Using the semantic web to navigate conceptual spaces: an application for the philosophical domain

Using the Semantic Web to Navigate Conceptual Spaces: an Application for the Philosophical Domain

Michele Pasin, Enrico Motta

Knowledge Media Institute,
Open University, UK
{m.pasin,e.motta}@open.ac.uk

Abstract. The semantic web offers new possibilities in the eLearning area, which are dependent at least on two factors: the increasing availability of resources on the web and the semantic power of their meta-descriptions. This paper looks at this second issue, and tries to determine the general framework and the important features of the metadata for a specific domain, philosophy. Thanks to an adequate conceptualization, in fact, it is possible to recollect resources and compose them into a novel narrative, in order to provide specific learning services.

1. Introduction

Computers are everywhere, and the Internet is regarded as the most powerful communication device ever seen in human history. The Internet, in fact, makes a vast amount of information available, but does this really facilitate human capability to learn, or does it just open a large number of *cyber paths*, without giving a proper *map* in order to make sense of them?

A key-word search engine like Google can help us find a list of resources, connected merely by a *string* similarity, and, as we know, many times it fails in answering our initial research question. Of course this happens because a computer can hardly understand the sense of our words, but treats them only syntactically, namely, taking into account their external shape and not their meaning.

If the normal web we browse daily can be seen as a huge repository of this kind of "meaningless" information, the project of the semantic web consists of a meta layer built on top, in order to describe it and make it more meaningful to an automated agent. Therefore, it ideally points to a situation where all data comes together with the description of how to use it, a situation which would realize a web of knowledge and not only a web of information.

Given these premises, it is possible to go one step forward and imagine a semantic search[1], namely, a search action that is informed by this meta layer above the web and that points to resources in the web or in the real world. In such a scenario, it is possible to look for the name of a person, and retrieve only results that correspond to

the class of entities classified as “people” within the semantic metadata. Furthermore, it is possible to perform some reasoning over the knowledge represented in the metadata, and retrieve results following some more complicated paths, in a way similar to the actual reasoning *our minds* perform.

While researchers are trying to find the most effective ways for *annotating* the "old" web, new perspectives are also emerging in the area of eLearning. In fact, in a scenario where resources are annotated and could be found on the web, instead of a normal search engine we could have an intelligent knowledge browser[2] that, given a goal, follows some pre-existing knowledge patterns, gathering a set of resources that fulfill the goal. For example, I could ask this software agent to help me understand a specific concept in physics, and receive a series of knowledge elements that, properly digested, will bring me to the understanding of that concept. In other words, the knowledge browser *bridges* on-the-fly the missing spaces between what I already know and what I would like to know, giving me the opportunity to contextualize a piece of information and actually learn how to get there semantically.

1.1 Scenario

Imagine Robert, a student doing research on the new frontiers of biology for his university course. While browsing the Internet looking for relevant material, he runs into an interesting article about a new discipline, sociobiology. Skimming through the text he realizes that he missed the overall point of the article, because there is a concept that still is unclear to him, “evolution”. What he needs is to locate the resources that would fill the gap that does not let him carry on his research. Unfortunately, a normal Google search is too vague, and retrieves information that will just reduce his focus. Therefore, through a browser plug-in (analogous to Magpie[3]), he highlights the unknown concept and selects, between different ones, a specific narrative path related to it, let’s say, “concept explanation”. The system, making use of a domain ontology (in this case a philosophical one) locates the position of the concept “evolution” and of the various theories associated with it, within different research areas. The correlated resources are gathered and, through a discourse and media ontology, formatted in order to create a particular learning path for the understanding of the concept (Figure 1).

1.2 Existing technologies

This vision, as pointed out by Stojanovic and others[4], becomes quite feasible in a semantic web scenario, where agents can reason on the semantics of the learning resources and therefore select them in order to achieve goals that the user specifies dynamically.

The technologies most used in the semantic web in order to construct semantic structures are XML, RDF[5], and ontologies[6]. The overall aim is always to express more machine-understandable meaning, nonetheless these technologies vary quite a lot in their expressive power: in fact they are depicted by Berners-Lee[7] as constituting a continuous framework where one stands on top of the other, taking

advantage of the representational power of the closest technology underneath and giving more human-like abstraction capability to the closest technology above. So, for example, while the XML layer represents the structure of data, the RDF layer represents the meaning of data; the Ontology layer, instead, represents the formal common agreement about meaning of data; above all these stands the Logic layer, which enables intelligent reasoning with meaningful data.

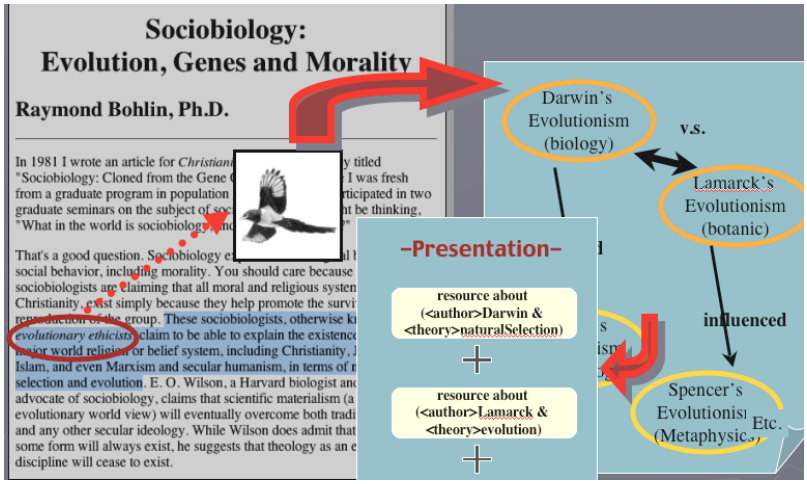

Fig. 1. Explanation of a concept through a learning web service

In our case, from the learning point of view, ontologies constitute one of the major contributions to the field. Since an ontology is a formal, shared conceptualization of a particular domain of interest, it facilitates enormously the communication between the people or the agents that act within that specific domain.

The "paths" of the learning process, e.g., the narratives that a learner follows, for example, do not have to be defined anymore by the publisher of a learning resource, but can be constructed on the fly through the usage of the meta-description associated with each single learning object. As in a lego system, different pieces match together depending on their shape.

The conventional metadata used to describe learning resources, up to now, are just a set of common tags (expressed in XML or RDF) that could be applied to any resource, in order to express how learning materials could be used in an interoperable way.

Three emerging metadata schemas represent the standards for e-learning: IEEE LOM[8], ARIADNE[9] and IMS[10]. These schemas, however, are basically taxonomies that describe either some generic features of the learning objects, or the kind of activities they could be used within. These metadata focus on course production generation, namely they describe how to put together a set of learning objects according to some general rules, without trying to represent their inner semantics. In other words, they work on a macro level, and doing so they lack the granularity and semantic expressiveness needed in order to describe the specific learning scenarios (the concepts and the relations linking entities within a domain).

Therefore, since domain knowledge is only shallowly treated, the models employed are not powerful enough to develop a suitable learning narrative.

1.3 Problem analysis

This lack of domain knowledge could be avoided through the usage of ontologies as a referring "universe of discourse" of learning resources. An ontology can describe concepts, the relations between them and also behaviors of entities (expressed through rules). This is what is needed in order to abstract the learning object itself from the characteristics of its usage. Following a structuralist approach[11], we can methodologically decompose the learning process into three different dimensions: content (what the learning material is about), discourse(in which form the material is connected, at the conceptual and rhetorical level), and media (how the final presentation transmits the material).

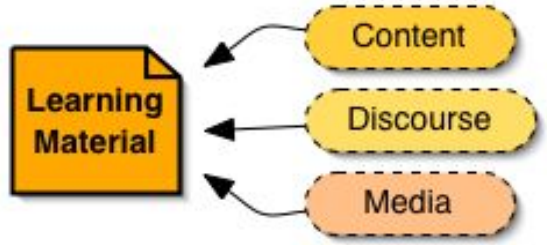

Fig. 2. Methodological division of the learning resource

Therefore, to sum up this brief overview, we are now able to spot the key problems in semantic web learning (Figure 2):

- when it comes to the **content**, there are a set of problems related to the appropriate definition of the metadata of a domain, e.g.: they should be faithful to the entities they represent but at the same time not too specific (for re-usability reasons); they should support understanding of the key features of the domain; also, depending on the representational language used, they may possess more or less semantic expressiveness;
- for that regards the **discourse**, the problems reside in the connection between concepts or resources (represented by concepts); the automatic generation of a narrative subsequent to the user-input is in fact determined by a set of rules and exploration facilities provided by a system, and strictly related to the kind of scenario we are working within,, e.g. learning or generic introduction;
- finally, for that regards the **media**, there is ongoing research[12] in order to develop adequate standards of the formatting of the learning material. Once the semantics of a set of resources is clear the visual rendering of the relations that link them becomes important, e.g., an example, a figure or a key-concept should stand out differently in the final presentation (maybe also through a video or audio clip), or a tangent concept could only be hyperlinked and explained in another page.

2. A navigational approach to the philosophical domain

This section introduces initial work aimed at the instantiation of the above ideas and techniques about a feasible and effective semantic web approach to learning. The domain chosen is philosophy, for a series of reasons that could be summarized in the following three points:

- the lack of an explicit semantic formalization of this domain, namely, of a consistent and detailed metadata definition;
- the widespread availability of philosophical resources (documents in a broad sense: text, image, video) on the web;
- the semantic richness of the domain, that could be translated into an advanced and non-trivial navigational capability.

2.1 Generic framework for the domain ontology

An ontology for philosophy can be seen as a categorization of the domain that should be, at the same time, quite precise and quite detached. Precise in focusing on the key points that we can find throughout the philosophical work, the underlying themes that guide the questioning and the explicit places where the research has historically condensed. Detached, since, in order to maintain a wider applicability, the categorization tries to be *a-philosophical* to the maximum level: it does not itself bring forward a philosophical standpoint (of course, at the extreme level, this is impossible) but its strength lies in its being a meta-philosophy.

This standpoint is clearly expressed by the difference in meaning of the concepts *systemic* and *systematic*. While the first one refers to an approach that tries to reduce everything to a single principle or set of principles, therefore explaining a whole set of phenomena as different manifestations of the same underlying reason, the second one refers to an approach that lets things 'live in themselves', namely, it maintains the specificity of things and does not try to reduce one to the other.

In doing so, we have imagined ourselves to be in a situation similar to that of a librarian who owns a large amount of philosophical books, and wants to cleverly organize them. The final aim of this categorization, therefore, is to provide a high level semantic structure that allows intelligent navigation of philosophical resources.

In a similar vein to Dieberger and Frank[13], a potentially useful metaphor we can use to describe philosophy is to see it as a wide *territory*, defined, firstly, by the needs of someone who has to go through it (e.g. a student). These needs are the *highways* that give sense and direction within the territory, and correspond to the problems that guide the research and that give birth to the philosophical questioning. These dimensions can be defined as the basic philosophical **problems**, and their answers can be located, following the work of Schulz[14], throughout the entire history of philosophy.

The strength of this approach relies in its being a-philosophical, in a sense. In fact, it starts from a normal person's experience, the everyday life, and from there it raises the philosophical departure points. The problems, therefore, are not yet *philosophical discussions*, but just the possible spark of them. Browsing through the problems is a way to connect philosophical resources in a quite detached and accessible way.

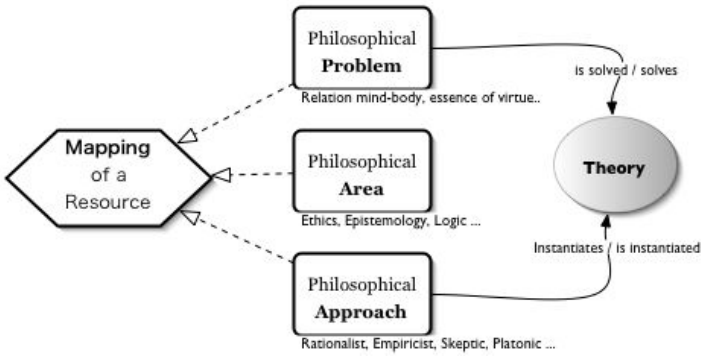

Fig. 3. The major dimensions of the mapping of a philosophical document

Within the territory, the highways (together with smaller roads - the sub problems) connect the different *regions*, and actually also define them. Following the initial metaphor, in fact, a region gathers around a set of communication channels, namely, around one or more problems.

Regions therefore stand for the institutionalized and historical ways to solve one or more problems, and correspond to the established philosophical **areas**. The same problem, in fact, can appear in different contexts and consequently also generates different solutions, depending on how it is tackled. An area gathers a knot of solutions (representations of the problem and of its resolution) that with time may become autonomous, and maintain their meaning independently of their "birth place".

These solutions correspond to the philosophical **theories**. A theory is a constellation of concepts, namely a model, a representation of a recognized (philosophical) standpoint in the community, As such, it is directly related to a specific problem. In fact, theories correspond to the other side of problems: they solve them and they are used in different areas (regions), therefore they are also well represented by the *highways*, in the territory. That is, theories are the main vehicle we can use in order to navigate through the philosophical space: the evolution of a theory and its movement from one domain to another within the history of thought is the most powerful semantic navigation we can define.

Moreover, within the areas, we can recognize another different force: the philosophical **approach**. An approach is defined as the application of a theory in a specific domain (thus, the theory by definition is related to more than one domain). Approaches correspond to the *cities* in the regions, since they can be similar to each other but still not the same, at the very least because they are instantiated in different areas.

A particular **location** (Figure 3) in the philosophical territory is therefore defined in terms of a problematic field, a philosophical area and an approach (i.e. the instantiation of a theory). These dimensions allow an adequate initial mapping of the semantic position of any resource.

The entire generic semantic framework that has been presented until now, however, constitutes only a part of what is needed in order to efficiently browse philosophical resources. We can name this kind of knowledge interpretative, since it

is defined by the person who annotates a document or by who inserts a conceptual schema. Eventually, this kind of knowledge is strictly subjective and relies on a person's standpoint and interpretation of a text.

On the other side, there is the factual knowledge, that is the set of 'indubitable' information about authors and their works, for example that related to dates, places, publishers or historical periods. This kind of knowledge is of course quite obvious and basically it is what stays at the root of the narrative of the most common books about the history of philosophy, nonetheless it is still precious for us, since it can be used in conjunction with the interpretative knowledge, to generate cross-narratives (for example, to show the evolution of a theory within a particular age).

3. Related Work

The work of **Schank**[15] constitutes a fundamental theoretical background when trying to grip the relationships between human's learning and storytelling abilities. Basically he draws a line that connects intelligence, understanding, conversational structures and stories. In his opinion, since our knowledge scales down to the set of stories we are able to tell, the most interesting question becomes how we manage to get from one story to the other, namely how we constantly index new stories and relate them to the corpus of stories we stored in the past.

Story Fountain[16] is a tool developed to support a community in the exploration of digital resources. Thanks to a domain ontology and a story and narrative ontology, different stories are annotated and stored in a database, for later being recollected in an enhanced way through the usage of the knowledge in the ontologies. Compared to a simple string matching retrieval, Story Fountain provides a great improvement towards the understanding of the stories, in fact it generates intelligent navigational paths through the novel connection of different concepts.

The **Topia**[17] project, instead, is focused on the production of hypermedia presentations from the semantics of potentially unfamiliar domains. Although the authors recognize the necessity of human insight in order to generate a story, they still feel that there is a subset of narrative and discourse concepts that one can automatically derive from semantics, and that includes, for example, the order of a presentation, the grouping of components into sections, etc.. This mapping is derived through a clustering algorithm, whose resulting sets are then linked to a set of properties defined in the discourse ontology, generating a so-called "structured progression".

Hardman and others[18] describe instead a system that makes use of the domain knowledge as an essential feature of an ontology-driven transformation process that generates a complex presentation design. The initial semantic graph gathers the information that is selected and ordered through a discourse ontology, with the aim of constructing a structured progression. In addition to discourse knowledge, the process here is also guided by media design knowledge from a design ontology.

References

1. R. Guha, R. McCool, E. Miller. *Semantic search.* in *WWW2003 --- Proc. of the 12th international conference on World Wide Web*. 2003. Budapest, Hungary: ACM Press.

2. Arthur Stutt, Enrico Motta, *Semantic Learning Webs.* Journal of Interactive Media in Education, 2004. **1**(Special Issue on the Educational Semantic Web).

3. M. Dzbor, J. B. Domingue, E. Motta. *Magpie - towards a semantic web browser*. in *Proc. Of the 2nd Intl. Semantic Web Conference*. 2003. Florida US.

4. Liljana Stojanovic, Steffen Staab, Rudi Studer. *eLearning based on the Semantic Web*. in *WebNet2001 - World Conference on the WWW and Internet*. 2001. Orlando, Florida, USA.

5. Decker, S.; Melnik, S.; van Harmelen, F.; Fensel, D.; Klein, M.; Broekstra, J.; Erdmann, M.; Horrocks, I., *The Semantic Web: the roles of XML and RDF*, in *Internet Computing*. 2000. p. 63-73.

6. Natalya F. Noy, Deborah L. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology*, in *Stanford Knowledge Systems Laboratory Technical Report*. 2001, Stanford University.

7. Berners-Lee, Tim, *What the semantic web can represent*. 2000.

8. *IEEE LOM*: http://ltsc.ieee.org/doc/wg12/LOM3.6.html.

9. *ARIADNE*: http://ariadne.unil.ch/Metadata/.

10. *IMS*: http://www.imsproject.org/metadata/imsmdv1p2/imsmd_infov1p2.html.

11. Chatman, Seymour, *Story and Discourse*. 1978: Cornell University Press.

12. Ossenbruggen, Jacco van, Hardma, Lynda, Geurts, Joost and Rutledge, Lloyd. *Towards a multimedia formatting vocabulary*. in *Proceedings of the twelfth international conference on World Wide Web*. 2003. Budapest, Hungary.

13. Andreas Dieberger, Andrew U. Frank, *A city metaphor for supporting navigation in complex information spaces.* Journal of Visual Languages and Computing, 1998(9): p. 597-622.

14. Schulz, W., *Philosophie in der veranderten Welt*. 1972.

15. Schank, R.C., *Tell Me A Story: A New Look at Real and Artificial Memory*. 1990, New York: Macmillan.

16. Paul Mulholland, Trevor Collins, Zdenek Zdrahal. *Story Fountain: Intelligent support for story research and exploration.* in *Proceedings of the 9th international conference on Intelligent user interface*. 2004. Funchal, Madeira, Portugal: ACM Press.

17. Lloyd Rutledge, Martin Alberink, Rogier Brussee, Stanislav Pokraev, William van Dieten and Mettina Veenstra. *Finding the Story – Broader Applicability of Semantics and Discourse for Hypermedia Generation.* in *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*. 2003. Nottingham, UK: ACM Press.

18. Hardman, J. Geurts S. Bocconi J. van Ossenbruggen L. *Towards Ontology-driven Discourse: From Semantic Graphs to Multimedia Presentations*. in *Proceedings of the Second International Semantic Web Conference (ISWC 2003)*. 2003.