Applicability assessment of Semantic Web technologies

Valentina Janev *, Sanja Vraneš

Mihajlo Pupin Institute, Volgina 15, 11060 Belgrade, Serbia

ABSTRACT

The Semantic Web is one of the fastest developing fields within the Information and Communication Technology sector and, as such, under constant examination by scientists and IT professionals. This article aims to provide a better understanding of the applicability of Semantic Web tools and technologies in practice. This aim will be achieved by surveying the recommended and emerging W3C standards, presenting an overview of the state-of-the-art in the Semantic Web research in the European Union, analysing the W3C collection of Case studies and Use Cases, and discussing the extent of adoption of Semantic Web technologies. The overall technology maturity level assessment has shown that Semantic Web technologies are finding their ways into real-world applications, and that, rather than being merely a fashionable research issue, the Semantic Web, slowly but surely, becomes our reality.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

New technologies, such as Semantic Web (SW) technologies, are usually subjected to experimentation, refinement, and increasingly realistic and exhaustive testing. This kind of information gathering, which aims to look beyond the immediately obvious and at analysing the ramifications of a given technology in as wide-ranging and far-sighted a manner as possible, is known as *technology assessment* (Braun, 1998). Technology assessment is usually based on different forecasting methods including extrapolation, expert opinion (the Delphi method) and modelling, cost-benefit analysis, cross-impact analysis, and others.

The term "Semantic Web" refers to the World Wide Web Consortium's vision of the Web of linked data (called also the Web of Data) as "…an extension of the current Web in which information is given a well-defined meaning, better enabling computers and people to work in cooperation" (Berners-Lee, Hendler, & Lassila, 2001). There are few international standardization organizations (associations or consortia) relevant for assessment of information technologies such as IEEE-SA (see The Institute of Electrical and Electronics Engineers Standards Association, http://standards.ieee.org/), OASIS (see The Organization for the Advancement of Structured Information Standards, http://www.oasis-open.org/), OMG (see The Object Management Group, http://www.omg.org/) and W3C (see The World Wide Web Consortium, http://www.w3.org/).

In order to determine the achievements in the Semantic Web field, especially its maturity status and the adoption of these technologies by the industry, as well as to predict the future development of SW technologies, we have studied and analyzed a substantial amount of various sources. These range from deliverables from projects financed by the European Commission within its 6th and 7th Framework Programmes (EC FP6 + FP7), through scientific papers from prestigious international journals and conferences, Semantic Web technologies and tools from the industry top vendors and open source communities

---

* Corresponding author. Tel.: +381 11 2774 024; fax: +381 11 2775 835.
E-mail addresses: Valentina.Janev@impcs.com (V. Janev), Sanja.Vranes@institutepupin.com (S. Vraneš).

(Janev, 2008), to Web resources of the W3C standardization body. This paper clarifies the applicability of Semantic Web technologies and identifies the major trends in the Semantic Web field, both from researchers' point of view and from the perspective of early adopters.

The article is organized as follows. After identifying the work related to Semantic Web technologies assessment in Section 2, we point to existing and emerging SW standards and discuss their adoption for developing new applications and for meeting different technical and application challenges. Using the W3C collection of Case Studies and Use Cases, we infer relations between the SW technologies and the gained benefits. Section 4 considers the state-of-the-art of SW technologies from several viewpoints such as development of SW languages, availability of data, content and ontologies on the Semantic Web, data and service integration, while in Section 5 we point to open issues such as stability of SW languages, interoperability issues, and scalability of SW applications.

2. Related work to Semantic Web technologies assessment

In literature and in practice, SW tools and technologies are named by using different keywords: ontology design/management/maintenance tools, semantic data management and integration platforms (Ahmad & Colomb, 2007), RDF triple storage systems, web services, SOA middleware platforms, semantic annotation tools (Reeve & Han, 2005), content indexing and categorization tools, semantic search and information retrieval technologies (Andrews, 2009; Crestani, Lalmas, Van Rijsbergen, & Campbell, 1998; Mangold, 2007), NLP, linguistic analysis and text mining algorithms, collaboration and other social networking technologies (Correndo & Alani, 2007; Gootzit, Phifer, Valdes, & Knipp, 2009), knowledge visualization/presentation technologies, ontology mediated portals, ontological querying/inference engines, rule-based engines, ontology learning methods (Gómez-Pérez & Manzano-Macho, 2003), ontology reasoners, etc. Exploring the business value of semantic technologies provided by 50 commercial companies (Davis, Allemang, & Coyne, 2004) identified the following four major functions: discover, acquire, & create semantic metadata; represent, organize, integrate, and inter-operate meanings & resources; reason, interpret, infer, & answer using semantics; and provision, present, & communicate, and act using semantics. Analysing the functionalities of more than 50 SW tools (see the Web4WeB repository of the Sixth Framework Program "Web Technologies for the West Balkan Countries", at www.web4web.org/portal/Semantic_Web_Tools) provided by 30 different commercial vendors and open source communities, Janev and Vraneš (Janev, 2008) established the following classification of key semantic technology segments: semantic modelling and development, semantic annotation, semantic data management and integration, semantic search and retrieval, semantic collaboration including portal technologies, learning and reasoning. According to this analysis, vendors have made recognizable progress towards the specification and acceptance of semantic standards, but still lack efficient reasoning support which is crucial for realization of the Semantic Web vision.

The literature review has showed that most of the scientific studies usually analyze a single aspect of semantic technologies, and that just a few studies provide the key trends in the Semantic Web field (see, e.g., Davis et al., 2004; Cali et al., 2005; Cardoso, 2007; d'Aquin et al., 2008) or investigate the adoption of semantic technologies in various scientific and industrial domains. While some surveys take an industrial economic approach towards semantic technologies (Davis et al., 2004; Provost, 2008), others concentrate on the evaluation of specific technologies and their potential success within specific application areas (Davis et al., 2004; Cuel, Delteil, Louis, & Rizzi, 2007; Pellegrini et al., 2009). In the "The Technology Roadmap of the Semantic Web" white paper the authors deliver a comprehensive analysis maturity, applicability and adoption of the SW technologies (Cuel et al., 2007) based on a survey conducted with selected experts from industry and academia. Using the Gartner Hype Cycle Curve (Linden & Fenn, 2003), they present expert opinions, i.e. results of estimation of years to mainstream adoption for different SW technologies and applications both from the researchers' and the business point of view. They found out that: "research community considers that developments from the past 10 years have resulted in some tools and standards, which are reliable and mature enough to be transferred to industry and successfully integrated into SW applications. The developers' community, however, is not yet fully aware of the availability of such tools, which, consequently, has to be promoted further, together with the innovative functionalities they can provide to software applications". In the "The Semantic Web Awareness Barometer" (Pellegrini & et al., 2009), the authors use statistical methods to summarize the results of the survey and the Chi-square test to compare expert opinions from industry and academia. The analysis indicated that "the expectations in Semantic Web technologies were very high in both groups. Generally both groups believe that organizational culture is not yet ready for the Semantic Web. Additionally, application-oriented participants believe that Semantic Web technologies are too complex. On the contrary, research-oriented participants believe that the lack of success stories, a lack of quality of available software, the problem of quantifying the benefits, the costs of implementation and the general heterogeneity of information are the biggest obstacles to the application of Semantic Web technologies."

3. Analysis of the W3C collection of Case Studies and Use Cases

3.1. Semantic Web standards

In the time since the SW's initial conception (Berners-Lee et al., 2001), the Semantic Web Activity Group of the World Wide Web Consortium (W3C) has accepted numerous Web technologies as standards or recommendations. In an attempt

to structure and relate these technologies, Tim Berners-Lee presented several versions of the SW architecture that layered them into a so-called stack of increasingly expressive languages for metadata specification (www.w3.org/2001/sw/). To represent information on the Web in the form of a graph, the SW uses the Resource Description Framework (RDF) as a general–purpose language. To ensure interoperability between applications that exchange machine-understandable information, RDF describes information in terms of objects (resources) and the relations between them via the RDF Schema, which serves as a meta-language or vocabulary to define properties and classes of RDF resources. To avoid ambiguity, the RDF Schema uses uniform resource identifier (URI) references for naming. Furthermore, since every Semantic Web resource has a unique URI, it is possible to establish links among existing models, re-use existing models and build new models upon them, thus creating a network of ontologies and enabling development of large-scale Semantic Web applications.

After standardizing RDF and the ontology layer (OWL), the SW research community's primary efforts in the past few years have been focused on standardizing technologies for SW services and provisioning technologies and tools that enhance interoperability and availability of content on the Semantic Web. This latter effort involves the development of rule languages, rule exchange languages, and engines that enhance reasoning, improvement of ontology languages, invention of new knowledge representation formalisms and elaboration of methods and tools for publishing linked data on the Web (Bizer, Heath, & Berners-Lee, 2009). Table 1 surveys the maturity status of technologies that were closely investigated in this research.

To summarize, some of the key benefits of ontology languages that are usually declarative compared to procedural languages include re-use and interoperability of models, flexibility in open and dynamic systems, consistency and quality checking across models.

3.2. The W3C collection of Semantic Web Case Studies and Use Cases

Most of the existing survey studies are descriptive and do not provide any empirical data about the usability of semantic technologies in different application domains, or data about the achieved benefits. This was the reason for choosing the W3C collection of Semantic Web Case Studies and Use Cases to find relationships between variables such as the enterprise area of activities, the application area of SW technologies, the SW technologies used, and the benefits of SW technologies. Retrieved from http://www.w3.org/2001/sw/sweo/public/UseCases/ on August 11, 2010, the collection includes 29 descriptions of systems used within a production environment (Case studies) and 13 examples of built prototype systems (Use cases). The employed method was *cross-tabulation*, also known as *contingency table analysis* which is often adopted for exploring categorical (nominal measurement scale) survey data. A cross-tabulation is a two- (or more) dimensional table that records the frequency of respondents with the specific characteristics described in the table cells. Our primary objective was to identify

Table 1
An overview of mature and emerging Semantic Web technologies.

Technology	W3C status	Date of publication
Mature SW technologies		
RDF Resource Description Framework	Recommendation	2004-02-10
RDF Schema	Recommendation	2004-02-10
SPARQL Query Language for RDF	Recommendation	2008-01-15
OWL 2 Web Ontology Language	Recommendation	2009-10-27
SKOS Simple Knowledge Organization System Reference	Recommendation	2009-08-18
RIF Rule Interchange Format	Recommendation	2010-06-22
RIF Framework for Logic Dialects	Recommendation	2010-06-22
WSDL Web Services Description Language	Recommendation	2007-06-26
SAWSDL Semantic Annotations for WSDL and XML Schema	Recommendation	2007-08-28
RDFa in XHTML	Recommendation	2008-10-14
SML Service Modelling Language	Recommendation	2009-05-12
POWDER Protocol for Web Description Resources	Recommendation	2009-09-01
GRRDL Gleaning Resource Descriptions from Dialects of Languages	Recommendation	2007-09-11
Emerging SW technologies		
SPARQL 1.1 Query Language	Working draft	2010-06-01
SPARQL 1.1.Update		
OWL 2 RL in RIF	group note	2010-06-22
OWL-S Semantic Markup for Web Services	Member submission	2004-11-22
RDFa Core 1.1	Working draft	2010-08-03
RIF In RDF	Working draft	2010-06-22
WSMO Web Service Modelling Ontology	Member submission	2005-06-03
WSML Web Service Modelling Language	Member submission	2005-06-03
SWRL Semantic Web Rule Language	Member submission	2004-12-21
Combining OWL and RuleML		
FOL RuleML: The First-Order Logic Web Language	Member submission	2005-04-11
SIOC (Semantically-Interlinked Online Communities) Core Ontology	Member submission	2007-06-12
Web Security Context: User Interface Guidelines	Proposed recommendation	2010-06-22
RuleML	Under development	
R2RML (RDB2RDF) Mapping Language	Under development	

the application areas as well as W3C SW technologies applied in a particular application domain and to indicate the benefits end-user organizations gain by utilizing SW technologies.

3.3. Application areas of the Semantic Web technologies

The scientific literature lacks a survey analysis of business applications of semantic technologies. In a technology assessment report based on findings from 35 semantic technology early adopters from many different verticals (Davis et al., 2004), the authors have identified the following application areas: infrastructure and integration, managing risk, customer-facing services, output management, smart products and services, design and manufacture, research, input management, supplier facing processes, intelligence and security. Analyzing the time to mainstream adoption of semantic technologies in different business applications from practitioners' point of view, European researchers (Cuel et al., 2007) found that the most mature business application areas were bioinformatics, knowledge management and e-learning. Business areas with less than 5 years of mainstream adoption were B2B (Business-to-Business) and B2C (Business-to-Customer), while the duration to mainstream adoption for e-government applications was estimated to be more than five years. Looking at the business areas of early adopters registered in the W3C collection of Semantic Web Case Studies and Use Cases, we found that 36% of the early adopters were public institutions and that 47% of them implemented e-government principles using semantic technologies. Further, 12% of them were applications in the health sector and exactly the same share had the applications in the IT industry, and then follow with less than 10% the applications in the telecommunications sector, life sciences (pharmaceutical industry), broadcasting services and library services, while only 2% were finance applications (Table 2).

Analyzing the W3C collection of Semantic Web Case Studies and Use Cases and looking at the technical problems solved by using semantic technologies we have compiled a list of the most mature application areas of semantic technologies, presented in Table 3 Semantic technologies are by far most often used for data integration and for improving the search. However, according to these statistics, they have untapped potential for dynamic customization/composition of services and, hence, for building flexible architectures.

Table 2
Relating Company activity area to Semantic Web technologies' application domains.

Company activity area		SW application areas						
		28 – Data integration	19 – Improved search	11 – Content discovery	10 – Semantic annotation	4 – Social network	2 – NL interface	2 – Service integration
	Coverage	67%	45%	26%	24%	10%	5%	5%
Public institution	36%	9	8	3	4	1		
e-Government	17%	2	5	2				
Health care	12%	5	2					
IT industry	12%	3	1		2	2	1	1
Telecommunication	10%	2	1	1	1		1	1
Life sciences	7%	3	1	1				
Energy	7%	3	1		1	1		
Broadcasting	7%	3	3	2				
Library	7%	2	2		2			
Automotive	5%	1						

Table 3
Semantic Web technologies' application domains.

Domain	Definition	Coverage percentages
Data integration	Process of retrieving, merging and storing of data originated in heterogeneous, expanding set of corporate/public data sources using ontologies and web services	67
Semantic Search	Methods (i.e. faceted-based) that aim to augment and improve traditional search results by using not just words, but concepts and logical relationships	45
Semantic content discovery	Using taxonomies and ontologies for describing different types of content and for dynamic linking between content items	26
Semantic annotation	Process of assigning unambiguous meaning to resources (e.g. assigning types to single terms or semantic relations to pairs of terms) in order to enable a more efficient discovery mechanism	24
Social networks	Technologies that are used to help people track, discover, and share content around topics they are interested in, thus forming social networks	10
Natural language interfaces	Methods and technologies that are used to provide natural, human-like interaction with the computer	5
Service integration	Adding semantics to Web services with the aims to augment service integration, i.e. discovery, composition, ranking, selection and mediation of services	5

Table 4
Relating Application area to SW technologies used.

SW technologies		SW application areas						
		28 – Data integration	19 – Improved search	11 – Content discovery	10 – Semantic annotation	4 – Social network	2 – NL interface	2 – Service integration
	Coverage	67%	45%	26%	24%	10%	5%	5%
34-RDF(S)	81%	21	13	10	8	4	2	1
27 – In-house vocabularies	64%	17	12	9	4		2	
20 – Public vocabularies	48%	15	10	8	5	2		
21 – OWL	50%	14	11	6	6	1	1	1
15 – SPARQL	36%	13	7	3	4	4	1	1
6 – Rules	14%	4		1	1		1	
4 – SKOS	10%	2	3	3	2			
3 – RDFa	7%	2	2	2	1			
2 – GRRDL	5%	1	1	1				
1 – WSMO	2%	1						1

Table 4 establishes relationships between the most frequently used Semantic Web technologies and application domains. The first column shows the number of items (SW applications) that use one specific technology, while the second column indicates the coverage percentages. This distribution shows that most of the applications (81%) are based on RDF(S) and half of them use OWL models for knowledge representation. This can be explained with the expressivity/performance trade-off (see Table 1. *Complexity of Tractable Fragments* on the following Web site http://www.w3.org/Submission/2006/SUBM-owl11-tractable-20061219/). OWL models are more expressive, and hence require more sophisticated reasoning algorithms (e.g. Concept Subsumption or Instance Checking in RDFS is PTIME while in OWL DL is NEXPTIME complete). In order to achieve better performance, developers tend to use less expressive languages and languages that are not purely declarative (e.g. F-logic). Further, SW applications are based on in-house vocabularies, half of which incorporate public vocabularies, and one tenth of which use the SKOS standard for representing and publishing classification schemes, thesauri, and controlled vocabularies as RDF graphs. It is interesting to note that investigation of the use of the SPARQL standard query language for RDF shows that it is more often used for data integration, i.e. for syntactic matching of different knowledge schemas, than for querying and retrieval. The search and content discovery functions rely on the public and in-house vocabularies, as well as on the SKOS models that, similarly to Knowledge Organization Systems in library and information sciences, are very suitable for retrieval and navigation tasks. Considering the use of rule languages, we can conclude that they are used in just a small number of applications in public institutions, health care and IT. It is encouraging to observe, however, that, besides the mature SW technologies (RDF and OWL), technologies such as OWL-S and WSMO that are still in the process of standardization are being considered for service integration.

3.4. Benefits from utilizing Semantic Web technologies

Looking at the W3C Early Adopter collection and analysing the reported benefits from utilizing Semantic Web technologies, we have found out that the companies achieved data share and re-use (57%), improved search (57%), incremental modelling (26%), explicit content relation (24%), identifying new relationships (17%), dynamic content generation (14%), personalization (10%), open model (12%), rapid response to change (10%), reduced time to market (5%), and automation (5%).

In order to provide further evidence which links SW technologies with gained benefits, we have analyzed two subsets of case studies: (I) a set of entries (25 out of 42) whose main application was "search and content discovery" and (II) a set of entries (30 out of 42) whose main application was "data integration". Both sets count 39 out of 42 or 93% of all entries. The results of the analysis are summarized in Fig. 1, where the left-hand-side illustrates the applied SW technologies in realization of the semantic solutions for both sets of semantic applications, while the right-hand-side illustrates the gained benefits. The results can be interpreted as follows. In the whole W3C dataset, 14% of "data integration" solutions use rule languages, hence rule languages are more suitable for data integration than for implementing the search function. Both subsets show that SW technologies are very suitable for data sharing and reusing (reported by 67% of applications in the second dataset), as well as for knowledge search (reported by 84% of applications in the first dataset). Using semantic annotation and automatic ontology learning techniques, unstructured text documents can be processed and semantic relations extracted, thus making the content relationships explicit (reported by 12–27% of applications) and, hence, machine-processable. Design based on public vocabularies and public datasets leads to open model architectures that enhance interoperability. Faceted navigation techniques and semantic relations shorten the search time, improve the relevance of search results, and deliver high-quality search services. Reasoning services can be used to identify new relationships (reported by 20–24% of applications) and dynamic content generation (reported by 20% of applications).
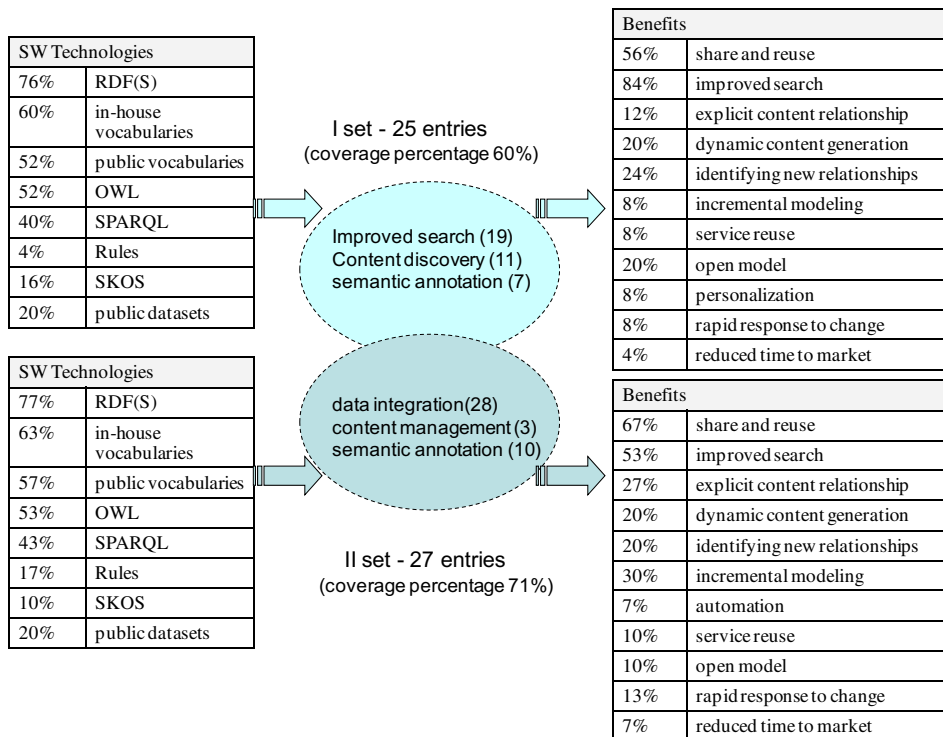
SW Technologies	
76%	RDF(S)
60%	in-house vocabularies
52%	public vocabularies
52%	OWL
40%	SPARQL
4%	Rules
16%	SKOS
20%	public datasets

I set - 25 entries
(coverage percentage 60%)

Improved search (19)
Content discovery (11)
semantic annotation (7)

Benefits	
56%	share and reuse
84%	improved search
12%	explicit content relationship
20%	dynamic content generation
24%	identifying new relationships
8%	incremental modeling
8%	service reuse
20%	open model
8%	personalization
8%	rapid response to change
4%	reduced time to market

SW Technologies	
77%	RDF(S)
63%	in-house vocabularies
57%	public vocabularies
53%	OWL
43%	SPARQL
17%	Rules
10%	SKOS
20%	public datasets

data integration(28)
content management (3)
semantic annotation (10)

II set - 27 entries
(coverage percentage 71%)

Benefits	
67%	share and reuse
53%	improved search
27%	explicit content relationship
20%	dynamic content generation
20%	identifying new relationships
30%	incremental modeling
7%	automation
10%	service reuse
10%	open model
13%	rapid response to change
7%	reduced time to market

Fig. 1. Relating Semantic Web benefits to SW technologies used.

4. State-of-the-art in the Semantic Web research in the European Union

The EU framework programmes (http://cordis.europa.eu/) – together with EUREKA (A Europe-wide Network for Market-Oriented R&D and Innovation, http://www.eurekanetwork.org/) and COST (an intergovernmental framework for European Cooperation in Science and Technology, http://www.cost.esf.org/) – is one of the three pillars of the joint European research initiatives. According to the Europe's Digital Competitiveness Report – Main achievements of the i2010 strategy (2005–2009), "Europe remains a global force in advanced information and communication technologies industry (European Commission, 2009)".

Using the established classification of Semantic Web technologies' application domains (see Table 3), herein, we present the results of the analysis of the main contributions of European research community to the Semantic Web development. As an entry point for this survey we have used the Web4WeB repository of EU funded Semantic Web related projects (http://www.web4web.org/portal/Research_Projects), that besides the basic information provides links to the organizations involved in these projects and mirrors the most important "research", "industry" and "education" deliverables. At present, the knowledge base stores information about 58 projects including REWERSE, KnowledgeWeb, SUPER, NeOn, DIP, NEPO-MUK, TRIPCOM, SWING, TAO, SemanticGov, Service Web 3.0, ACTIVE, INSEMTIVES, LarKC, OKKAM, KIWI, LarKC, SOA4ALL, IKS, MATURE, SEALS and others.

4.1. SW methodologies

Ontology building exhibits a structural and logical complexity comparable to the production of software artifacts. Most mature methodologies, i.e. methodologies with the greatest adoption among ontologists are those developed fifteen years ago such as Methontology (13.7%), followed by On-To-Knowledge methodology (7.4%), and Mike Uschold and Martin King's method (4.2%) (Cardoso, 2007). These methodologies aimat building corporate semantic applications. Our analysis of methodologies used in recently terminated and ongoing EU FP6 and FP7 projects indicated that each project developed its own methodology. The evaluation of the SUPER Semantic Business Process Management Methodology (Fantini et al., 2007) for example, showed that "the methodology is almost complete and accepted as a general guideline and that problem areas are mainly related to the current tool support along the whole SBPM lifecycle" (SUPER, http://www.ip-super.org/res/Deliverables/M18/D2.3.pdf). Therefore, we conclude that unlike conventional software development, in which most methodologies have matured and are usually facilitated by well-defined modelling languages and computer-aided software engineering tools, SW methodologies (especially the ones for development of large scale applications) are still in their inception phase. The most promising methodologies and tools for large scale applications were delivered or are still under development in the framework of the SemanticGov, NeOn, TripCom, LarKC and other projects.

4.2. Development of SW languages

Our analysis of semantic languages and formalisms leveraged in European research projects has revealed several important insights. The research efforts in this field have been focused on the development of many existing SW languages, especially to WSMO, WSML, RIF, and SAWSDL, within the framework of the Data, Information, and Process Integration with Semantic Web Services project (DIP, http://projects.kmi.open.ac.uk/dip/) and REWERSE Rule Markup Language (R2ML) within the framework of Reasoning on the Web with Rules and Semantics project (REWERSE, http://rewerse.net/). It has been also revealed that besides RDFS and OWL, researchers tend to use WSML and its variants that differ in expressiveness as well as in their underlying logical formalism (Description Logics, Logic Programming, First-Order Logic or F-Logic). In the framework of the TripCom project, which aimed to further develop the machine-to-machine Web service communication, researchers have proposed a new Triple Reasoning and Rule Entailment Engine language (TRREE, http://www.tripcom.org/docs/del/D2.3.pdf) while software developers have integrated it into the existing OWLIM semantic repositories (see http://www.ontotext.com/owlim/). In the framework of the LarKC project (http://www.larkc.eu), which has the objective to develop a platform for Web-scale reasoning, researchers propose more lightweight representation formalism (e.g., the L2 language in Fischer, Unel, Bishop, and Fensel (2010)) that takes into consideration the expressivity, complexity, and practical feasibility of current knowledge representation languages, as well as to further elaborate the Continuous SPARQL (or simply C-SPARQL (Barbieri, Braga, Ceri, Della Valle, & Grossniklaus, 2010)). Finally, within the framework of the Lifecycle support for networked ontologies project (NeOn, http://www.neon-project.org/), researchers have developed new formalisms for distributed knowledge representation and reasoning, e.g. the Integrated Distributed Description Logics (IDDL, http://www.neon-project.org/web-content/images/Publications/neon_2009_d144.pdf); implemented new ontology modularization tools; suggested new approach for efficient reasoning with large data volumes and conjunctive query answering with PTIME Data Complexity, and others.

4.3. Data integration, service integration and publishing linked data on the Web

Ontologies are the backbone of the Semantic Web, a semantic-aware version of the World Wide Web. Together with commonly used vocabularies, they provide reference models for data and process integration as well as knowledge sharing between systems and people in ways that facilitate machine reasoning and inference. The European Commission's FP6 and FP7 projects have produced and made publicly available many high-quality domain ontologies in areas such as life science, business organization, e-government, and so on. Considering the most often used languages and formalisms for building ontologies, we found out that some business process and geo-spatial ontologies were based on the WSML-Flight knowledge representation language; TripCom's Triple Space Ontology has three variants (RDFS, OWL-Lite, and WSML-Flight); and the NeOn project chose the Meta Object Facility (MOF) as a model-driven integration framework for definition, manipulation, and integration of modules of the Networked Ontology Metamodel. The Networked Ontology Metamodel can be grounded on OWL, SWRL, F-logic, and other ontology languages.

As our analysis previously showed, ontology-based approaches to data and service integration have already been well established in corporate settings. During the last five years, research efforts have been dedicated to extending the conventional Web, and making the existing content available on the Semantic Web as well. Aimed at publishing structured data in the RDF format there have been several open community projects coordinated by EU experts such as DBpedia (http://dbpedia.org/), Semantic Web Dog Food (http://data.semanticweb.org/) and SensorMasher (http://sensormasher.deri.org/), as well as open source tools developed by EU researchers, e.g. the D2RQ server (Bizer & Cyganiak, 2006), and the Triplify toolkit (Auer, Dietzold, Lehmann, Hellmann, & Aumueller, 2009). Taking the advantage of the Open Linked Data initiative (http://linkeddata.org/), large ontologies have been defined, for example, for the Health Care and Life Sciences domains we can mention MesH (http://www.nlm.nih.gov/mesh/), Disease (http://diseaseontology.sourceforge.net/), EHR RM (http://trajano.us.es/~isabel/EHR/), and GO (http://www.geneontology.org/). As a result of the Open Government Initiative (Alani et al., 2008; Auer, Lehmann, & Hellmann, 2009), the government domain has emerged as one of the promising application area for Semantic Web technologies (e.g. see http://data.gov.uk/).

4.4. Semantic search and content discovery

Considering the applicability of SW technologies for search and content discovery, we can distinguish three different use scenarios of utilization of SW technologies:

- meaningful search in domain-specific applications, e.g. discovery and retrieval of geo-spatial information or indexing, cataloging, or searching for biomedical and health-related information and documents (Doms & Schroeder, 2005),
- visual search, navigation and querying of large public interlinked datasets within the Web of Data (e.g. see tools such as the SW indexing engine Sindice at http://www.sindice.com/, the Semantic Information MAshup SIG.MA at http://sig.ma/ or the VisiNav system at http://visinav.deri.org/),
- hybrid search as a search method that supports both document and knowledge retrieval via the flexible combination of ontology-based search and keyword-based matching (Bhagdev, Chapman, Ciravegna, Lanfranchi, & Petrelli, 2008).

As was previously stated, SW approaches proved their advantages in domain-specific applications on the search side. Considering the knowledge extraction side, the focus of research has moved from semi-automatic annotation (Haslhofe, Momen, Gay, & Simon, 2010) and text-based topic discovery to semantic knowledge extraction from multimodal data sources (Ahmed & et al., 2010).

Comparing the SW search and metadata service engines such as Sindice, SWSE (http://swse.deri.org/), Swoogle (http://swoogle.umbc.edu/), and Watson (http://kmi-web05.open.ac.uk/WatsonWUI/) with the traditional ones such as Google and Yahoo, we conclude that the traditional search engines mainly search for unlimited topics in unstructured contents, while semantic search engines search for limited/unlimited topics in structured/unstructured contents.

4.5. Semantic annotation

Semantic annotations are machine-processable tags that are associated with a word, phrase, text, picture or video with the aim to describe that portion of data uniquely. SW annotation systems utilize text/video processing approaches to assign semantics, expose the annotations as linked data and to create links to contextually relevant resources on the Web (e.g. DBpedia). The semantics of words differ from domain to domain, and therefore efficient semantic annotation requires domain learning or manual domain specification in a form of taxonomy. Taking into consideration the type of annotation method used, Reeve and Han (2005) distinguish two primary categories of semantic annotation platforms (Pattern-based and Machine Learning-based). Building minimally-supervised and domain-independent methods for the population/construction of ontologies (named also ontology learning) is still an open issue (Manzano-Macho, Gómez-Pérez, & Borrajoy, 2008).

Worth mentioning in this context is the European Commission's support for several cross-national projects aimed at creating resources for the community, e.g. the CALBC project ("Collaborative Annotation of a Large-Scale Biomedical Corpus", http://www.calbc.eu/) or Europeana (http://www.europeana.eu), a platform that links to millions of digital items (texts, images, videos, sounds) in cultural organizations all over Europe.

4.6. Natural language interfaces

Natural language search interfaces (NLIs) hide the formality of an ontology-based knowledge base as well as the executable query language from end-users by offering an intuitive and familiar way of query formulation. In a study with 48 real-world users and four interfaces featuring four different query languages, Kaufmann revealed that "NLIs to ontology-based knowledge bases can be considered to be useful for casual or occasional end-users" (Kaufmann & Bernstein, in press). Studying the human involvement in the process of semantic content creation Siorpaes and Simperl (2010) came to conclusion that "Interacting with semantic technologies today requires specific skills and expertise on subjects which are not part of the mainstream IT knowledge portfolio". These findings lead to the conclusion that semantic technologies community has to pay more attention to the design of semantic applications.

4.7. Social networks

Social network technologies such as wiki-based systems and social tagging systems aim at interactive information sharing, facilitating interoperability, user-centered design, and improving the collaboration on the World Wide Web. The basic features of social networking sites are profiles, friends' listings and commenting, messaging, discussion forums, blogging, media uploading and sharing. From the Semantic Web perspective, one of the limitations of the social network technologies is that metadata that is added to the Web content is based on freely chosen keywords (folksonomies) instead of a controlled vocabulary thus producing ambiguity in the meaning of words or phrases on the Web. Therefore, research efforts during the last several years have been devoted to bridging the gap between the Web 2.0 and Semantic Web paradigms and to developing and implementing SW tools and technologies into Social Web applications. Semantic MediaWiki (http://semantic-mediawiki.org), for example, extends the MediaWiki Collaborative Knowledge Management System. Twarql (http://twarql.sf.net) is an infrastructure translating microblog posts from Twitter (http://twitter.com/) as Linked Open Data in real-time and enabling expressive queries and flexible analysis of microblog data for sensemaking tasks (Mendes, Passant, & Kapanipathi, 2010).

5. Future development of Semantic Web technologies

Although many Semantic Web related technologies have emerged or have been elaborated in the last few years and have already been adopted in corporate environments, a lot still has to be done until the Berners-Lee's vision of the Semantic Web becomes a reality (Bizer et al., 2009; Shadbolt, Berners-Lee, & Hall, 2006).

Investigating the adoption of the SW technologies by enterprises, we (Janev, 2008) identified the following major open issues: scalability and run-time support, interoperability between different knowledge organization schemas, data synchronization between OWL/RDF based knowledge bases and the traditional persistence mechanisms, migration from traditional to semantic-enabled technologies. Herein, we will discuss the future development of SW technologies from the following

perspectives: (i) availability of content, (ii) interoperability, (iii) standardization and stability of SW languages, (iv) scalability, and (v) large-scale adoption.

5.1. Availability of content

W3C recommended several technologies including GRRDL, RDFa, and microformats aimed at extending the Web with structured data. The process of online conversion of the existing unstructured contents on the web into a format understandable by computers is not a trivial, and not generally solvable task. However, we have to note here that majority of content on the web is generated dynamically, e.g. exported from RDBMS.

When the content stored in RDBMS is exported into RDFS, the problem of consistency and synchronization appears, because the RDF store should be updated each time the RDBMS is updated. Therefore, it is a recent trend to use the RDBMS as a SW endpoint and SPARQL + SPARQL/Update for ontology-based read and write access to relational data.

5.2. Interoperability issue

One of the most challenging and important tasks of the SW ontology engineering is the integration of ontologies with the purpose of building a common ontology for all Web sources and consumers in a domain. The available ontologies often exhibit different conceptualizations of similar or overlapping domains, thus leading to the interoperability problem. The problem could be overcome by detecting semantic relations between concepts, properties or instances of two ontologies, i.e. ontology matching. Due to the increasing number of methods available for schema matching/ontology integration, the Ontology Alignment Evaluation Initiative (see http://oaei.ontologymatching.org/) was started with the aim to compare systems and algorithms on the same basis and to allow anyone to draw conclusions about the best matching strategies. The recent and future trends in overcoming schema heterogeneity between linked semantic repositories, i.e. ontology matching consider using reasoning languages (e.g. Distributed Description Logics) to reason about ontology alignments in distributed environments.

5.3. Standardization and stability of languages

While the W3C makes extensive efforts to define and standardize the upper layers of the W3C SW architecture model that refers to logic, inference, and reasoning, the research communities come out with new SW languages (as was presented above). Reasoning is a distinctive feature of the Semantic Web. Currently, the advanced reasoning tools and technologies are mainly delivered by open source communities, while the contemporary commercial SW development frameworks offer integration with few years old reasoning engines such as Pellet, KAON2, or Jess.

5.4. Scalability of SW applications

Scalability is one of the key SW issues that relates to large ontology creation and maintenance, semantic metadata extraction of massive and heterogeneous content and inference mechanisms (Sheth & Ramakrishnan, 2003). The scalability issue was identified very early in the SW research and adequately addressed. However, despite the huge number of SW applications today, advanced SW technologies such as reasoning under open-world assumptions are hardly applicable in real-time on web scale. The recent efforts toward development of parallel inference engines for scalable distributed reasoning (see WebPIE at http://www.few.vu.nl/~jui200/webpie.html) give promising results.

5.5. Large-scale adoption

Many constructive critics argued that "large-scale adoption of semantic technologies is still to come (Siorpaes & Simperl, 2010)" and that "the central deficiency of the Semantic Web is their static model of knowledge (ontologies), which implies static and predefined meaning of web-content (Pohjola, 2010)". Thus, we come to the problem of widespread adoption of SW technologies in situations where existing Web technologies already proved useful. In our opinion, additional investments are needed to mature SW technologies, especially, to optimize the querying and reasoning strategies. Concepts and methods already proven in information retrieval (e.g. the MapReduce framework introduced by Google (Dean & Ghemawat, 2010)) and data base processing (e.g. query adaptive result caching, view maintenance and query subsumption (Auer, 2008)) should be adopted to improve the performance of distributed systems and achieve scalable distributed reasoning. Instead of competing, the Semantic Web should integrate features from other similar research communities, i.e. the Social Web and the Pragmatic Web (http://www.pragmaticweb.info/). While the Semantic Web promotes "cooperation through shared models of knowledge", the Pragmatic Web "does not rely merely on shared models of knowledge, but also on shared ways of socially cultivating the ways of representing knowledge (Pohjola, 2010)". Thus, the relevant topics of future investigation and development include ontology negotiations, the integration of contextual ontologies, and developing pragmatic patterns for defining issues such as communication, information and tasks.

6. Conclusion

The study presented in this article aims to answer the question "Is the Semantic Web merely a fashionable research topic or rather our reality and future?" It analyses the current status and trends in the Semantic Web and discusses the adoption of Semantic Web technologies in practice.

The results obtained in this article show that semantic-based technologies have been steadily increasing their relevance in recent years in both the research and business worlds. W3C, together with universities and IT research organizations, and in cooperation with the major software companies and governmental agencies have already accepted many specifications, guidelines, protocols, software, and tools which are the basis for the realization of the Semantic Web vision. Innovative enterprises, interested in catching new opportunities from the Semantic Web, and also developing new business models, are involved in research projects and are introducing semantic technologies that facilitate data integration and interoperability, as well as improve search and content discovery. Considering the benefits from introducing SW applications, the analysed early adopters from the W3C collection of Case Studies and Use Cases prove that SW technologies are useful for sharing and reusing data, improving search and establishing explicit content relations. To summarize, based upon the overall analysis we have performed, we can conclude that SW technologies are finding their ways into applications, and that rather than being another research project, the Semantic Web becomes a reality.

References

Ahmad, M. N., & Colomb, R. M. (2007). Overview of ontology servers research. *Webology, 4*(2).

Ahmed, A. et al (2010). Structured literature image finder: Parsing text and figures in biomedical literature. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web, 8*(2-3), 151–154.

Alani, H. et al (2008). Building a pragmatic semantic web. *IEEE Intelligent Systems, 2008*(May/June), 61–68.

Andrews, W. (2009). Magic Quadrant for Information Access Technology (2009). Gartner RAS Core Research Note G00169927 from September 2, 2009. Gartner, Inc. <http://www.gartner.com/technology/media-products/reprints/microsoft/vol7/article2/article2.html> Retrieved 15.09.10.

Auer, S. (2008). Methods and applications of the social Semantic Web. In S. Vraneš (Ed.), *Semantic Web and/or Web 2.0: Competitive or complementary?* (pp 100–128). Belgrade, Serbia: Academic Mind.

Auer, S., Lehmann, J., & Hellmann, S. (2009). LinkedGeoData: Adding a spatial dimension to the web of data. In *Proceedings of the international Semantic Web conference* 2009 (pp. 731–746).

Auer, S., Dietzold, S., Lehmann, J., & Hellmann, S., Aumueller, D. (2009). Triplify: Light-weight linked data publication from relational databases. In *Proceedings of the 18th international conference on World Wide Web* (pp. 621–630).

Barbieri, D., Braga, D., Ceri, S., Della Valle, E., & Grossniklaus, M. (2010). Stream reasoning: Where we got so far. In S. Ceri, E. Della Valle, J. Hendler, & Z. Huang (Eds.), *Proceedings of the 4th workshop on new forms of reasoning for the Semantic Web: Scalable & dynamic* (pp. 1–7). <http://www.nefors10.larkc.eu/>.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. Scientific American. <http://www.sciam.com/article.cfm?id=the-semantic-web>.

Bhagdev, R., Chapman, S., Ciravegna, F., Lanfranchi, V., & Petrelli, D. (2008). Hybrid search: Effectively combining keywords and semantic searches. In *European Semantic Web Conference – ESWC 2008, Tenerife, Spain*.

Bizer, C., & Cyganiak, R. (2006). D2R server – Publishing relational databases on the Semantic Web (Poster). Poster at the 5th international Semantic Web conference, Athens, USA, November 2006.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems, 5*(3), 1–22.

Braun, E. (1998). *Technology in context: Technology assessment for managers* (p. 23). London: Routledge.

Cali, A., et al. (2005). State of the art survey. TONES. D.01. FP6-7603 Thinking OntolgiES.

Cardoso, J. (2007). The Semantic Web vision: Where are we? *IEEE Intelligent Systems*(September/October), 22–26.

Correndo, G., & Alani, H. (2007). Survey of tools for collaborative knowledge construction and sharing. In *Workshop on collective intelligence on Semantic Web (CISW 2007)*, 2–5 November 2007, Fremont, CA, USA.

Crestani, F., Lalmas, M., Van Rijsbergen, C. J., & Campbell, I. (1998). Is this document relevant? Probably: A survey of probabilistic models in information retrieval. *ACM Computing Surveys, 30*(4), 528–552.

Cuel, R., Delteil, A., Louis, V., & Rizzi, C. (2007). The technology roadmap of the Semantic Web. KnowledgeWeb White Paper. FP6-507482. <http://www.knowledgeweb.semanticweb.org/o2i/menu/KWTR-whitepaper-43-final.pdf> Retrieved 22.10.09.

d'Aquin, M. et al (2008). Toward a new generation of Semantic Web applications. *IEEE Intelligent Systems, 23*(3), 20–28.

Davis, M., Allemang, D., & Coyne, R. (2004). Evaluation and market report. IST Project 2001-33052 WonderWeb: Ontology infrastructure for the Semantic Web.

Dean, J., & Ghemawat, S. (2010). System and method for efficient large-scale data processing. United States Patent 7,650,331.

Doms, A., & Schroeder, M. (2005). GoPubMed: Exploring PubMed with the GeneOntology. *Nucleic Acid Research, 33*(Web Server Issue), W783–W786.

European Commission (2009). Europe's digital competitiveness report – Main achievements of the i2010 strategy 2005–2009. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Brussels, COM (2009) 390 from 4.8.2009.

Fantini, P., et al. (2007). Semantic business process life cycle. SUPER D2.2 IST Project-026850. <http://www.ip-super.org/res/Deliverables/M12/D2.2.pdf> Retrieved 22.10.09.

Fischer, F., Unel, G., Bishop, B., & Fensel, D. (2010). Towards a scalable, pragmatic knowledge representation language for the Web. In A. Pnueli, I. Virbitskaite, & A. Voronkov (Eds.), Perspectives of systems informatics. Lecture notes in computer science, 5947, 124–134.

Gómez-Pérez, A., & Manzano-Macho, M. (2003). A survey of ontology learning methods and techniques. OntoWeb D.1.5. IST Project IST-2000-29243.

Gootzit, D., Phifer, G., Valdes, R., & Knipp, E. (2009). Magic quadrant for horizontal portal products. Gartner RAS core research note G00170709 from September 17, 2009. Gartner, Inc. <http://www.gartner.com/technology/media-products/reprints/microsoft/vol10/article2/article2.html> Retrieved 15.09.10.

Haslhofe, B., Momen, E., Gay, M., & Simon, R. (2010). Augmenting Europeana content with linked data resources. In A. Paschke, N. Henze, T. Pellegrini (Eds.), Proceedings of I-SEMANTICS 2010. ACM international conference proceedings series.

Janev, V. (2008). A comparison and critical assessment of cutting edge Semantic Web products and technologies. In S. Vraneš (Ed.), *Semantic Web and/or Web 2.0: Competitive or complementary?* (pp 148–165). Belgrade, Serbia: Academic Mind.

Kaufmann, E., (2009). Talking to the Semantic Web: Natural language query interfaces for casual end-users. PhD Thesis at the Faculty of Economics, Business Administration and Information Technology of the University of Zurich. <http://www.ifi.uzh.ch/pax/uploads/pdf/publication/1384/Kaufmann_2007.pdf> Retrieved 18.11.10.

Linden, A., & Fenn, J. (2003). Understanding Gartner's hype cycles. Gartner research R-20-1971. Gartner, Inc. <www.paulohenrique.com-a.googlepages.com/3606402_gg_understanding_gartners_hy.pdf> Retrieved 15.09.10.

Mangold, C. (2007). A survey and classification of semantic search approaches. *International Journal of Metadata, Semantics and Ontology, 2*(1), 23–31.

Manzano-Macho, D., Gómez-Pérez, A., & Borrajoy, D. (2008). Unsupervised and domain independent ontology learning. Combining heterogeneous sources of evidence. In *Proceedings of the sixth international conference on language resources and evaluation (LREC'08)*. Marrakech.

Mendes, P. N., Passant, A., & Kapanipathi, P. (2010). Twarql: Tapping into the wisdom of the crowd. In A. Paschke, N. Henze, T. Pellegrini (Eds.), Proceedings of I-SEMANTICS 2010. ACM international conference proceedings series.

Pellegrini et al (2009). Semantic Web awareness barometer 2009 – Comparing research- and application-oriented approaches to social software and the Semantic Web. In A. Paschke, H. Weigand, W. Behrendt, K. Tochtermann, & T. Pellegrini (Eds.), *Proceedings of I-KNOW '09 and I-SEMANTICS '09*. Graz, Austria: J.UCS.

Pohjola, P. (2010). The pragmatic web: Some key issues. In A. Paschke, N. Henze, T. Pellegrini (Eds.), Proceedings of I-SEMANTICS 2010. ACM international conference proceedings series.

Provost, D. (2008). On the cusp: A global review of the Semantic Web industry. <http://www.scribd.com/doc/6379635/Semantic-Web-Industry-Review-sept2008> Retrieved 22.10.09.

Reeve, L., & Han, H. (2005). Survey of semantic annotation platforms. In *Proceedings of the 20th annual ACM symposium on applied computing, web technologies and applications track*, Santa Fe, New Mexico.

Shadbolt, N., Berners-Lee, T., & Hall, W. (2006). The Semantic Web revisited. *IEEE Intelligent Systems, 21*(3), 96–101.

Sheth, A., & Ramakrishnan, C. (2003). Semantic (Web) technology in action: Ontology driven information systems for search, integration and analysis. *IEEE Data Engineering Bulletin, Special Issue on Making the Semantic Web Real*(December), 40–48.

Siorpaes, K., & Simperl, E. (2010). Human intelligence in the process of semantic content creation. *World Wide Web WWW Journal, 13*(1), 33–59.