Jorge Cardoso, “**The Semantic Web Vision: Where are We**?” IEEE Intelligent Systems, September/October 2007, pp.22-26, 2007.

The Semantic Web Vision: Where are We?

Jorge Cardoso

Department of Mathematics and Engineering
University of Madeira, 9000-390, Funchal, Portugal
jcardoso@uma.pt

Abstract. The full vision of the Semantic Web has yet to be fully accomplished, but there has been considerable progress in the development and use of standards, languages, technologies and applications. While mainstream adoption is still five to ten years away, there is an imperative need for a survey that will paint a clear picture for practitioners describing the “semantic” road that is being followed. The aim of this article is to present a snapshot that can capture key trends in the Semantic Web, such as application domains, tools, systems, languages and techniques being used, and a projection on when organizations will put their full-blown systems into production.

Keywords: I.2.4.d, I.2.4.i, J.8

1. INTRODUCTION

In the past 6 years our community has seen continuous, sustained growth in the deployment of Semantic Web inventiveness in large and small organizations of many types. Much effort has been made towards achieving the vision of a machine-readable World Wide Web—the Semantic Web. Many researchers and practitioners are already starting to believe that a new Web is emerging based on the ongoing research and developments. Some industries, and its main players, are shifting from a “wait-and-see” approach (Cardoso, Miller et al. 2005) to the real-world deployment of applications that will give them added value and a competitive advantage. To see how the Semantic Web is becoming real, one need only look to companies such as Oracle, Vodafone, Amazon.com, Adobe, Yahoo, and Google. All of these organizations are working for a smarter Web and some of their implementations are already available on the Web. For example, Oracle has introduced the industry’s first RDF (Decker, Melnik et al., 2000) management platform, targeting application areas such as life sciences (Stephens, Morales et al. 2006), data and content integration, enterprise application integration, and supply chain integration. Oracle is already working on extending the platform with OWL support. Vodafone, a leading mobile phone company, has used RDF to describe and search ring tones, games and pictures in their Web site. As a result, the page viewed per download has decreased by 50% and revenues have risen by 20%. Many more successful examples could be enumerated.

We have witnessed a significant evolution of standards as improvements and innovations allow the delivery of more complex, more sophisticated and more far-reaching semantic applications. For the Semantic Web vision to become reality in everyday life, it is indispensable at this stage to present a snapshot that will capture certain key trends in the Semantic Web, the current developments and determine how researchers and practitioners are using and interrelating semantic technologies. Therefore, the results of a survey are presented here, so that we may keep a finger on the pulse of our community and demonstrate the variability and dynamism of the work being done on the Semantic Web, by looking at the picture that this survey paints for us.

2. SURVEY OVERVIEW

The aim of the survey is to give an account of current Semantic Web practices. The findings reported in this article are based on 627 surveys that were filled in and conducted within a two month period. The data was collected from 12 December 2006 to 26 January 2007. The participants were requested to answer 14 questions related to particular aspects of the Semantic Web and its technologies. The study covered close-ended questions. The participants were allowed to choose from multiple choice answers. Out of the 14 questions, 8 contained an option for "other" to be filled in. The survey was carried out electronically and an e-mail was sent to several mailing lists and discussion groups to request volunteers. The e-mail was sent to the following entities: DBWorld, SemanticWeb googlegroups, comp.ai.nat-lang, comp.ai, Music Ontology Specification googlegroups, comp.lang.lisp, and semantic-web@w3.org. Additionally, we sent approximately 40 personal invitations to people involved in academic and industrial research.

3. SURVEY DESIGN

The survey can be divided into five categories: Demographics, Tools and Languages, Ontology, Ontology Size, and Production. The **demographics** category had three questions intended to describe the most important facts about each participant that answered the survey. Namely, we asked in which type of organization they were working, their primary role, and the number of years of professional experience. The second category, **tools and languages**, was composed of three questions and focused explicitly on the tools used to build ontologies and the ontology languages used. The third part of the survey, **ontology**, contained 4 questions. The interviewees indicated for which industry or domain they are developing ontologies, which methodology they use to develop their ontologies, and why they use ontologies. The last question assessed the use of mapping, aligning, merging, and integrating ontologies. The fourth category, **ontology size**, contained 3 questions all related to the size of the ontologies developed. Participants were asked how many concepts had the smallest, a typical, and the biggest ontology in their organization. The last part of the survey, **production**, comprised only one question which focused on the assessment of the estimated timeframe to put the developed ontologies and associated systems into production.

Organization – We were impressed by both the size of the sample – the number of people who responded – and the distribution of the sample among academia and industry, and geographical locations. These elements all suggest that the data will be representative of the broad scope of Semantic Web projects today. Each respondent was asked in what type of organization they worked. Sixty-six percent of the respondents identified themselves as affiliated to academia, 18.4% to industry, and 16% worked in academia and industry (see Figure 1).

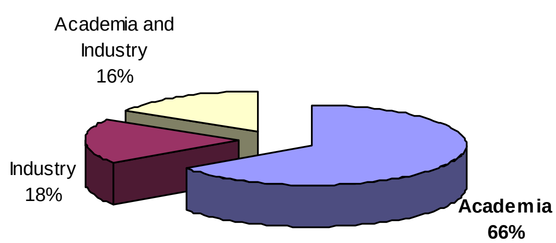

Figure 1. Respondent's organization type

The nature of the respondents' organization means that this survey tilted slightly toward an academia perspective, as opposed to an industrial or commercial perspective. We believe that this reflects the market in general. We were particularly impressed by the number of respondents that work actively in academia and industry. This may indicate a strong technology transfer from universities to companies in the next few years. There seems to be an emergent call to identify research results from the Semantic Web for potential commercial interests and to develop strategies for how best to exploit the Semantic Web in industry.

Participant Role – Each respondent was asked to describe his role within his organization. Almost twenty-two percent of the respondents identified themselves as researchers. Roughly equal numbers identified themselves as professors (12.5%), knowledge engineers (12%), programmers (12%), project manager (11.1%), or system analysts and designer (11.1%). Only 9.2% identified themselves as architects and 5% as students. More than five percent identified themselves as belonging to another category (e.g. consultant, CTO, Business Analyst). (See Figure 2.)

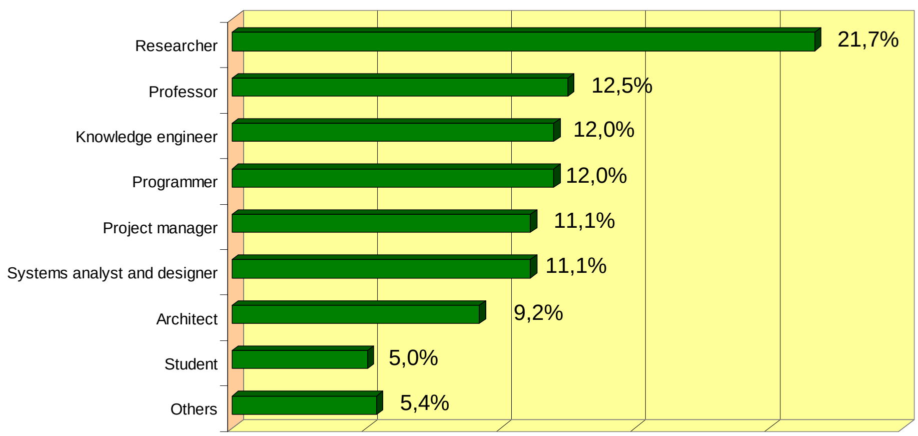

Figure 2. The role of respondents in their organizations.

The nature of the respondents' role demonstrates that there is a good sample distribution of roles in this survey which includes architects, programmers, systems analysts and professors. Nevertheless, the survey tilted slightly toward a research perspective as seen in Figure 2.

Professional Experience – Each respondent was asked the number of years of professional experience they had had in their role. Only a small fraction of the participants (6.8%) had less than one year of experience. Most had more than 6 years of experience in their role (35%). The number of participants with experience of between 1-2 (21.5%), 3-4 (23.1%), and 5-6 (13.6%) years was also significant; they comprised 58.2%. This diversity suggests that the data will be representative of the broad scope of professionals with different experience (Figure 3).

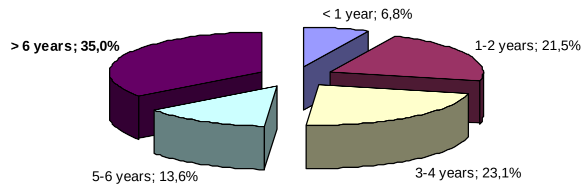

Figure 3. How many years of professional experience respondents had in their role?

Ontology Editors. We asked respondents to tell us which ontology editors they were currently using in their organizations, and their responses have been summarized in Figure 4. The editor most frequently cited was Protégé with a market share of 68.2%. The results clearly point out that Protégé is ahead of all other editors. Approximately equal numbers of respondents use SWOOP (13.6%), OntoEdit (12.2%) (OntoEdit is now called OntoStudio), and Altova SemanticWorks (10.3%). Surprisingly, 10.3% of users make use of simple text editors, such as emacs and vi, to create their ontologies. Other editors with some statistic expressions included OilEd (7.3%), OntoStudio (5.5%), IsaViz (4.9%), WebODE (3.7%), OntoBuilder (3.7%), WSMO Studio (2.8%), TopBraid Composer (1.8%), and pOWL (1.6%). Other editors that are in use include ORM and SemTalk. A good survey of the most popular ontology editors can be found in (Escórcio and Cardoso, 07).

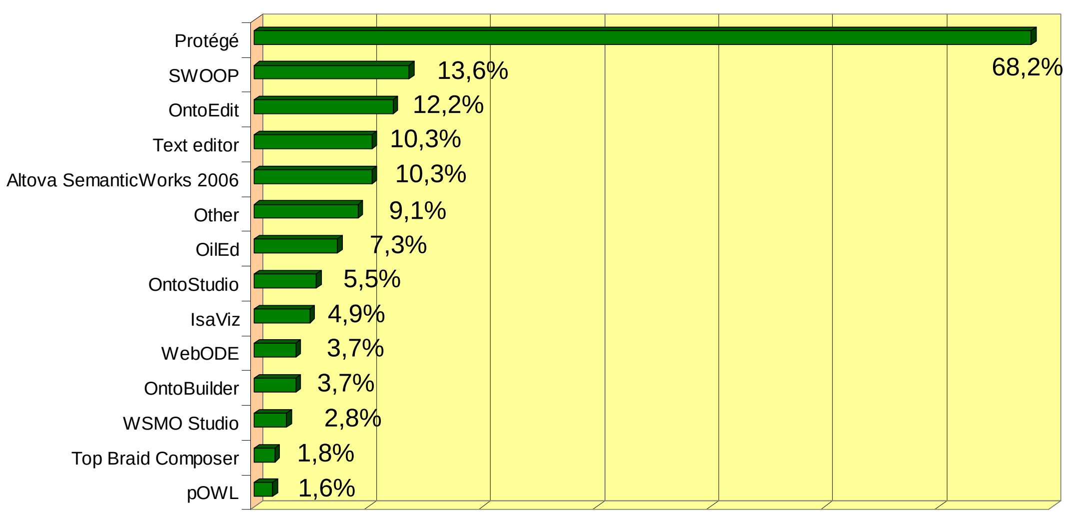

Figure 4. Respondents' use of ontology editors

Ontology Languages. Several ontology (Chandrasekaran, Josephson, et al., 1999) languages have been developed during the last few years. In 2002, Gomez-Perez and Corcho presented a study on *Ontology Languages for the Semantic Web*. The languages studied – XOL, SHOE, OML, RDF(S), OIL, DAML+OIL – were considered the most promising back then and it was thought that they would surely become ontology languages in the context of the Semantic Web (Gomez-Perez and Corcho, 2002). Our study revealed something somewhat different. In fact, XOL (0.9%), SHOE (1.9%), and OML (0%) languages show extremely low adoption among ontologists. On the other hand, the Gomez-Perez and Corcho study was correct in identifying RDF(S) and DAML+OIL as potential languages for the semantic Web. RDF(S) and DAML+OIL have a penetration rate higher than 64% and 12%, respectively. The language with the strongest impact in the Semantic Web is without a doubt OWL (which is derived from DAML+OIL and builds upon the Resource

Description Framework). More than 75% of ontologists have selected this language to develop their ontologies. Curiously, Description Logic and FLogic are also being used with a penetration rate of 17% and 11.8%, respectively. A recent language, WSML (Web Service Modeling Language), has also gained some popularity (3.7%). Approximately an equal numbers of respondents use Ontolingua/KIF (2.6%), Common Logic (2.6), Semantic Net (2.2%), SHOE (1.9%), OKBC (1.9%), and CycL (1.7%). Less popular languages include XOL, OCML, LOOM, CanonML, Topic Maps, SKOS, and XCML.

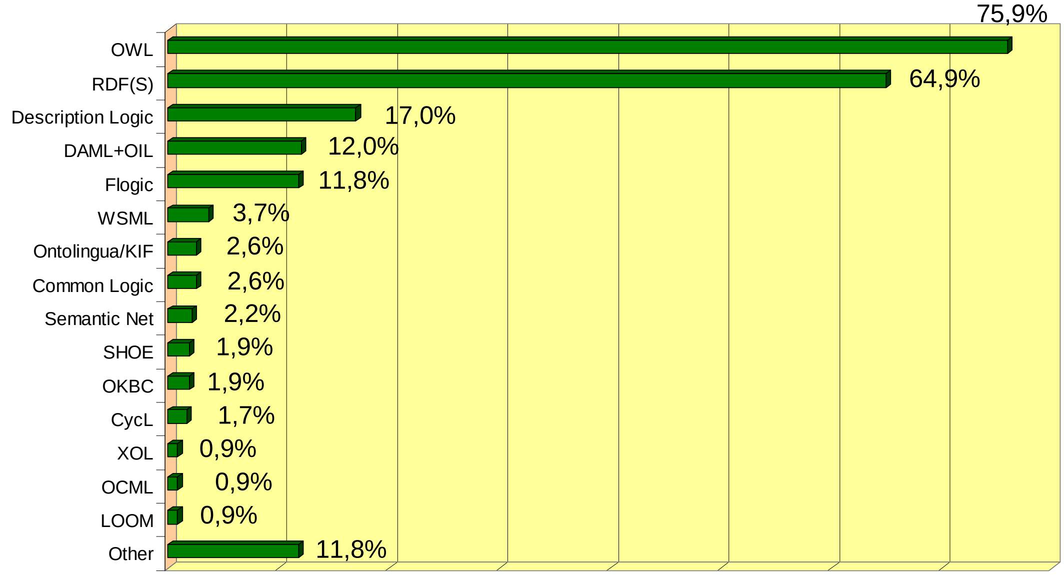

Figure 5. Ontology languages currently used by users.

Reasoning Engines. The use of inference engines in the Semantic Web allows applications to inquire as to why a particular conclusion has been reached (inference engines, also called reasoners, are software applications that derive new facts or associations from existing information). For example, consider an ontology that describes relationships between programming languages. Let us assume that Sara wants to search for a document in the Web about an imperative programming language. Her enhanced browser tries to search for a Web page with the progLangType property with the value 'imperativeLanguage'. A particular Web site advertises its content type as 'Ada'. These cannot be matched as keywords or even using a thesaurus, since 'imperativeLanguage' and 'Ada' are not equal in any syntactic perspective. The programming ontology establishes a relationship between these two concepts that makes things clearer, i.e. the following relationship exists: Ada rdfs:SubClassOf imperativeLanguage. By using an inference engine, Sara can successfully find a Web site that she is looking for.

We asked all respondents to indicate the reasoning engines they were using (Figure 6). The largest segment (53.6%) indicated that they were using Jena (McBride, 2002). Smaller groups indicated they were using Racer (28%) and Pellet (23.7%). FaCT++ (13.3%) and OWLJessKB (8.1%) have also gained preference by a small group of participants. Finally, engines such as OntoBroker, JTP, KAON2, TRIPLE, F-OWL, and SweetRules have a very small market share. We should notice that 4.3% of respondents do not use any reasoning engine. Other engines, far less popular (<0.5%), included OntoEngine, OpenCyc, OWLAPI, Sesame, SWI-Prolog, and WSMX.

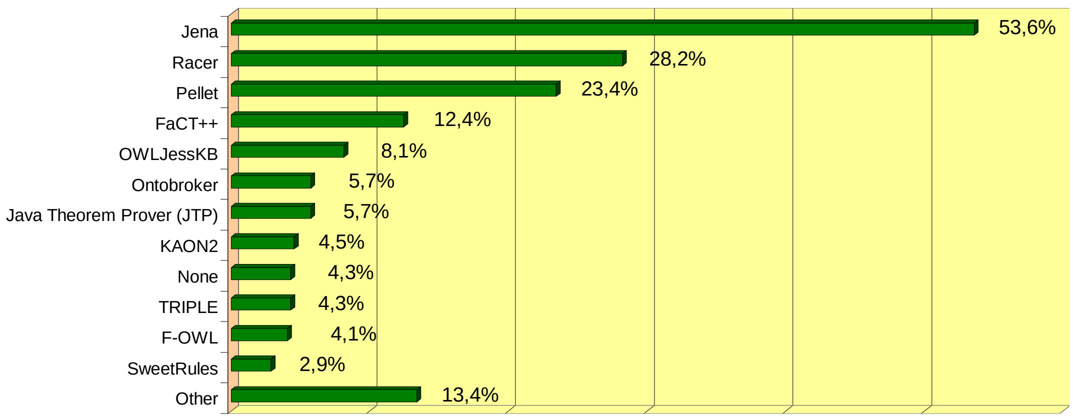

Figure 6. Reasoning engines used.

Ontology Domains. In (Cardoso, J., Sheth, A., 2006), the authors identify the state of the art of the applications that use semantics and ontologies. They identified various applications ranging from the use of semantic Web services and discovery, semantic integration of tourism information sources, semantic digital libraries, semantic grids, semantic enterprise information integration, semantic Web search to the development of bioinformatics ontologies.

To determine the actual trend in the development of ontologies for particular domains, we asked all respondents to indicate for which industries they were representing knowledge with ontologies. The survey provides the perspective of individuals from a wide range of industries. Education and Computer Software are the best represented industries (31% and 28.5%, respectively), followed by Government (17%) and Business Services (17%). Other notable industries are the Life Sciences (16.5%), Communications (13%), the Media (12.8%) and Healthcare providers (11.3%). See Figure 7. Other less significant/prominent industries (not shown in the graph) include Accessibility, Agriculture, Archaeology, eRecruitment, Geography, Geosciences, Human Resource Management, Microelectronics, Physics, Satellites, Sports, and the Toy industry.

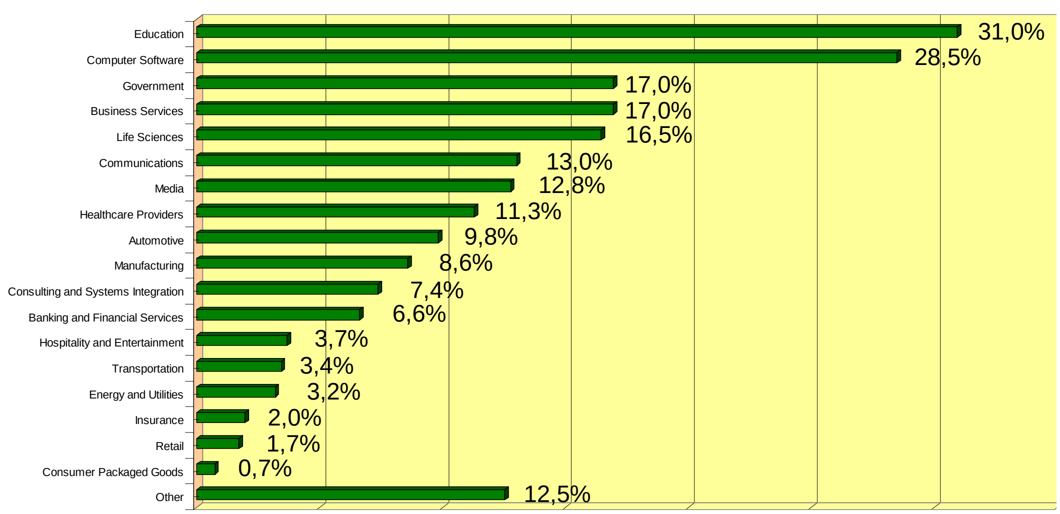

Figure 7. Industry or domain for which respondents develop ontologies.

Methodologies. In software engineering to build a software product a methodology needs to be adopted and followed. Since ontologies are considered to be software products, methodologies, methods and techniques are indispensable to document a development process for building ontologies. Ontological Engineering (OE) is the discipline that studies such methodologies (Gómez-Pérez, A., M. Fernández-López, and O. Corcho, 2007). It as been considered that OE is still a relatively immature discipline and most work groups utilize their own methodology.

We asked all respondents to indicate which methodology or method they were employing to develop their ontologies. We were overcome by the percentage of respondents (60%) that develop ontologies without using any methodology. The methodologies with greatest adoption among ontologists are METHONTOLOGY (13.7%) followed by On-To-Knowledge methodology (7.4%) and Uschold and King's method (4.2%). Other less popular/less commonly used methods include Cyc method, Gruninger and Fox's methodology, DILIGENT method, KACTUS method, SENSUS method, and Noy and McGuinness method.

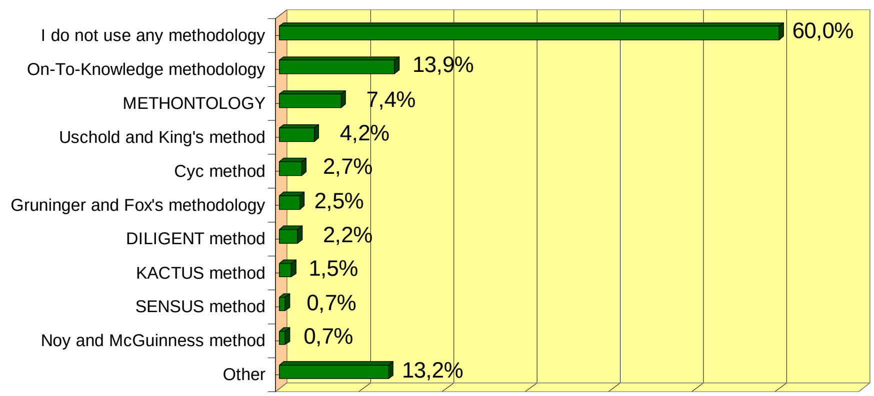

Figure 8. Methodologies used to develop ontologies.

Purpose of using ontologies. Organizations use ontologies for different purposes. We asked participants to tell us the reasons, identified by Noy and McGuinness (2001), that motivated them to use ontologies.

The vast majority of participants (69.9%) answered that they use ontologies to share common understanding of the structure of information among people or software agents, so models can be understood by humans and computers. The second reason for using ontologies is to enable reuse of domain knowledge (56.3%). Approximately equal numbers of respondents use ontologies to make domain assumptions explicit (36.9%) and to analyze domain knowledge (34.2%). A smaller number of participants answered to separate domain knowledge from operational knowledge (29.9%). Finally, other answers (12.4%) indicated that ontologies are also used for code generation, data integration, data publication and exchange, document annotation, information retrieval, search, reasoning, annotating experiments, building common vocabularies, web service discovery or mediation, and enabling interoperability. The results are shown in Figure 9.

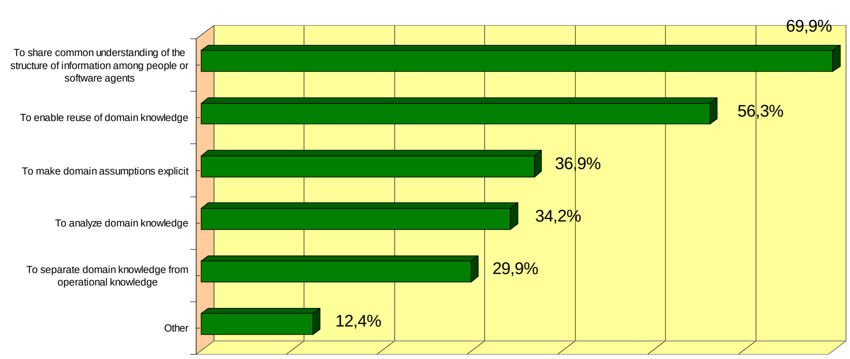

Figure 9. Why to use an ontologies.

Techniques used with ontologies. We asked respondents to tell us if they used any specific technique to manipulate ontologies such as mapping, aligning, merging and integration. The results are shown in Figure 10. The largest segment of respondents (67.1%) indicated that they use ontology mapping, i.e. relating similar concepts or relations from different sources to each other by an equivalence relation. Roughly equal numbers of participants indicated that they were integrating ontologies (40.7%), i.e. building a new ontology reusing other available ontologies (assemble, extend, specialize), and merging ontologies (33.9%), i.e. merging different ontologies about the same subject into a single one that unifies all of them. Twenty six percent stated that they align ontologies, i.e. they bring two or more ontologies into mutual agreement, making them consistent and coherent. Finally, more than fifteen percent of respondents indicated that they perform ontology learning, ontology matching, ontology mediation, ontology pruning, and ontology versioning.

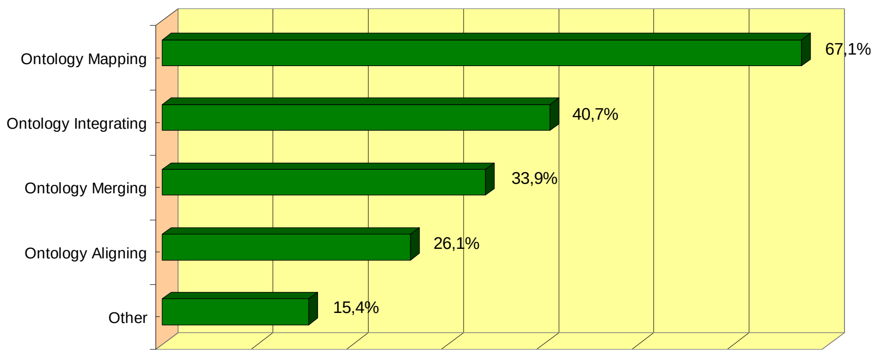

Figure 10. Which ontology techniques do respondents use?

Purpose of using ontologies. A quick look at real-world ontologies that have already been developed makes us wonder about the size of ontologies developers are creating. From life sciences, we can find examples of ontologies with as few as 195 concepts (protein-protein interaction) and 227 (MGED), to ontologies with 2222 concepts (BRENDA), and even large ontologies with 200.002 concepts (GO).

We asked each respondent to indicate the average size of the smallest, typical, and biggest ontologies they were working with. The results are shown in Figure 11. A vast majority of respondents (72.9%) indicated that their smallest ontologies had less than 100 concepts and 20.5% considered that they had between 100 and 1000 concepts. When asked about typical ontologies, forty four percent of respondents stated that such types of ontology had between 100 and 1000 concepts and 35% considered that typical ontologies in their organization have less than 100 concepts. Finally, when asked about the biggest ontologies being deployed, the majority of respondents, 33.5%, considered that this type of ontology had between 100 and 1000 concepts. Roughly equal numbers of respondents, 17.9%, 15.8%, and 16.6%, stated that their biggest ontologies had a number of concepts within the ranges 0-99, 1001-2000, and >10000, respectively. One surprising conclusion from this data is that the ontologies being developed are much smaller in size than can be ascertained from many research papers and conference keynotes and talks.

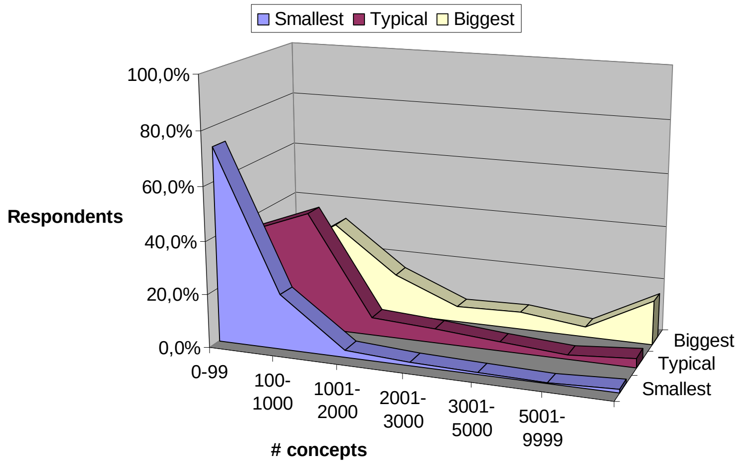

Figure 11. Size of ontologies.

Production. Each respondent was asked to indicate the estimated time before putting an ontology-based system into production. More that 70% of all respondents indicated that they planned to adopt ontology-based systems (Figure 12). A considerable number of participants (27.9%) indicate that they do not have any plans to use such types of systems in the future. More than twenty five (25.4%) of the respondents indicated that their organization was currently active in the development and installation of ontology-based systems. Almost 21% stated that they will put their ontology-based systems into production within 6 months, while 13.7% will have to wait one more year. Clearly many organizations will be active in this area in 2007. The remaining respondents answered that they will install ontology-based systems in 18 months (3.7%) and in 24 months (8.2%).

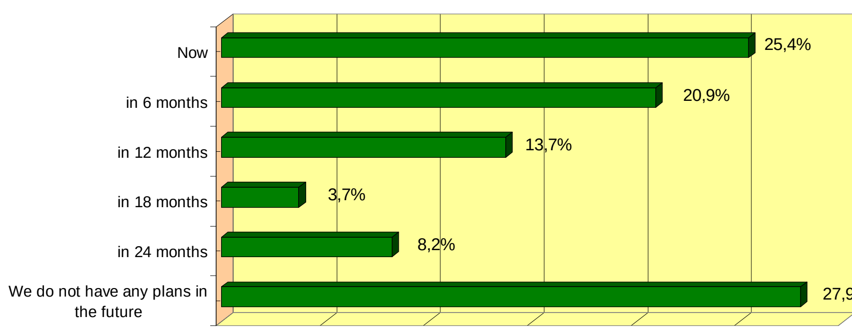

Figure 12. Timeframe to put systems into production.

4. CONCLUSIONS

The Semantic Web vision: Where are we? Where are we heading? While some people think the Semantic Web is just research, this is a misconception since it is already in use and starting to reach the market. The study presented in this article is a fundamental checkpoint to frame the current status and trends in the Semantic Web. We can easily observe that it has already caused a paradigm change from an idea to the development and use of real and concrete solutions. But what lessons can we learn from this study? Can we extrapolate from them what will be important in the future Semantic Web? These are the answers this article provides. For example, this study shows that the Semantic Web does not require complex ontologies. In fact, the large majority of ontologies developed are rather small. It shows that the Semantic Web does not even need OWL and can achieve important objectives such as data-sharing and data-integration using just RDF alone. From the different uses of ontologies, respondents used them mainly to share common understanding of the structure of information among people or software agents, so models can be understood by humans and computers. It also demonstrates that 70% of people working on the Semantic Web are committed to deploying real-world systems that will go into production in less than 2 years.

5. REFERENCES

Cardoso, J., J. Miller, et al. Academic and Industrial Research: Do their Approaches Differ in Adding Semantics to Web Services. Semantic Web Process: powering next generation of processes with Semantics and Web services. Heidelberg, Germany, Springer-Verlag. LCNS 3387: 14-21, 2005.

Cardoso, J. and A. Sheth, "Semantic e-Workflow Composition". Journal of Intelligent Information Systems (JIIS), Springer, Vol. 12, No. 3 (November 2003) pp. 191-225.

Cardoso, J., Sheth, A., "Semantic Web Services, Processes and Applications", Springer, 2006. ISBN: 0-38730239-5.

Chandrasekaran, B., Josephson, J. R., Benjamins, V. R. (1999) "What are Ontologies, and Why Do We Need Them?". IEEE Intelligent Systems. 14(1), pp. 20-26.

Escórcio, L. and Cardoso, J. "Editing Tools for Ontology Construction", in Semantic Web Services: Theory, Tools and Applications, Idea Group, March 2007. ISBN: 1-59904-045-X

Fernández, L. M., "Overview of methodologies for building ontologies", Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends, International Joint Conference on Artificial Intelligence, Stockholm, pp4-1 to 4-13, 1999.

Gomez-Perez, A. and Corcho, O. "Ontology languages for the Semantic Web", IEEE Intelligent Systems, January/February 2002, 17(1), pp. 54-60.

Gómez-Pérez, A., M. Fernández-López, and O. Corcho, Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. Advanced Information and Knowledge Processing. Springer, 2007.

McBride, B. (2002) "Jena: A Semantic Web Toolkit", IEEE Internet Computing, 6(6):55-59.

Noy, N.F., McGuinness, D. L., Ontology Development 101: A Guide to Create Your First Ontology. Knowledge Systems Laboratory Technical Report KSL-01-05, Stanford University. 25p., 2001.

Decker S., Melnik S., van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., Horrocks, I. (2000), "The Semantic Web: The Roles of XML and RDF", IEEE Internet Computing, 4(5), pp. 63-74.

Susie Stephens, Alfredo Morales, and Matthew Quinlan (2006) "Applying Semantic Web Technologies to Drug Safety Determination", IEEE Intelligent Systems, January/February 2006, 21(1).