Semantic Web in 10 years: Semantics with a purpose

[Position paper for the ISWC2012 Workshop on
"What will the Semantic Web look like 10 years from now?"]

Marko Grobelnik, Dunja Mladenić, Blaž Fortuna
J. Stefan Institute, Ljubljana, Slovenia
Marko.Grobelnik@ijs.si, Dunja.Mladenic@ijs.si, Blaz.Fortuna@ijs.si

Introduction

Semantic Web, had its evolution in the last 10-15 years from very basic ideas, mostly initially converted from "Knowledge representation and reasoning" (KRR) area towards its very specific targets, developed along the way relevant in the recent years. We can observe this evolution along several dimensions – for sure, the dimension of the depth and strength of the operators applied on the structured data is the most characteristic for the Semantic Web (through the eyes of ISWC research community). Initially (after 2000), the strength of the operators (mostly inherited from KRR and logic) was relatively high, enabling deep reasoning – this had its own beauty, but lack of applications. In the later years (after 2007) the community (and funders in particular) recognized this deficiency and the area started splitting into 'deep' and 'shallow' Semantic Web, where 'shallow' Semantic Web was targeting possible applications where semantics could get into the game.

Maturity: Crossing the chasm

Trying to extrapolate what might happen in the mid future, one could say, the trend of making Semantic Web closer to where it is needed will get even more emphasized – this can be already seen in the funding trends, where Semantic Web became more like an infrastructural topic (not being funded by itself). We can say, this is a sign of maturity, but still the area needs to cross the real 'chasm' ([1], see Figure 1) to get widely adopted by other research communities and especially the commercial lines of development. As a key goal, this should include adoption of the 'semantic mind set' across research and technologies being in operation.

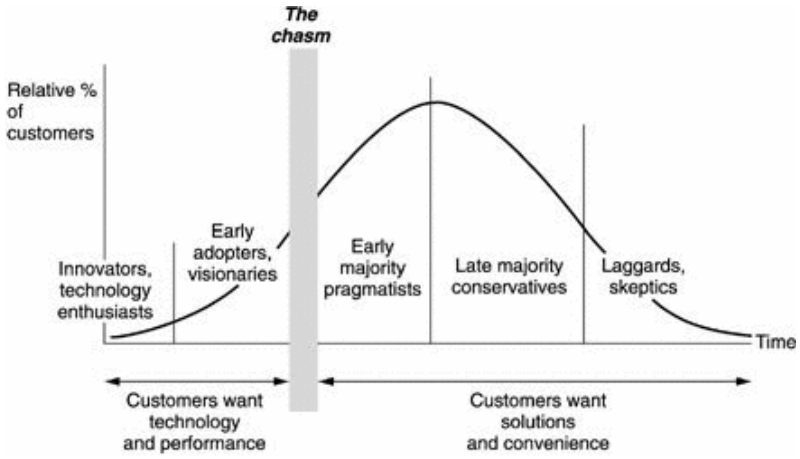

Figure 1: The chasm to be crossed [1]

The mission: Interoperability

Semantic Web in its most general aim is about **interoperability** being needed in almost all areas of research and businesses – it is just not called like this in most cases. Probably, the most important mission of the Semantic Web is the notion of interoperability, defining it on different levels along different operators and delivered through algorithms, tools and systems. At this stage we are only at the beginning of this process covering just some of more popular and not so difficult cases.

Overlap with other research areas

One necessary step to cross the chasm towards maturity should be to get proper overlap with other research communities. Such an overlap can have multiplicative effect in the 10 years' time. At the present time (2012) relations that exists are rather weak, but seem to have an increasing trend.

Since such an overlap is hard to measure, we made just a quick test on the recently very popular and hyped area of "Big Data". We checked the number of web page hits on Google for the keywords "big data" (20 million), "semantic web" (9 million) and combined "big data" & "semantic web" (0.3 million). This kind of measurements has its own problems and one needs to take it with care, nevertheless it provides some idea on the situation.

Another measurement was to search for the number of appearances of the word "semantics" in the four leading books on "Big Data" published in 2011 where we see a rather disappointing presence of semantics in "Big Data" community (see Table 1).

Mining of Massive Datasets [2] (340 pages)	zero appearances
Understanding Big Data [3] (166 pages)	three appearances
Big data: The next frontier [4] (156 pages)	zero appearances
Big Data Glossary [5] (62 pages)	one appearance

Table 1: number of appearances of the words *semantics* in four leading books on "Big Data"

Following the data

Dealing with semantics is essentially dealing with data + abstractions. If we try to structure the domain (in a simplified way) of data, we can split it into the main three dimensions: (1) modalities, (2) operators, and (3) additional properties one needs to take into account (Figure 2). We can say that the individual areas of data related research are sub-cubes of the cube from Figure 2.

Semantics, as a general concept, can theoretically appear as a founding principle in most parts of the data cube. The current practice is that it appears only in some parts. The key here would be to adapt to the needs of particular areas of the data cube and introduce new operators supporting the required applications.

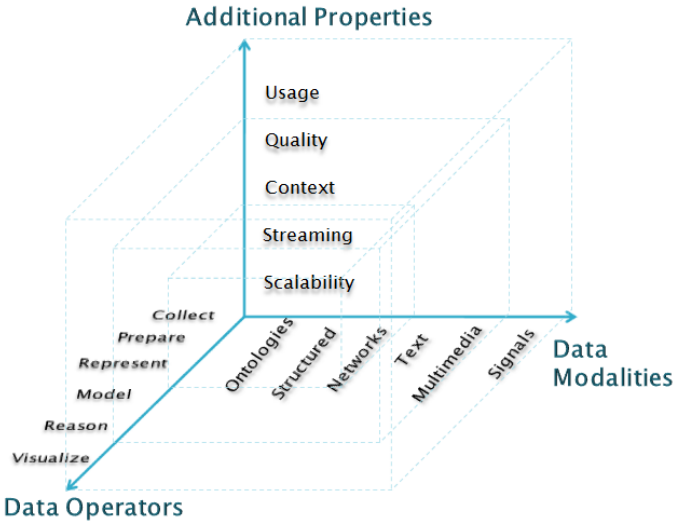

Figure 2: Structuring the domain of data along three dimensions

Another view to the data is across some of the more relevant industrial sectors (see Figure 3) where we see variety in topics and data modalities. In terms of topical interoperability (mainly when we talk about textual data), there is some noticeable progress in mapping terminologies. Since interoperability can serve also as glue for connecting data modalities, this can be an obvious topic for the future of the Semantic Web. Unfortunately some of the data modalities (e.g. video, images, audio) are still in relatively early stage of extracting structured data from unstructured representations; therefore, this can still take some time till the proper uptake.

General comment about the uptake of the Semantic Web in various industries is that each industry solves its interoperability problems in its own way. Structured and standardized approach as offered by Semantic Web penetrated so far only some of them (depending on the business, historical and other constraints).

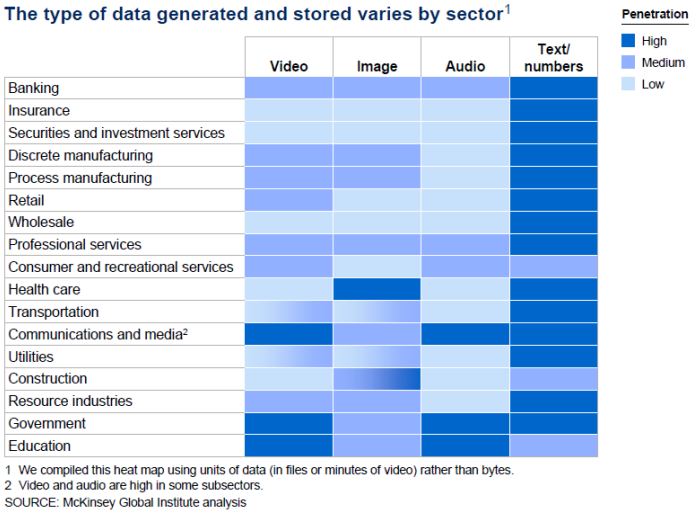

Figure 3: The type of data across industrial sectors

Possible future in 10 years

For the future there can be many important scenarios, all of them in conjunction with some other domain of application. We will address just two of the areas of our core expertise.

Computational linguistics is an area which still suffers a lot from the lack of semantics. Because its traditional approaches, which were more rule driven, and later more statistical driven it came to a position where significant progress can be made only through the use of semantic approaches (logic based or probability based). At the present stage only lighter semantic approaches are being used. Our prediction is that semantics will get heavily involved in the areas of text understanding, as well as machine translation. The need to go towards the old AI dream to understand text is getting higher and the amount of funding to bridge the gap between simple 'text mining' towards deeper 'text understanding' is significantly increasing. Also the changed circumstances like availability of big data and high performance computer infrastructures will contribute a lot. Semantics can play here a role of a glue to connect bits and pieces of human language in a proper ways.

Extremely important trend for the future is "**Internet of Things**"/IoT (see Figure 4 for projected growth). Some of the key properties of IoT systems are (a) unstructured nature of data being generated, (b) the speed and streaming nature of data, (c) fuzziness of data, (d) dynamics and heterogeneity of technical architectures. The variety which needs to be controlled in such a systems is very high and therefore seems as a natural application for semantic technologies. At the present stage only first steps have been made towards standardizations of some of the lower levels of IoT architectures; in the same way, speed, streaming and fuzziness are not properly addressed yet in Semantic Web. Prediction for the future is that the need to control complex IoT systems will push development to resolve uncovered features and in 10 years we can expect whole cities to be smart and ran by semantic technology.

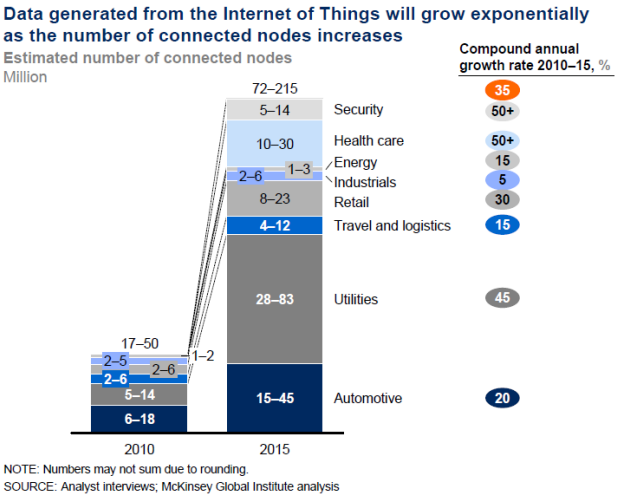

Figure 4: Projected growth for the "Internet of Things"

Summary

In the position paper we presented some of the (possibly key) problems in today's Semantic Web and how to approach them. In particular we addressed:

- Maturity of the Semantic Web which still needs to evolve to cross the chasm.
- In terms of positioning, we structured the data domain along the three dimensions where Semantic Web should search for its future challenges.
- Predictions for the future were made for the domains of Computational Linguistics and Internet of Things where we expect significant developments will happen in the next 10 years.

References

[1] Geoffrey A. Moore: Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers, HarperCollins Publishers, 1999

[2] Anand Rajaraman, Jeffrey D. Ullman: Mining of Massive Datasets, Cambridge University Press, 2011

[3] Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos: Understanding Big Data, McGraw-Hill, 2011

[4] Big data: The next frontier for innovation, competition, and productivity, 2011, McKinsey Global Institute, 2011

[5] Pete Warden: Big Data Glossary, O'Reilly Media, 2011