaluation. In order to more directly achieve close co-operation between researchers and industry, each research activity in the network was invited to select a use case whose requirements closely correlated to what would be achieved in their research work. Results have been collected and reported in July 2007[6].

Currently in 2007, this Industry Board consists of about 50 members (e.g., France Telecom, British Telecom, Institut Français du Pétrole, Illy Caffe, Trenitalia, Daimler Chrysler, Thalès, EADS, …) from across 14 nations and 13 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, a total of 16 use cases were collected from 12 companies. In 2007, through many workshops and Industry forum sessions at major Semantic Web conferences, more than a hundred use cases were available or illustrative of the current trend to introduce Semantic Web technology in the main stream.

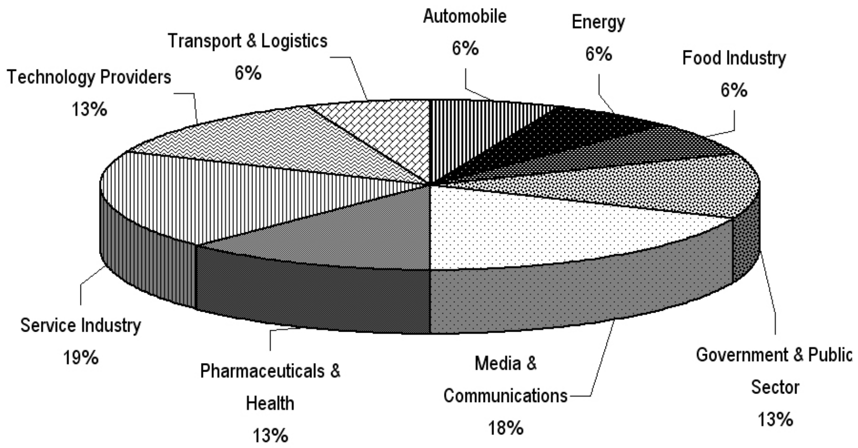

Figure 1. *Breakdown of use cases by industry sector*

As shown in Figure. 1, where the use cases are broken down according to the industry sector, collected cases are from 9 industry sectors, with the highest number of the use cases coming from the service industry (19%) and media & communications (18%) respectively. This initial collection of use cases can be found in (Nixon L. *et al.,* 2004), while a constantly growing and updated selection is available on the Knowledge Web Industry portal[7].

The co-operations have been a very challenging activity, given the early state of much cutting edge Semantic Web research and the differences in perspective between academia and business. However, successes have been reported , not only in the production of some prototypical solutions and demos which can be shown to industry and business top managers, but also in making researchers more aware of the importance of their work to solving business problems and the earlier recognition by academics of industry needs and expectations and so integrating them to their research agenda.

Hence, the Industry-Research co-operations in NoE-Knowledge Web and NoE-REWERSE must be seen as a significant first attempt to align the ambitious cutting edge work on Semantic Web technologies done by leading researchers in Europe and the real world business problems encountered by the european industry which may find a potential solution in those same Semantic Web technologies. Given a continued rise in awareness among Semantic Web researchers of the applicability of their work to industry and the continued rise in awareness among industry of the potential of the work of Semantic Web researchers, which has been begun in IST-NoEs, in IST-R&D projects, but also clearly in industry (SMEs and large companies), the technology transfer is gradually evolving.

---

[6] Knowledge Web Deliverable D 1.1.4v3
http://knowledgeweb.semanticweb.org/semanticportal/deliverables/D1.1.4v3.pdf

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

1.1.2 Analysis of Use Cases by Expert Estimations

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, the service functionalities they would be offered and the technological locks they encountered, and eventually how they expected that Semantic Web technology could resolve those locks. As a result, we have gained an overview of:

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

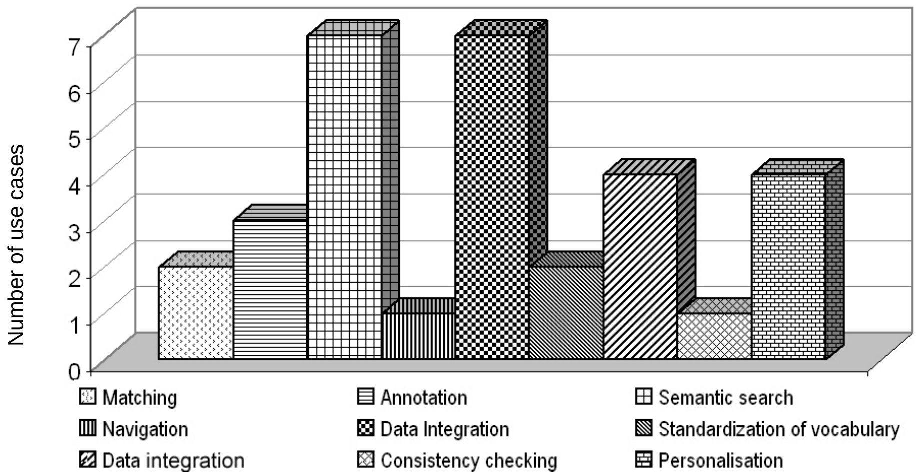

Figure 2. *Breakdown of use cases by industry sector*

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.

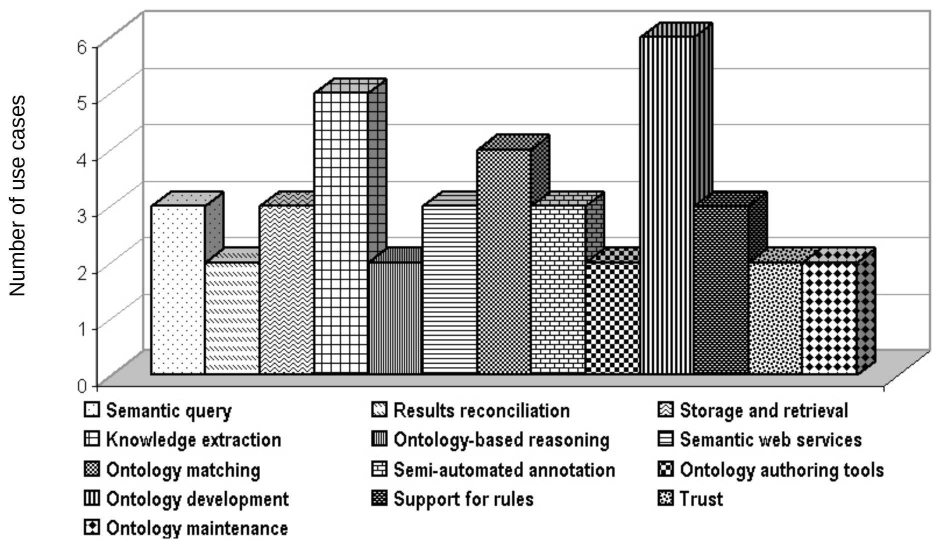

Figure 3. *Preliminary vision of technology locks in use cases*

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. These are: 1) ontology development, i.e., modeling of a business domain, authoring, reusing existing ontologies; 2) knowledge extraction, i.e., populating ontologies by extracting data from legacy systems; and 3) ontology mapping, i.e., resolving semantic heterogeneity among multiple ontologies.

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

The company generates a 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 document retrieval and to enable personalization services for 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.

Eventually, this analysis (by experts estimations) provides us with a preliminary understanding of scope of the current industrial needs and the current 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.

1.1.3 Knowledge Processing Tasks and Components

Based on the knowledge processing needs identified during the technical use cases analysis (Shvaiko P. *et al.*, 2004), we built a typology of common knowledge processing tasks and a library of high level components for realizing those tasks, see Table 1. 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, 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 (Dou D. *et al.*, 2005) (McGuiness D. *et al.*, 2000) (Protégé[8]), OAEI-2007Ontology Alignment Evaluation Initiative[9], NeOn[10] (Networked Evolving Ontologies) and Ontology Matching survey site[11].

Ontology Matching, Matching Results Analysis, Producing Explanations and Match Manager. These tasks and component are in charge of (on-the-fly and semi-automatic) determination of semantic mappings between the entities of multiple schemas and ontologies, see (Rahm E. *et al.*, 2001) (Shvaiko P. and Euzenat, 2005), (Euzenat J. and Shvaiko P., 2007). 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 (Billig A. *et al.*, 2002) (Ehrig M. *et al.*, 2004) (Euzenat J. *et al.*, 2004) (Do H. H. *et al.*, 2002) (Zhong J. *et al.*, 2002) or in the form of a logical relation (e.g., equivalence, subsumption), see (Giunchiglia F. *et al.*, 2003) (Giunchiglia F. *et al.*, 2004). The mappings might need to be ordered according to some criteria, see (Di Noia T. *et al.*, 2003) (Do H. H. *et al.*, 2002).

Finally, explanations of the mappings might be also required, see (Dhamankar R. *et al.*, 2004) (Shvaiko P. *et al.*, 2005). 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 and 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 (Hull R. 1997) (Petrini J. *et al.*, 2004) (Velegrakis Y. *et al.*, 2005) (Halevy A. et al., 2006). Here, mappings are taken as input (for example, from the match

[8] http://protege.stanford.edu/index.html

[9] http://oaei.ontologymatching.org/2007/

[10] http://www.neon-project.org

[11] http://www.ontologymatching.org/

manager component) and are the support for generating query expressions that perform the required semantic and syntactical manipulations with data instances coming from heterogeneous environment..

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 (Preguica N. *et al.*, 2003).

N°	Knowledge processing tasks	Components
1	Ontology Management	Ontology Manager
2	Ontology Matching	Match Manager
3	Ontology 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	Ontology Merging	Ontology Manager
11	Producing explanations	Match Manager
12	Personalization	Profiler

Table 1. *Typology of knowledge processing tasks & components*

Composition of Web Services and Planner. This task and component is in charge of automated composition of web services into executable processes (Orchestration). Composed web services perform new functionalities by specific on demand interaction with pre-existing services that are published on the Web, see surveys from (Chan et al., 2007) (Berardi et al., 2005) (Hull et al., 2005) (Pistore et al., 2005) (Roman et al., 2005) (Traverso P. *et al.*, 2004) (Cardoso et al., 2003) (McIlraith et al., 2001). From a business viewpoint, it remains a key challenge to be overcome, as the businesses react very positively to the need for a very effective integration technology and for more agility in a very competitive worldwide economy. In the meantime, reducing interoperability problems will open opportunities for easier innovative solutions and for the increase in cooperation between enterprises. This should result in re-combinations of businesses the technology provides and so will have a profound impact on business and economic workflows.

Content Annotation and Annotation Manager. This task and component is in charge of automatic production of metadata for the contents, see aceMedia[12] for multimedia annotation. 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 offer the user the possibility to enrich the content annotation by adding extra information (e.g., title, name of a location, title of an event, names of people) that could not be automatically detected.

Automated Reasoning. This task and component is in charge of providing logical reasoning services (e.g., subsumption, concept satisfiability, instance checking tests), see (Haarslev V. *et al.*, 1999-2007). 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. This can certify that the objects detected in the multimedia content correspond semantically to the concepts defined in the domain ontology. For example, in the racing domain, the automated reasoner should check whether a car is located on 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 posed by a human being or a machine, by using terms which are explicitly specified in the model of domain knowledge in order to provide semantics preserving query answering, see (Mena E. *et al.*, 1996) (Halevy et al., 2001) (Calvanese et al., 2002) (IST-IP aceMedia 2004). Examples of queries are

[12] http://www.acemedia.org

"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 sample content e.g. an 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 (Antoniou G. *et al.*, 2004). 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 contacts recommendation. Also, a particular form of personalization, which is media adaptation, requires knowledge-based technology for a suitable delivery of the contents to the user's terminal (e.g., PDA, mobile phone, portable PC).

2 Key application sectors and typical technology problems

2.1 Healthcare and biotechnologies

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 very complex: medical knowledge is difficult to represent in a computer format, making the sharing of information even more difficult. Semantic Web solutions become very promising in this context.

One of the main mechanisms of the Semantic Web - resource description using annotation principles - is of major importance in the medical informatics (or sometimes called bioinformatics) domain, especially as regards the sharing of these resources (e.g. medical knowledge on the Web or genomic database). Through the years, the IR area has been developed by medicine: medical thesauri are enormous (e.g., more than 1,600,000 terms in Unified Medical Language System, UMLS[13]) and are principally used for bibliographic indexation. Nevertheless, the MeSh thesaurus (Medical Subject Heading) or UMLS have been used to provide data semantics with varying degrees of difficulty. Finally, the web services technology allows us to imagine some solutions to the interoperability problem, which is substantial in medical informatics. Below, we will describe current research, results and expected perspectives in these biomedical informatics topics in the context of Semantic Web.

2.1.1 Biosciences resources sharing

In the functional genomics domain, it is necessary to have access to several databases and knowledge bases which are accessible separately on the Web but are heterogeneous in their structure as well as in their terminology. Among such resources, we can mention SWISSPROT[15] where the gene products are annotated by the Gene Ontology[16], Gen-Bank[17], etc. When comparing these resources it is easy to see that they propose the same information in different formats. The XML language, which acts as a common data structure for the different knowledge bases, provides at most a syntactic Document Type Definition (DTD) which does not resolve the semantic interoperability problem.

One of the solutions comes from the Semantic Web with a mediator approach (Wiederhold G., 1992) which allows for the accessing of different resources with an ontology used as the Interlingua pivot. For example and in another domain than that of genomics, the NEUROBASE project (Barillot C. *et al.*, 2003) attempts to federate different neuro-imagery information bases situated in different clinical or research areas. The proposal consists of defining an 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

[13] http://www.nlm.nih.gov/research/umls/umlsmain.html

[15] http://us.expasy.org/sprot/

[16] http://obo.sourceforge.net/main.html

[17] http://www.ncbi.nlm.nih.gov/Genbank/index.html

several databases. The mediator of NEUROBASE has been tested on decision support systems in epilepsy surgery.

2.1.2 Ontologies for coding systems

The main usage of ontologies in medical domain is as index of coding system: after using thesauri for indexing medical bibliography (PubMed with the Mesh[18]), the goal is to index Electronic Health records with medical concept in order to enhance information retrieval or to allow epidemiological studies. For that purpose, several countries intend to use the SNOMED, an international classification of concepts organized in eight axes (Spackman *et al.*, 2002). Except the problem of languages, this classification exists in two versions: a classification of 160,000 concepts (SNOMED-I V3.5) and an ontology, which is the evolution of the preceding one, of 330,000 concepts, SNOMED CT. The use of ontologies of such a size is difficult. Some authors describe them as *reference ontology* which cannot be accessed without an *interface ontology* (Rosenbloom *et al.*, 2006).Notwithstanding, UK national health system (NHS) is integrating SNOMED CT and it will be interesting to examine the results of this industrial deployment[19].

2.1.3 Web services for interoperability

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

The patient envelope is a proposition of the Electronic Data Interchange for Healthcare group (EDI-Santé[20]) with an active contribution from the ETIAM[21] society. The objective of the work is on filling the gap between "free" communication, using standard and generic Internet tools, and "totally structured" communication as promoted by CEN (in the Working Group IV "Technology for Interoperability"[22]) or HL7[24]. After the 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 E. *et al.*, 2003). This proposal has reached widespread adoption 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 (see items 1 and 2 below) and medical (see items 3 and 4 below) properties:

1. **Separate what is mandatory** to the transport and management of the message (e.g., patient identification from what constitutes the "job" part of the message.
2. **Provide a "container" for the message**, collecting the different elements, texts, pictures, videos, etc.
3. **Consider the patient as the unique object of the transaction**. Such an exchange cannot be anonymous. It concerns a sender and an addressee who are involved in the exchange and who are responsible. A patient can demand to know the content of the exchange in which (s)he is the object, which implies a data structure which is unique in the form of a triple {sender, addressee, patient}.
4. **The conservation of the exchange semantics.** The information about a patient is multiple in the sense that it comes from multiple sources and has multiple forms and supporting data (e.g., database, free textual document, semi-structured textual document, pictures). It can be

---

[18] http://www.ncbi.nlm.nih.gov/sites/entrez

[19] http://www.connectingforhealth.nhs.uk/

[20] http://www.edisante.org/

[21] http://www.etiam.com/

[22] http://cen.iso.org/ and http://www.tc251wgiv.nhs.uk/

[24] http://www.hl7.org/

fundamental to maintain the existing links between elements, to transmit them together, e.g., a scanner and the associated report, and to be able to prove it.

The interest of such an approach is that it prepares the evolution of the transmitted document from a free form document (from proprietary ones to normalized ones as XML) to elements respecting HL7v3 or EHRCOM data types. In France, the GIP-DMP[25] retains such an approach (in conjunction with the Clinical Document Architecture of HL7[26]) for the implementation of the exchanges of the *Dossier Médical Personnel* (a future national electronic health record).

2.1.4 What is next in the healthcare domain?

These different projects and applications highlight the main consequence of the Semantic Web being expected by the medical communities: the sharing and integration of heterogeneous information or knowledge. The answers to the different issues are the use of mediators, a knowledge-based system, and ontologies, which should be based in the mid term on normalized languages such as RDF, OWL but also in addition to come OWL-S, SAWSDL, WSML, SWRL, or RuleML. The work of the Semantic Web community must take into account these expectations, see for example the FP6 projects[27,28,29]. Finally, it is interesting to note that the Semantic Web is an integrated vision of the medical community's problems (thesauri, ontologies, indexation, and inference) and provides a real opportunity to synthesize and reactivate some research (Charlet J. *et al.*, 2002).

2.2 Knowledge Management

2.2.1 Leveraging Knowledge assets in companies

Knowledge is one of the key success factors for enterprises, both today and in the future. Therefore, company knowledge management has been identified as a strategic tool. However, if for KM information technology is one of the foundational elements, KM in turn, is also interdisciplinary by its nature. In particular, it includes human resource management as well as enterprise organization and culture[30].We view KM as the management of the knowledge arising from business activities, aiming at leveraging both the use and the creation of that knowledge for two main objectives: capitalization of corporate knowledge and durable innovation fully aligned with the strategic objectives of the organization.

Conscious of this key factor of productivity an ever faster changing ecosystem, the European KM Framework (CEN/ISSS[31], KnowledgeBoard[32]) has been 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 of the 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.

2.2.2 Knowledge-based KM benefits

The knowledge backbone is made up of ontologies that define a shared conceptualization of an application domain and provide the basis for defining metadata that have precisely defined semantics, and

---

[25] http://www.d-m-p.org/docs/EnglishVersionDMP.pdf

[26] http://www.hl7.org/Special/Committees/structure/index.cfm#Mission

[27] http://www.cocoon-health.com

[28] http://www.srdc.metu.edu.tr/webpage/projects/artemis/index.html

[29] http://www.simdat.org

[30] Some of the well-known definitions of KM include: (Wiig 1997) " 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"; (Hibbard 1997) "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"; (Pettrash 1996) "KM is getting the right knowledge to the right people at the right time so they can make the best decision"

[31] http://www.cenorm.be/cenorm/index.htm

[32] http://www.knowledgeboard.com

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 technology a complete success for KM solutions:

- Industrial KM applications **have to avoid any kind of overhead as far as possible**. **A seamless integration of knowledge creation** (i.e., content and metadata specification) and knowledge access (i.e., 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 that can be supported through ontology learning, which should reduce the current time consuming task of building-up and maintaining ontologies.
- **Access to as well as presentation of knowledge has to be context-dependent**. Since the context is setup 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 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 the 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 at the right level of granularity and from the right perspective at the right time.

The development of knowledge portals serving the needs of companies or communities is still 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 mechanisms *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 technology is a promising approach to meet the needs of mobile environments, like 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*.

Social networking (W2.0), combined with Semantic Web technology, will be a strong move towards getting rid of the more centralized KM approaches that are currently used in ontology-based solutions. W2.0 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 with *e-learning* (e.g., skill adaptation in companies) 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 portals KM solutions have already reached their limits, whereas semantic technology opens the way to meet KM requirements in the future.

2.2.3 Knowledge-based KM applications

In the context of geographical team dispersion, multilingualism and business unit autonomy, usually a company wants a solution allowing for the identification of strategic information, the secured distribution of this information and the creation of transverse working groups. Some applicative solutions allowed for the deployment of an Intranet intended for all the marketing departments of the company worldwide, allowing for a better division of and a greater accessibility to information, but also capitalisation on the total knowledge. There are four crucial points that aim at easing the work of the various marketing teams in a company: (i) Business intelligence, (ii) Skill and team management[33], (iii) Process management and (iv) Rich document access and management.

Thus, a 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, knowledge representation). The result is a dynamic Semantic Web system of navigation (research, classification) and collaborative features.

At the end from a functional point of view, a KM system organises skill and knowledge management within a 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, and offers advanced features of capitalisation (Bonifacio M. *et al.*, 2005) (Brunschweig B. *et al.*, 2005) (Nordheim D. *et al.*, 2005).

Finally, the semantic backbone makes possible to cross a qualitative gap by providing cross-lingual data.

2.3 E-Commerce and E-Business

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 unified 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) heading. 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) see (Guarino N. 1999) (Zyl J. *et al.*, 200) (Lehtola A. et al., 2003) (Heinecke J. et al., 2003) (Benatallah B et al., 2005).

2.3.1 Knowledge-based E-Commerce and E-Business value

At present, ontology and more generally knowledge-based systems appear as a central issue for the development of efficient and profitable e-commerce and e-business solutions. However, because of the actual situation i.e. the partial standardization of business models, processes, and knowledge architectures, it is currently difficult for companies to achieve the promised ROI from knowledge-based e-commerce and e-business.

Moreover, a technical barrier exists that is delaying the emergence of e-commerce, lying in the need for applications to *meaningfully share information*, taking into account the lack of reliability, security and eventually trust in the Internet. This fact may be explained by the variety of e-commerce and e-business systems employed by businesses and the various ways these systems are configured and used. As an important remark, such *interoperability problems* become particularly severe 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 indicator.

---

[33] Semantic Web, Use Cases and Challenges at EADS, http://www.eswc2006.org Industry Forum.

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 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 widely, and consequently, cannot always be aligned for a variety of technical, practical, organizational and political reasons.
- The complexity of a 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 not always possible 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 therefore, potentially reduce their ability to fully participate in e-commerce.

A knowledge-based approach has the potential to significantly accelerate the penetration of electronic commerce within vertical industry sectors, by *enabling interoperability at the business leve*l. This will enable services to adapt to the rapidly changing business ecosystem.

2.3.2 Knowledge-based E-Commerce and E-Business applications

The Semantic Web brings opportunities to industry to create new services[34], extend markets, and even develop new businesses since it enables the inherent meaning of the data available in the Internet to be accessible to systems and devices able to interpret and reason at the knowledge level. This in turn leads to new revenue opportunities, since information becomes more readily accessible and usable. For example, a catering company whose web site simply lists the menus available is less likely to achieve orders compared to one whose menus are associated with related metadata about the contents of the dishes, their origin (e.g., organic, non-genetically modified, made with local produce), links to alternative dishes for special diets, personalised ordering where a user profile can be established which automatically proposes certain menu combinations depending on the occasion (e.g., wedding banquet, business lunch). The latter case assumes that both provider-side knowledge generation and knowledge management tools are available, such that the asset owner can readily enhance their data with semantic meaning, and client-side tools are available to enable machine interpretation of the semantic descriptions related to the products being offered, such that the end user can benefit from the available and mined knowledge. Examples of some possible application areas were studied by the Agent Cities project[35].

In the e-business area Semantic Web technology can improve standard business process management tools. One prototypical case is in the area of logistics. The application of knowledge technology on top of today's business management tools enables the automation of major tasks of business process management[36] see (Semantic Web Case Studies for eBusiness 2005).

In one of the Knowledge Web Industry-Research co-operations, a number of scenarios within the **B2B integration scenario** were identified, involving data mediation, discovery, and composition of services. All of these use cases have been evaluated according to a community-agreed methodology defined by the SWS challenge methodology with satisfying success levels defined by the methodology. This is an important step when proving the added value of the Semantic Web service technology applied to B2B integration domain. In addition, the standardization process has been partially finalized within the OASIS Semantic Execution Environment Technical Committee (OASIS SEE TC) and W3C Semantic Annotations for WSDL and XML Schema (W3C SAWSDL WG). However, the standardization process in both groups is still ongoing, but under business pressure should rapidly conclude.

The Industry-Research co-operation has **demonstrably solved a business case from the B2B domain**. We have shown how the technology deals with requirements from B2B domain and how this technology reacts to changes in back-end systems which might occur over the system's lifetime.

The research is not yet ready for industry. It must be shown how the technology is layered on the existing infrastructure and how it interacts with existing systems. For this purpose some parts of the technology need to be standardized (such as grounding mechanisms built on SAWSDL or the architecture). In particular, the grounding mechanism built on SAWSDL provides a "common interface"

---

[34] E.g. see the EU Integrated project "DIP Data, Information, and Process Integration with Semantic Web Services", http://dip.semanticweb.org/

[35] agentcities RTD project http://www.agentcities.org/EURTD/

[36] Semantic Business Automation, SAP, Germany http://www.eswc2006.org Industry Forum

between semantic descriptions and non-semantic descriptions (in our case WSDL). The standardization is still ongoing while at the same time, the alignment of service semantics with this grounding mechanism must be further finalised. While it has been demonstrated how this is possible to be done and what the added value of this approach is, **the complexity of business standards still needs to be addressed**.

In addition, a prototype is available[37] and has been provided to NoE-Knowledge Web industry partners see Fig. 4.

The following scenarios have been realised as part of the Semantic Web Services Challenge:

1. **Mediation Scenario** http://sws-challenge.org/wiki/index.php/Workshop_Budva. Addressing the mediation scenario for B2B integration when proprietary back-end systems of one company needed to be integrated with a partner using RosettaNet standard. Whole scenario has been successfully addressed.
2. **Discovery Scenario**. http://sws-challenge.org/wiki/index.php/Workshop_Athens. Addressing discovery scenario when a supplier needed to be discovered and selected from suitable ones. Whole scenario has been successfully addressed.
3. **Composition Scenario**. http://sws-challenge.org/wiki/index.php/Workshop_Innsbruck. Addressing composition scenario when more services can satisfy the user need. Whole scenario has been successfully addressed.

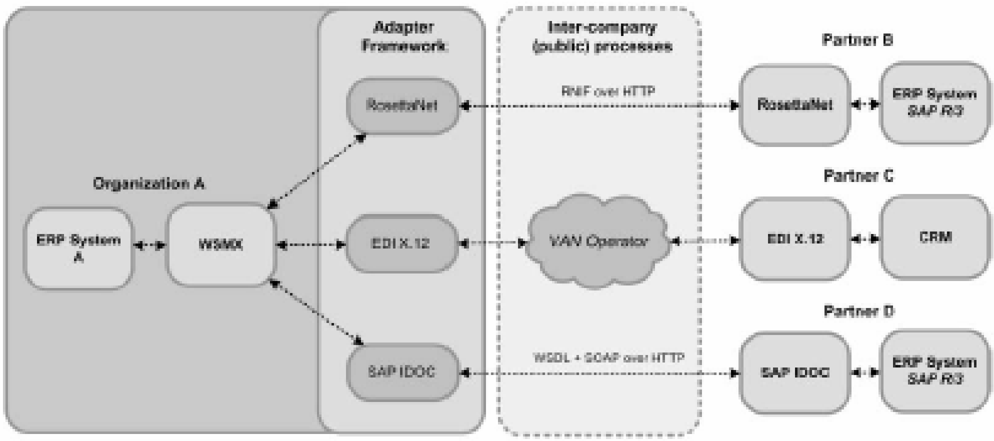

Figure 4. *Semantic Web Services integration in B2B*

Work will continue and the co-operation plans to address additional scenarios of the SWS challenge, namely scenarios when services can be filtered based on non-functional properties (QoS, financial, etc.). In addition, a tutorial was given on SWS in the context of business process management at ICWS'07 conference, and the authors co-organize the workshop on service composition and SWS challenge held at the Web Intelligence conference[38] (Vitvar T. *et al.*, 2007a) (Vitvar T. et al., 2007b) (Hasselwanter T. *et al.*, 2007).

2.4 Multimedia and audiovisual services

2.4.1 Multimedia and audiovisual services

Practical realisation of the Semantic Web vision is actively being researched by a number of experts, some of them within European collaborative projects, and others within company specific initiatives.

[37] http://sws-challenge.org/2006/submission/deri-submission-discovery-phase3/
http://sws-challenge.org/2006/submission/deri-submisson-mediation v.1/
http://sws-challenge.org/2006/submission/deri-submisson-mediation v.2/

[38] http://events.deri.at/sercomp2007/

Earlier projects such as SEKT[39] and DIP, mainly focused on enhancing text based applications from a knowledge engineering perspective. Although significant benefits in unlocking access to valuable knowledge assets are realised via these projects, in various domains such as digital libraries, enterprise applications, and financial services, it was soon recognised that there was a challenging and potentially highly profitable area of research into the integration of multimedia and Semantic Web technologies for multimedia content based applications. Projects such as aceMedia, BOEMIE, and MESH are examples of initiatives aiming to advance the use of semantics and reasoning for improved multimedia applications such as automated annotation, content summarisation, and personalised content syndication.

The drive for application of semantic technologies in the multimedia and content domains comes from a proliferation of audiovisual devices and services which have led to an exponential growth in available content. Users express dissatisfaction at not being able to find what they want, and content owners are unable to make full use of their assets. Service providers seek means to differentiate their offerings by making them more targeted toward the individual needs of their customers. Semantic Web technology can address these issues. It has the potential to reduce complexity, enhance choice, and put the user at the center of the application or service, and with today's fast mobile data services and availability of wifi, such benefits can be enjoyed by consumers and professional users in all environments using all their personal devices, in the home, at work, in the car and on the go.

Semantic Web technologies can enhance multimedia based products to increase the value of multimedia assets such as content items which are themselves the articles for sale (songs, music videos, sports clips, news summaries, etc) or where they are used as supporting sales of other goods (e.g. promotional images, movie trailers etc). Value is added in search applications, such that returned items more closely match the user's context, interests, tasks, preference history etc, as well as in proactive push applications such as personalised content delivery and recommendation systems, and even personalised advertising. However, applications such as content personalisation, where a system matches available content to the user's stated and learned preferences, thereby enabling content offerings to be closely targeted to the user's wishes, rely on the availability of semantic metadata describing the content in order to make the match. Currently, metadata generation is mostly manual, which is costly and time consuming. Multimedia analysis techniques which go beyond the signal level approach to a semantic analysis have the potential to create automatic annotation of content, thereby opening up new applications which can unlock the commercial value of content archives (Stamou *et al.,* 2006).

Automated multimedia analysis tools are important enablers in making a wider range of information more accessible to intelligent search engines, real-time personalisation tools, and user-friendly content delivery systems. Such automated multimedia analysis tools, which add the semantic information to the content, are critical in realising the value of commercial assets e.g. sports, music and film clip services, where manual annotation of multimedia content would not be economically viable, and are also applicable to users' personal content (e.g. acquired from video camera or mobile phone) where the user does not have time, or a suitable user interface, to annotate all their content.

Multimedia ontologies are needed to structure and make accessible the knowledge inherent in the multimedia content, and reasoning tools are needed to assist with identification of relevant content in an automated fashion. Although textual analysis and reasoning tools have been well researched, and despite the projects funded by the European Commission in the 6th framework, fewer tools are available for semantic multimedia analysis, since the problem domain is very challenging. However, automated multimedia content analysis tools such as those being studied within aceMedia[40] are a first step in making a **wider range of information more accessible to intelligent search engines, real-time personalisation tools, and user-friendly content delivery systems**.

Furthermore, interoperability of multimedia tools is important in enabling a wide variety of applications and services on multiple platforms for diverse domains. The W3C Multimedia Semantics Incubator Group reported on interoperability issues[41] and it is clear that a common framework using Semantic Web languages tools is essential for full exploitation of the potential of multimedia assets. Interoperability is essential in achieving commercial success with semantic multimedia applications, since it enables multiple manufacturers, content providers and service providers to participate in the market. This in turn enables consumer confidence to be achieved, and a viable ecosystem to be developed.

---

[39] Semantically Enabled Knowledge Technologies http://www.sekt-project.com/

[40] http://www.acemedia.org

[41] http://www.w3.org/2005/Incubator/mmsem/

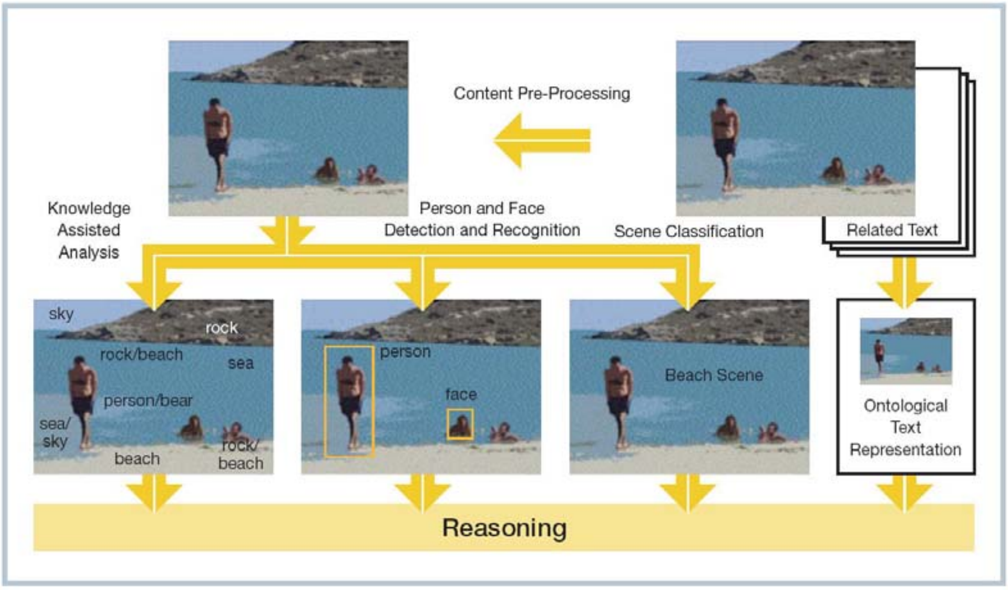

Figure 5. ***Automated Semantic Annotation in aceMedia***

2.4.2 Knowledge enhanced multimedia services

In aceMedia the main technological objectives are to discover and exploit knowledge inherent in multimedia content in order to make content more relevant to the user; to automate annotation at all levels (see Fig. 5) ; and to add functionality to ease content creation, transmission, search, access, consumption and re-use.

Users access multimedia content using a variety of devices, such as mobile phones and set-top-boxes, as well as via broadband cable or wireless to their PC. Through exploitation of Semantic Web tools, aceMedia has created a system which provides new and intuitive ways for users to enjoy multimedia content, such as intelligent search and retrieval, self-organising content, and self-adapting content. For example, using aceMedia's automatic metadata generation, a user can annotate content taken with her mobile phone, then seamlessly upload it to her PC where further automatic metadata generation takes place. aceMedia tools enables the content to be automatically organised into thematic categories, according to the user's preferences, and using extensions to DLNA/UPnP standards, the content can be automatically pushed to other users (as specified by the content owner) according to chosen rules. For example, our user might automatically receive new pictures of herself on her mobile phone or PC which were acquired and annotated on the device of one of her friends or family.

The aceMedia use case highlighted a number of future direction, issues and new challenges with respect to semantic multimedia content analysis and manipulation within a Semantic Web framework. Apart from the requirements with respect to formal uncertainty representations and more effective reasoning and management tools support, two dimensions of significant importance include:

- **cross-media analysis**, where additional requirements are posed due to the multimodality of knowledge considered, and their semantic modelling and integration, and

- **non-standard approaches to reasoning**, as purely deductive reasoning alone proves not sufficient

Other projects which can use the results of this co-operation: particularly K-Space[45], X-Media[46], BOEMIE[47] and MESH[48] constitute research consortiums working on the same topic. As, in the case of

[45] http://www.kpace-noe.net

[46] http://www.x-media-project.org

[47] http://www.boemie.org

[48] http://www.mesh-ip.eu

aceMedia, the main research directions focus on the exploitation of formal explicit knowledge and (possibly extended) inference services for the extraction of semantic descriptions from multimedia content. Additional aspects include among other scalability, logic programming and DL-based reasoning integration for non-standard inference support, and ontology evolution (Stoilos G. et al, 2005) (Petridis K. et al., 2006) (Dasiopoulou S. et al., 2007).

Another interesting reported multimedia experiment is MediaCaddy (Garg S. *et al.*, 2005) aiming at providing **movie or music recommendations** based on published online **critics, user experience and social networks**. Indeed, for the entertainment industry, traditional approaches to delivering meta-content about movies, music, TV shows, etc. were through reviews and articles that were done and published in traditional media such as newspapers, magazines and TV shows. With the introduction of the Internet, non-traditional forms of delivering entertainment started surfacing. The third quarter of 2003 in the U.S was the best ever for broadband penetration bringing such services as content on-demand and mobile multimedia. As of today more than 5000 movies and 2,500,000 songs are available on line. In the next couple of years this figure is expected to grow in leaps and bounds. With such a phenomenal rise in content over IP, a new need for secondary metacontent related to the movies/music emerged. Initially this was through movie reviews or music reviews published on web portals such as Yahoo, MSN and online magazine portals as well as entertainment sales sites such as Netflix.com and Amazon.com.

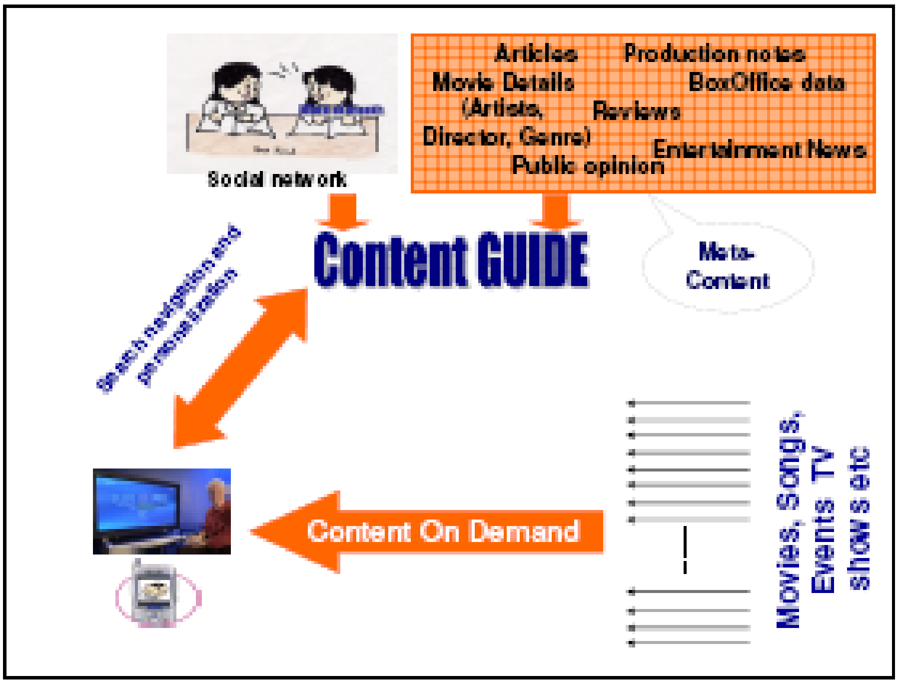

Figure 6. *Conceptual Model of Content Navigation System from the MediaCaddy project*

Most consumers today get information about media content primarily from reviews/articles in entertainment/news magazines, their social network of friends (one user recommends a song or movie to a friend), acquaintances and advertisements. In most of the cases, one or all of the above influence user's opinion about any content (s)he chooses to consume. In addition, a new breed of customizable meta-content portal has emerged, which specifically targets the entertainment industry. Examples of such portals include Rotten Tomatoes[49] and IMDB[50]. However, these services today are typically accessed via Web portals thereby limiting the interactions and access to the information for a user in a non-PC environment.

MediaCaddy is a recommendation and aggregation service built around a self-learning engine, which analyzes a click stream generated by user's interaction and actions with meta-content displayed through a UI. This meta-content (Music /Movies/ TV reviews/ article/ synopsis/ production notes) is accessed from

[49] http://www.rottentomatoes.com

[50] http://www.imdb.com/

multiple Internet sources and structured as an ontology using a semantic inferencing platform. Fig. 6 illustrates the conceptual model of MediaCaddy

This provides multiple benefits, both allowing for a uniform mechanism for aggregating disparate sources of content, and on the other hand, also allowing for complex queries to be executed in a timely and accurate manner. The platform allows this information to be accessed via Web Services APIs, making integration simpler with multiple devices and UI formats. Another feature that sets MediaCaddy apart is its ability to achieve a high level of personalization by analyzing content consumption behaviour in the user's personal Movie/Music Domain and his or her social network and using this information to generate music and movie recommendations.

2.5 Other prominent applications

Finally we list some excellent illustrations of the applications of Semantic Web technology, as selected from a worldwide competition[51] which offers participants the opportunity to show the best of the art.

- MultimediaN E-Culture demonstrator, is to demonstrate how novel semantic-web and presentation technologies can be deployed to provide better indexing and search support within large virtual collections of cultural heritage resources, 1st Prize 2006, http://e-culture.multimedian.nl/demo/search
- CONFOTO, Essen, Germany. CONFOTO is an online service which facilitates browsing, annotating and re-purposing of photo, conference, and people descriptions. 1st Prize 2005: http://www.confoto.org/
- FungalWeb, Concordia University, Canada. "Ontology, the Semantic Web, Intelligent Systems for Fungal Genomics".2nd Prize 2005: http://www.cs.concordia.ca/FungalWeb/
- Bibster – A semantics-based Bibliographic P2P system. http://bibster.semanticweb.org
- CS AKTive space – Semantic data integration. http://cs.aktivespace.org (Winner 2003 Semantic Web challenge)
- Flink: SemWeb for analysis of Social Networks. http://www.cs.vu.nl/~pmika (Winner 2004 Semantic Web challenge)
- Museum Finland: Sem Web for cultural portal. http://museosuomi.cs.helsinki.fi (2nd prize 2004 Semantic Web challenge)
- Also see Applications and Demos at W3C SWG BPD. http://esw.w3.org/mt/esw/archives/cat_applications_and_demos.html

3 Conclusions and Future Trends

In 2000, three prominent authors in the Semantic Web activity expounded in a seminal Scientific American paper (Berners-Lee T. *et al.*, 2001) the Semantic Web vision. In the time since then, the Semantic Web has become real. Currently, there are hundreds of millions of RDF triples, on tens of thousands of Web pages, and thousands of ontology pages have been published using RDF schema and OWL, with a growing level of industrial support. This very active support from industry is witnessed at worldwide key conference[52] very focused on the applications of the Semantic Web Technology. Indeed, about 100 talks on industry experience in testing and deploying the technology and about 30 technology showcases and 10 workshops or tutorials were actively followed by hundreds of attendees (300 at STC 05, 700 at STC 06, 730 at STC 07 and 210 at the 1st ESTC 2007) mostly from the industry.

However, the Semantic Web is still in its early days and there are many exciting innovations on the horizon.

[51] Annual Semantic Web applications challenge: http://challenge.semanticweb.org

[52] Semantic Technology Conference http://www.semantic-conference.com/ ; European Semantic Technology Conference http://www.estc2007.com/

A keynote speech[53] foresaw (Hendler J. & Lassila O., 2006) a "re-birth of AI" (or the end of the AI-winter) thanks to big-AI applications (Deep Blue, Mars Rover, Deep Space 1, Sachem-Usinor) and Web AI (IR, NLP, Machine Learning, Services, Grid, Agents, social nets) needed due to the tremendous amount of data continuously available on the Web and the emergence of new ways of doing things (loose coupling of distributed applications or services, new business models, etc.).

From 2000 to 2007, three major endeavours have paved the way for the future: DARPA, W3C and EU IST where DARPA and EU IST funded projects particularly were clearly forces towards production of recommendations to W3C (RDF-S, OWL, Rules, …), for fast adoption in industry. In the meantime, 2003 saw early government adoption and emerging corporate interest, in 2005 the emergence of commercial tools, lots of open source software and even good progress in the problem of scalability (tractable reasoning over 10 million triples has already been claimed by Oracle company!).

So, **a significant corporate activity is clearly noticeable today compared to 7 years ago**:

– Semantic (Web) technology companies are starting and growing: Cerebra, Siderean, SandPiper, SiberLogic, Ontology Works, Intellidimension, Intellisophic, TopQuadrant, Data Grid, Software AG, OntoText, etc.
– Semantic Web technology appears in the strategic plans of large corporations: Nokia, SAP AG, IBM, HP, Adobe, Cisco, Oracle, Sun, Vodaphone, Renault, AGFA etc.
– Outreach to industry is also demonstrated through a newly launched W3C initiative (2007): "Semantic Web Education and Outreach Interest Group - Case Studies and Use Cases". Case studies include descriptions of systems that have been deployed within an organization, and are now being used within a production environment[54].
– Government projects in and across agencies: US, EU, Japan, Korea, China, FAO, etc.
– Life sciences/pharma is an increasingly important market, e.g. the Health Care and Life Sciences Interest Group at W3C[55]
– Many open source tools are available: Kowari, RDFLib, Jena, Sesame, Protégé, SWOOP, Wilbur etc. see the W3C SWAD initiative[56]
– Related technologies are taking off: Google Base (taxonomies for resource descriptions), web 2.0 initiatives for mash-up applications, etc.
– Enterprise Web 2.0 can be the catalyst for a more collaborative business environment[57]. The BBC World Service had done a lot of work to try to create a more collaborative work environment. As it turned out, the BBC's internal forums, which only cost the company about 200 pounds, got the company to be more collaborative than the more formal initiatives did.

Then, it is also witnessed that adding a few semantics to current web applications - meaning "not harnessing the full picture at once but step by step" – gives a significant push in applications: richer metadata, data harvesting and visualization, web-based social network, digital asset management, scientific portals, tools for developers, and so gradually closing the semantic gap.

Semantic Web lessons: What has been learned from AI?

– Cross-breeding with AI succeeded, stand-alone AI did not!
– Tools are hard to sell (needed too much skill and education)
– Reasoners are a means, not an end (a key component but not the end)
– Knowledge engineering bottleneck (Ontology development and management)

Semantic Web lessons: What has been learned from the Web?

– The magic word: Distribute, interconnect and Share Roadmap!
- PC era [1980-1990] – autonomous computing and Ethernet
- Internet 1st generation [1990-2000] - Web 1.0), "read-only web", Web sites and Companies' portals
- Social Networks [2000-2010] - Web 2.0, corporate Knowledge Management and social nets

[53] SemWeb@5: Current status and Future Promise of the Semantic Web, James Hendler, Ora Lassila, STC 2006, 7 March 2006, San José, USA

[54] http://www.w3.org/2001/sw/sweo/public/UseCases/

[55] http://www.w3.org/2001/sw/hcls/

[56] Semantic Web Advanced Development for Europe http://www.w3.org/2001/sw/Europe/

[57] Forester, Erica Driver, October 2007

- Semantic Web [2007 – 2020] - Web 3.0 – merging social net with automated semantic web
- Web OS [2020-2030] - Web 4.0

However, it must be clear that there are still **Key technology locks** identified today that needs academic research and R&D investments for a full uptake of the technology *(Cuel et al., 2007)*:

Ontology and reasoning

– **The development of medium size to large ontologies is a challenging task**: e.g. modelling of business domains, unified and industry-grade methodology, best practices and guidelines, re-use of existing ontologies and simple tools to use ;
– **Automated update of ontologies and knowledge bases**: e.g. ontology maintenance by extraction of new knowledge facts and concept reasoning (abduction, learning), knowledge base population from legacy databases, data warehouse and data on the web, consistency checking ;
– **Ontologies interoperability**: overcome inevitable heterogeneity in spite of KR standards via e.g. automated mapping (concepts, properties, rules, graphs, …) in particular in the open context of the Web and the social Web (Web 2.0) ;
– **Scalability**: be capable to process business and real applications needs e.g. approximate reasoning, reasoning on inconsistent ontologies, ontology and reasoning modularization, distribution and cooperation ;
– **KR Expressivity and Tractability trade-off**: maintaining the just needed KR expressivity to avoid tractability and decidability issues, there are many open problems in this area. Look for reasoning algorithm optimizations (!), measure experimental complexity and lastly may be relax the completeness property ;
– **Rules - Logic Programming and KR:** moving towards a deeper and broader automation of business process intra- and inter-enterprise require the addition of semantic rules technology. E.g. Rules in communicating applications, Rules to describe / represent service process models, Rules for policies and contracting, etc. (see e.g. RuleML W3C[58])

Semantic Web Services and Sevices Oriented Computing (Papazoglou *et al.*, 2006),

– **Discovery**: Automated service discovery, reasoning on non functional parameters like QoS and cost ;
– **Composition**: business and industrial processes automated. I/O signature and behavioural composition (Inputs, Outputs, pre-conditions, effects and message protocols). Robustness in a versatile and inconsistent context. Composition driven by high level business needs (Business Rules) ;
– **Management**: web services supervision, self-healing service, self-optimizing service, self-configuring service, self-protecting service ;
– **Design tools**: Unified Design Principles for engineering service applications, Associating a service design methodology with standard software development and business process modelling techniques, Service governance, test and proof checks ;
– **Pilots and standard platforms**: the most prominent (2007)
- WSMX**[59]** (Fensel *et al.*, 2005) probably the most complete architecture to date, experimented on business cases and in transfer to OASIS.
- METEOR-S[60] SAWSDL: some running prototypes, industrial pilots and transfer to W3C (Sivashanmugam, 2003).
- OWLS (Ankolenkar, 2004).
- IRS III[61]
- SWSF[62]

In summary, the Semantic Web is “an interoperability technology”, “a architecture for interconnected communities and vocabularies” and “a set of interoperable standards for knowledge

---

[58] http://www.w3.org/2005/rules/wg/wiki/RuleML

[59] http://www.oasis-open.org/committees/semantic-ex/faq.php, http://www.wsmx.org
http://sourceforge.net/projects/wsmx

[60] http://lsdis.cs.uga.edu/projects/meteor-s/

[61] http://kmi.open.ac.uk/projects/irs/

[62] http://www.w3.org/Submission/SWSF/

exchange"[63]. Firstly, layman users facing the unmanageable growth of data and information, and secondly companies facing the huge amounts and volatility of data, applications and services, all require urgently automated means that master this growing complexity. In such de-facto context, no individual is able to identify knowledge patterns in their heads, no company (and employees!) is able to always shorter its products and service lifecycle and self adapt rapidly enough to survive.

The performance of semantic technologies clearly shows efficiency gain, effectiveness gain and strategic edge. Those facts are based on a survey[64] of about 200 business entities engaged in semantic technology R&D for development of products and services to deliver solutions and also recently witnesses at the ESTC 2007 industry oriented major event. From an academic and technological viewpoint, most things that have been predicted have happened - the semantic chasm is closing. Some things happened faster than anticipated like – triple store scaling, cooperation tools, enhanced SOA middleware, meta-data tagging tools, semantically-aware development environments and **last but not the least the unpredicted huge rise of web 2.0 user-oriented technology** – and others still need to be realized: ontologies are there (but very little interlinking and the need is huge especially in the healthcare domain), public information sources and public re-usable ontologies (as RDF, OWL, etc.), standard Rules (SWRL, WSML, etc.) and Logic Programming integration to Ontology languages, scalable and robust reasoners, technology transparency for the final user and the practitioners, and these technologies must mature into enterprise-class products, etc …. **Pervasive computing is just emerging**.

Acknowledgements

This work has been possible thanks to the three large European consortia REWERSE, Knowledge Web and aceMedia. Acknowledgements are also for the large gathering of international conferences mixing research results and prospects from academia and industry: ESWC, ISWC, ASWC, ICWS, WWW, STC etc. Lastly, credits go also directly to the numerous people, in research labs in academia and in industry who are contributing so strongly to make semantic technology a real momentum in industry.

IST-REWERSE is a NoE supported by the European Commission under contract FP6-506779 http://rewerse.net

IST-Knowledge Web is a NoE supported by the European Commission under contract FP6-507482 http://knowledgeweb.semanticweb.org

IST-aceMedia is an Integrated Project supported by the European Commission under contract FP6-001765. http://www.acemedia.org

References

Ankolenkar A., Paolucci M., Srinivasan N., and Sycara K., 'The owl services coalition, owl-s 1.1 beta release', Technical report, July 2004.

Antoniou G., Baldoni M., Baroglio C., Baumgartner R., Bry F., Eiter T., Henze N., Herzog M., May W., Patti V., Schindlauer R., Tompits H., Schaffert S. (2004). Reasoning Methods for Personalization on the Semantic Web. Annals of Mathematics, Computing & Telefinformatics, 2(1), 1-24.

Benatallah B., Hacid M. S., Léger A., Rey C., Toumani F. (2005). On automating Web services discovery. VLDB Journal 14(1):84-96.

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American Journal* (May 2001)

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., Inkingnehun S., Malandain G., Matsumoto J., Papadopoulos D., Pélégrini M., Richard N., Simon E. (2003). Neurobase: Management of distributed knowledge and data bases in neuroimaging. *Human Brain Mapping*, 19, 726-726, New-York, NY.

[63] ESTC 2007 Keynote speech from Susie Stephens (Oracle)

[64] Semantic Wave 2006, Part-1 Mills Davis

Berardi D., Calvanese D., De Giacomo G., Lenzerini M., and Mecella M., Automatic Service Composition based on Behavioral Descriptions. International Journal of Cooperative Information Systems, 14(4):333--376, 2005.

Billig A. and Sandkuhl K. (2002). Match-making based on semantic nets: The xml-based approach of baseweb. *1st workshop on XML-Technologien fur das Semantic Web*, 39–51.

Bonifacio M. and Molani A.. Managing Knowledge needs at Trenitalia. *Proceedings of the Second European Semantic Web Conference* ESWC 2005, Industry Forum proceedings. http://www.eswc2005.org/industryforum.html

Brunschweig B. and Rainaud J.F. (2005). Semantic Web applications for the Petroleum industry. *Proceedings of the 2nd European Semantic Web Conference*, ESWC 2005 Industry Forum http://www.eswc2005.org/industryforum.html

Cardoso J. and Sheth A., "Semantic e-workflow composition". Journal of Intelligent Information Systems, 21(3) :191–225, 2003. Technical description of the composition of a workflow.

Chan May K.S., Bishop Judith and Baresi Luciano, Survey and Comparisaon of Planning Techniques for Web Services Composition, Technical Report 2006.

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), 37-62.

Cordonnier E., Croci S., Laurent J.-F., Gibaud B. (2003). Interoperability and Medical Communication Using "Patient Envelope"-Based Secure Messaging. *Medical Informatics Europe Congress*, 95, 230-235.

Cuel R., Deleteil A., Louis V., Rizzi C. (2007), Knowledge Web Technology Roadmap "The Technology Roadmap of the Semantic Web", to appear early 2008.

Dasiopoulou S., Saathoof C., Mylonas Ph., Avrithis Y., Kompatsiaris Y., Staab S. (2007). Introducing Context and Reasoning in Visual Content Analysis: An Ontology-based Framework, in Semantic Multimedia and Ontologies: Theories and Applications, Paola Hobson, Ioannis Kompatsiaris (Editors), Springer-Verlag, 2007, to appear.

Description Logics for Information Integration, (2002) LNCS; Vol. 2408, Computational Logic: Logic Programming and Beyond, pp 41 – 60, Publisher Springer-Verlag London, UK

Dhamankar R., Lee Y., Doan A., Halevy A., Domingos P. (2004). iMAP: Discovering complex semantic matches between database schemas. *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 383 – 394.

Di Noia T., Di Sciascio E., Donini F. M., Mongiello M. (2003). A system for principled matchmaking in an electronic marketplace. *Proceedings of the 12th international conference on World Wide Web*, WWW 2003, 321–330.

Do H.H. and Rahm E. (2002). COMA - a system for flexible combination of schema matching approaches. In Proceedings of Very Large Databases VLDB 2002, 610–621.

Dou D., McDermott D., Qi P. (2005). Ontology translation on the Semantic Web. *Journal on Data Semantics*, 3360, 35–57.

Ehrig M. and Staab S. (2004). QOM: Quick ontology mapping. Third International Semantic Web Conference, ISWC 2004, LNCS 3298, 683–697.

Euzenat J. and Valtchev P. (2004). Similarity-based ontology alignment in OWL-lite. *Proceedings of European Conference on Artificial Intelligence* ECAI 2004, 333–337.

Euzenat J. and Shvaiko P. (2007). Ontology Matching, Springer-Verlag, ISBN: 3-540-49611-4, 2007.

Fensel, D., Kifer, M., de Bruijn, J., Domingue, J.: Web service modeling ontology (wsmo) submission, w3c member submission. (2005).

Garg S., Goswami A., Huylebroeck J., Jaganathan S., Mullan P. (2005). MediaCaddy - Semantic Web Based On-Demand Content Navigation System for Entertainment. *Proceedings of the 4th International Semantic Web Conference*, ISWC 2005. LNCS 3729, 858 – 871.

Giunchiglia F. and Shvaiko P. (2003). Semantic matching. *Knowledge Engineering Review Journal*, 18(3), 265–280.

Giunchiglia F., Shvaiko P., Yatskevich M. (2004). S-Match: an algorithm and an implementation of semantic matching. In Proceedings of the *First European Semantic Web Symposium*, ESWS2004, LNCS 3053, 61–75.

Guarino N., Masolo C., Vetere G. (1999). OntoSeek: Content-Based Access to the Web, *IEEE Intelligent System*, 14(3), 70--80, (May 1999).

Haarslev V., Moller R., Wessel M. (1999-2007). RACER: Semantic middleware for industrial projects based on RDF/OWL, a W3C Standard. http://www.sts.tu-harburg.de/~r.f.moeller/racer/

Halevy A., Answering Queries Using Views: A Survey, VLDB Journal, Vol. 10, Issue 4, December 2001.

Halevy A., Rajaraman A., Ordille J., Data Integration: The Teenage Years, 10-year best paper award, VLDB 2006.

Hasselwanter T., Kotinurmi, P., Moran M., Vitvar T., Zaremba M. (2006). WSMX: a Semantic Service Oriented Middleware for B2B Integration , In proceedings of the 4th International Conference on Service Oriented Computing , Springer-Verlag LNCS series, December, 2006, Chicago, USA.

Heinecke, J. and Cozannet, A. (2003). Ontology-Driven Information Retrieval. a proposal for multilingual user requests. *Workshop on Ontological Knowledge and Linguistic Coding at the 25th annual meeting of the German Linguistics*, Feb. 26-28, 2003.

Hibbard, J. (1997). Knowledge management-knowing what we know. *Information Week* (October 20).

Hull R., (1997), Managing Semantic Heterogeneity in Databases : A Theoretical Perspective, Tutorial at PODS 1997.

Hull R. and Su J., “Tools for composite Web services: A short Overview. ACM SIGMOD Record, Vol. 34, N°2, June 2005.

Lehtola, A., Heinecke, J., Bounsaythip, C. (2003). Intelligent Human Language Query Processing in mkbeem. Workshop on Ontologies and Multilinguality in User Interface. In the *Proceedings of Human Computer Interface International*, HCII 2003, 4 , 750-754.

McGuinness D. L., Fikes R., Rice J., Wilder S. (2000). An environment for merging and testing large ontologies. Proceedings of the *Seventh International Conference on Principles of Knowledge Representation and Reasoning* (KR2000), Breckenridge, Colorado, 483–493.

Mcllraith S.A., Son T.C., and Zeng H.. “Semantic web services". IEEE Intelligent Systems, Special Issue on the Semantic Web, Volume 16, pages 46–53, 2001.

Mena E., Kashyap V., Sheth A., Illarramendi A. (1996). Observer: An approach for query processing in global information systems based on interoperability between pre-existing ontologies. *Proceedings of the First International Conference on Cooperative Information Systems* CoopIS’96, 14–25.

Nixon L., Mochol M., Léger A., Paulus F., Rocuet L., Bonifacio M., Cuel R., Jarrar M., Verheyden P., Kompatsiaris Y., Papastathis V., Dasiopoulou S. & Gomez Pérez A. (2004). *Prototypical Business Use Cases*. (Technical report Deliverable 1.1.2), Knowledge Web IST-NoE.

Norheim D. and Fjellheim R. (2006). Knowledge management in the petroleum industry. *Proceedings of the 3rd European Semantic Web Conference*, ESWC 2006 Industry Forum. http://www.eswc2006.org/industry.html

Papazoglou M., Traverso P., Dustdar S. and Leymann F., Service-oriented computing research roadmap. Technical report, 2006.

Petridis K., Bloehdorn S., Saathoff C., Simou N., Dasiopoulou S., Tzouvaras V., Handschuh S., Avrithis Y., Kompatsiaris I., Staab S. (2006). Knowledge Representation and Semantic Annotation of Multimedia Content, IEEE Proceedings on Vision Image and Signal Processing, Special issue on Knowledge-Based Digital Media Processing, Vol. 153, No. 3, pp. 255-262, June 2006.

Pistore M., Roberti P., and Traverso P.. “Process-Level Composition of Executable Web Services: "On-thefly" Versus "Once-for-all" Composition”. The Semantic Web: Research and Applications. Proceedings of ESWC 2005, Heraklion, Crete, Greece, May/June 2005. LNCS 3532, Springer Verlag, 2005.

Preguica N., Shapiro M., Matheson C. 2003). Semantics-based reconciliation for collaborative and mobile environments. In Proccedings of the Eleventh International Conference on Cooperative Information Systems CoopIS 2003, LNCS 2888, 38-55.

Petrash G. (1996). Managing knowledge assets for value. Proceedings of *the Knowledge-Based Leadership Conference*, Linkage Inc., Boston, MA, October 1996.

Petrini J. and Risch T. (2004). Processing queries over RDF views of wrapped relational databases. In *Proceedings of the 1st International workshop on Wrapper Techniques for Legacy Systems*, WRAP 2004, Delft, Holland, 2004.

Rahm E. and Bernstein P. (2001). A survey of approaches to automatic schema matching. *Very Large Databases Journal*, 10(4), 334–350, (Dec. 2001).

Dumitru Roman, Uwe Keller, Holger Lausen, Jos de Bruijn, Ruben Lara, Michael Stollberg, Axel Polleres, Cristina Feier, Christoph Bussler, Dieter Fensel: Web Service Modeling Ontology, Applied Ontology Journal, 2005, Volume 1, Number 1, Pages 77-106

Rosenbloom ST, Miller RA, Johnson KB. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *Journal of the American Medical Informatics Association* 2006

Semantic Web Case Studies and Best Practices for eBusiness (SWCASE) at ISWC 2005 on line version http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-155/

Shvaiko P., Giunchiglia F., Pinheiro da Silva P., McGuinness D. L. (2005). Web explanations for semantic heterogeneity discovery. In *Proceedings of the Second European Semantic Web Conference* ESWC 2005, 303-317.

Shvaiko P., Léger A., Paulus F., Rocuet L., Nixon L., Mochol M., Kompatsiaris Y., Papastathis V., & Dasiopoulou S. (2004). *Knowledge Processing Requirements Analysis*. Technical report Deliverable D 1.1.3, Knowledge Web IST-NoE.

Shvaiko P. and Euzenat J. (2005). A survey of schema-based matching approaches. *Journal on Data Semantics* (JoDS) 4, 146–171.

Spackman K. A., Dionne R., Mays E., Weis J., « Role grouping as an Extension to the Description Logic of Ontylog Motivated by Concept Modeling in SNOMED », *Proceedings of the AMIA Annual Symposium 2002*, San Antonio, Texas, p. 712-6, November, 9-13, 2002.

Stamou, Giorgos, Jacco van Ossenbruggen, Jeff Pan and Guss Schreiber. (2006). Multimedia annotations on the Semantic Web. IEEE Multimedia, 13(1):86--90, January-March 2006.

Stoilos G., Stamou G., Tzouvaras V., Pan J.Z., Horrocks I. (2005). A Fuzzy Description Logic for Multimedia Knowledge Representation, in Proc. of the International Workshop on Multimedia and the Semantic Web, pp 12-19, ESWC 2005, Heraklion, Grece, June 2005 .

Traverso P. and Pistore M. (2004). Automated composition of Semantic Web services into executable processes. In Proceedings of the Third International Semantic Web Conference ISWC 2004, 380–394, 2004.

Velegrakis Y., Miller R. J., Mylopoulos J. (2005). Representing and querying data transforma-tions. *Proceedings of the 21st International Conference on Data Engineering* ICDE 2005, 81-92.

Vitvar T., Zaremba M., Moran M. (2007a). Dynamic Service Discovery through Meta-Interactions with Service Providers , In Proceedings of the 4th European Semantic Web Conference (ESWC2007) , Springer-Verlag LNCS series, June, 2007, Innsbruck, Austria.

Vitvar, T. Zaremba M., Moran M., Haller A., Kotinurmi P. (2007b). Semantic SOA to Promote Integration of Heterogeneous B2B Services , The 4th IEEE Conference on Enterprise Computing, E-Commerce and E-Services (EEE07) , IEEE Computer Society, July, 2007, Tokyo, Japan.

Zhong J., Zhu H., Li J., Yu Y. (2002). Conceptual graph matching for semantic search. In *Proceedings of the 10th International Conference on Computational Science*, 2393 (2002), 92-106

Zyl J., and 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.

Wiederhold G. (1992). Mediators in the architecture of future information systems. *IEEE Computer*, 25(3), 38-49.

Wiig K. (1997). Knowledge management: where did it come from and where will it go? *Journal of Expert Systems with Applications*, 13(1), 1–14.

Additional Reading

- The Semantic Web: research and Applications LNCS series:

  LNCS 2342 (ISWC 2002), LNCS 2870 (ISWC 2003), LNCS 3053 (ESWS 2004), LNCS 3298 (ISWC 2004), LNCS 3532 (ESWS 2005), LNCS 4011 (ESWC 2006), LNCS4273 (ISWC 2006), LNCS 4519 (ESWC 2007), LNCS 4825 (ISWC 2007).

- Journal of Web semantics (Elsevier)
- Annual Semantic Web applications challenge: http://challenge.semanticweb.org
- W3C http://www.w3.org/2001/sw/