PhiloSURFical: browse Wittgenstein's world with the Semantic Web

Michele Pasin, Milton Keynes, United Kingdom

How many resources about Wittgenstein exist on the web? How do they relate to each other? What is the most productive way to navigate them, from the point of view of a learner? With the development of the PhiloSURFical tool, we aim at investigating these and other related issues. PhiloSURFical is a software environment which builds on Semantic Web technologies in order to facilitate the navigation and understanding of Wittgenstein's first work, the Tractatus Logico-Philosophicus. By relying on an ontology created to describe the philosophical domain at various levels of abstraction, PhiloSURFical presents the original text and other associated resources in a contextual manner. This can be achieved through a process of narrative pathway generation, that is, the active linking of resources into a learning path that contextualizes them with respect to one another. In this article we introduce the learning paths which PhiloSURFical makes available and highlight some of the modelling issues which emerged as fundamental in supporting such navigations, in the emerging web of data.

Vision: a semantic web *for* philosophers?

How could a web navigation enlighten or affect a philosophy scholar? Especially within an educational scenario, is the constantly increasing number of philosophical web materials a source of confusion, or an advantage? In our work we have been investigating the requirements and features of the possible navigation mechanisms a philosophy student could benefit from. In particular, in the context of the Semantic Web [1], we have identified some of the "learning pathways" which can be used for dynamically presenting these materials within a meaningful context.

For example, imagine that from the paragraph 7 of Wittgenstein's Tractatus, by selecting an *interpretative* navigation path, you could easily jump to Max Black's detailed commentary on it. And from there, being interested on Black's interpretation and wanting to gather information on its possible origins, you were able to query the web using a *comparative* navigation path, aimed at highlighting what Wittgenstein and Black had in common. Two main results are returned: both studied at Cambridge, both worked in the philosophy of language area. You decide to focus your attention on Cambridge, click on it, select an *historical* perspective and see that while in Cambridge, in the 20's, Black had the opportunity to listen to and meet some of the major scholars of the time: Russell, Moore and Ramsey were among them. Now you may want to reorganize these results, according to a *theoretical* perspective. Thus you discover that another link among all these philosophers is their interest in the philosophy of mathematics, and that actually Black's first book was centred on this topic. So you drift away for a moment, select again a *theoretical* navigation, pull up a small map of the important views in philosophy of mathematics in the last century, and see that among them there is also the first philosophy of Wittgenstein. You click on it, select a *textual* navigation and automatically you are taken back to the Tractatus, but this time to paragraph 6.2.

The sort of links that would make possible such navigations are of a slightly different nature than the famous hyperlink which, together with other things, made the fortune of the web. And if Google [2] does a great job in meaningfully organizing for us the web of hyperlinks, it cannot do much if we wanted to query directly the web of relations existing among our world's entities. In order to do so, resources need to be indexed and described not only at the syntactic level (e.g. with respect to their status as an image, a text file or a video), but also at the semantic one, i.e. with respect to their content. The Semantic Web effort, or web of data, brings forward the ambitious vision of creating and maintaining this "semantic layer" of the web, so to allow software agents (e.g. programs like the navigation tool described above) to accomplish various operations which would not be otherwise possible.

Often, with great and inspiring visions, also come great and challenging difficulties. The Semantic Web makes no exception here [3]. During our work with the PhiloSURFical tool we faced many of them, some with an exquisitely philosophical flavour, others of an inherent technical character. In the rest of this article, we want to draw the readers' attention to some lessons learned during the construction of the ontology and hopefully show how, despite the various limitations, the benefits of such an approach make the enterprise worth pursuing.

The PhiloSURFIcal tool

PhiloSURFical is a pedagogical application which allows the contextual navigation of a semantically-enhanced version of Wittgenstein's Tractatus Logico-Philosophicus [4]. By relying on an ontology created to describe the philosophical domain at various levels of abstraction, users can benefit from multiple perspectives on the text and on related resources. Moreover, users can take advantage of the application for storing their own annotations about the Tractatus and possibly contribute to the creation of a network of philosophical resources centred around the text and its author. For the moment, as the availability of free and adequate semantic data on the web is still limited, PhiloSURFical strongly relies on an internal knowledge base, but its architecture attempts to be open and extensible so to allow future integration and querying of different repositories, using the appropriate web standards (e.g. RDF [5], SPARQL [6], OWL [7]).

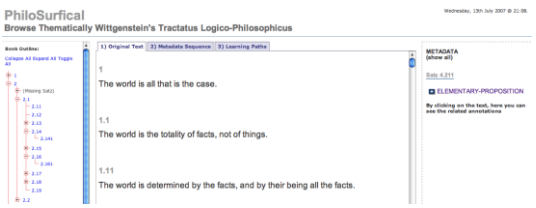

Figure 1. Screenshot of the PhiloSURFical prototype

In general, the usage of PhiloSURFical can be framed within the educational activity of learning through discovery of related resources. According to doctrines such as constructivism and situated cognition [8], this learning style is particularly effective because it pushes students towards the active exploration of a subject and the subsequent discovery of the interlinked nature of all knowledge. By constructing their own "paths" through the available learning materials, students engage directly with a subject mat-

ter and are more likely to actively construct a meaning out of it.

Thus, the semantic model behind PhiloSURFical has been designed with a clear purpose: the model should support the reconstruction of the history of ideas, by relying on structured information about the practical domain and the theoretical domain of thinkers. Our approach takes the notion of a learning pathway as a "system of specially stored and organized narrative elements which the computer retrieves and assembles according to some expressed form of narration" [9] and attempts to transpose it within the specific scenario made up of philosophical entities.

However, these pathways cannot be completely open-ended. Instead, they should be semi-structured, so to avoid phenomena such as information overload or inconclusive navigations. As a possible solution, we have formalized a number of *generic* learning paths. These represent the most interesting ways to browse the ontology, across one of its dimensions (or more of them simultaneously). So, for example, we can have a *theoretical* learning path (which focuses on the relations among ideas), a *textual* learning path (which attempts to retrieve related information objects), a *historical* learning path (which keeps results in chronological order) a *geographical* one etc. Of course, the paths can also be specialized: within the theoretical path, there can be a *disambiguation* one (which highlights concepts having the same name, but being actually defined by different views), a *contrast* one (which highlights opposing views) etc.

In conclusion, the data from both the local knowledge base and other (previously mapped) information sources can be dynamically reorganized and presented with relevance to the actual context. As the emerging Semantic Web makes available a larger number of queriable resources (e.g. the DBpedia [10], a structured version of the Wikipedia), so the navigation mechanisms will develop with regards to their complexity and interestingness.

Issues in modelling the philosophical domain

As said above, the PhiloSURFical system relies on an ontology. In AI terms, an ontology is often defined as an "explicit specification of a shared conceptualization" [11] and practically consists of a rich formal taxonomy augmented with typed relations, quantifiers and rules. The key feature of ontologies, is that computers can *process* it, so to *infer* some new relationships among data. In the context of the Semantic Web, ontologies can be viewed as a sort of "web deduction mechanism", that is, a reasoning backbone for the web of data. But first of all, ontologies provide a way for guaranteeing the semantic interoperability among different information providers. We do not want here to delve into the many problems involving the ontological representation *capabilities* and *limitations*. It is noteworthy that these problems are possibly increasing when trying to represent philosophical ideas, and the relations among them. Instead, we would like to stress that, as claimed by the authors of a recent project for the indexing of the Stanford Encyclopedia of Philosophy "while no single ontology can possibly capture the full richness and interrelatedness of philosophical ideas, we are operating on the principle that having (at least) one ontology is better than none" [12].

The specific approach used to realize the PhiloSURFical ontology has at its centre the decision to employ the CIDOC Conceptual Reference Model [13] as a starting point for our formalizations. The CRM ontology was originally an attempt of the CIDOC Committee of the International Council of Museums (ICOM) to achieve semantic interoperability for museum data. Since 1996, the formal model has improved considerably till becoming in 2006 an ISO standard (version 4.2). The choice of using the CRM was motivated by two reasons. Firstly, for its widely recognized status as a standard for interpreting cultural heritage data. In fact, by reusing and extending an existing and internationally recognized ontology, we will give our tool's users more chances to benefit from the emerging Semantic Web infrastructure. Secondly, for its extensive event-centred design. This design rationale, in fact, appeared to be appropriate also when trying to organize the history of philosophy: even if it is common to see it as an history of ideas, stressing the importance of the *theoretical* (i.e. meta-historical) dimension, this cannot be examined without an adequate consideration of the *historical* dimension. That is, a history of the events related (directly or indirectly) to these ideas.

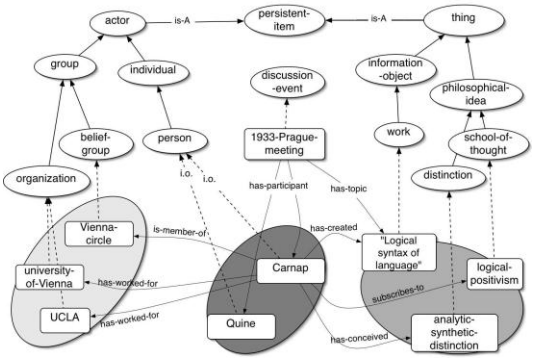

Figure 2. Example of an event-based representation

As an example, in figure 2 we can see an event-centred representation in the PhiloSURFical ontology. The *persistent-item* class, which is one of the five classes composing CIDOC's top layer (together with *time-specification*, *dimension*, *place* and *temporal-entity*) subsumes *thing* and *actor*. The two branches of the ontology departing from them can have various instances, which are related by taking part (in various ways) to the same event ("1933-Prague-meeting"). This kind of modelling, in the context of the PhiloSURFical tool, is extremely useful because of the multiple navigational pathways it can support (e.g. we could move to another event having the same topic, or to another topic treated during the same event, etc.).

In order to provide support for representing the multiple facets a philosophical fact can have, the ontology has been created by integrating other already existing models. In particular, we included knowledge about the domain of publications from the AKT reference ontology [14] and knowledge about information objects from the related module [15] of the DOLCE foundational ontology [16]. Moreover, as we are dealing with a domain where bibliographic resources are central, we have also attempted to build a model that is possibly compliant with a cataloguing standard. To this purpose, we are providing mappings and reusing notions from the Functional Requirements for Bibliographic Records (FRBR) specifications [17], which are a very influential standard for librarians. Finally, a large portion of the PhiloSURFical ontology is constituted by a se-

ries of new concepts and relations, mostly aimed at the description of philosophical events and ideas.

We can see with another example how these different formalizations can be used together. As CIDOC is not providing an easy way to model the social and intellectual activities of philosophers, we created various classes for this purpose, which are grouped under *social-activity* and *intellectual-activity*. Within the first group, we have five subclasses: *discussion*, *joining-a-group*, *educational-activity*, *close-social-contact* and *social-gathering*. Partially inspired by some AKT formalizations, these entities have let us extend the already supported event-based kind of reasoning. By instantiating such a model, as shown in figure 3, we can specify that the book by Kimberley Cornish (titled "The Jew of Linz" [18]) has as subject the fact that Wittgenstein, while studying at the Linz Realschule, had Hitler as one of his young school-fellows. Such a modelling can easily bring to a learning path which intertwines *publications* and *events* in the philosophical world.

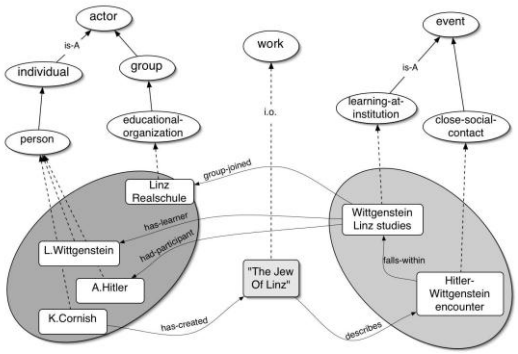

Figure 3. Representing the content of a work through events

Of course, there are many other modelling issues which we could not present here, for space reasons. This is not a surprise, if we just consider the size and complexity of the philosophical domain. In particular, the modelling of concepts regarding ideas and their relations is difficult and mostly overlooked in the literature [19]. At the moment, the ontology is undergoing a refinement phase thanks to the feedback given by various domain experts, but thanks also to users' feedback on the narrative pathways PhiloSURFical makes available.

Conclusions

In this article we presented PhiloSURFical, a software tool which takes advantage of various Semantic Web technologies to support the learners' task of finding relevant resources. The tool is prototyped with Wittgenstein's Tractatus Logico-Philosophicus, one of the most influential philosophical texts of the twentieth century. We have described its operating principles and shown how the ontology it is based on can support various navigation features. As the modelling of a domain such as philosophy is subtle and challenging, we have briefly discussed our approach and provided references to other useful semantic models we integrated. The PhiloSURFical tool and ontology are still in the evaluation phase, but are available online at http://philosurfical.open.ac.uk.

Literature

1. Berners-Lee, T., J. Hendler, and O. Lassila, The Semantic Web, in Scientific American. 2001.

2. Beavers, A., Searching for Philosophy: A Review of Google Scholar and Google News. Teaching Philosophy, 2005. 28(4).

3. Halpin, H. The Semantic Web: The Origins of Artificial Intelligence Redux. in Third International Workshop on the History and Philosophy of Logic, Mathematics, and Computation, (HPLMC-04 2005). 2004. Donostia San Sebastian, Spain. .

4. Wittgenstein, L., Tractatus Logico-Philosophicus. 1921: Routledge & Kegan Paul.

5. W3C. RDF Primer. 2004; Available from: http://www.w3.org/TR/rdf-primer/.

6. W3C. SPARQL Query Language for RDF. 2007; Available from: http://www.w3.org/TR/rdf-sparql-query/.

7. W3C. OWL Web Ontology Language Overview. 2004; Available from: http://www.w3.org/TR/owl-features/.

8. Brown, J.S., A. Collins, and P. Duguid, Situated Cognition and the Culture of Learning. Educational Researcher, 1989. 18(1): p. 32-42.

9. Brooks, K.M. Do Story Agents Use Rocking Chairs? The Theory and Implementation of One Model for Computational Narrative. in ACM Multimedia. 1996. Boston MA, USA.

10. Auer, S. and J. Lehmann. What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content. in ESWC. 2007.

11. Gruber, T.R., A translation approach to portable ontology specifications. Knowledge Acquisition, 1993. 5(2): p. 199-220.

12. Niepert, M., C. Buckner, and C. Allen. A dynamic ontology for a dynamic reference work. in Joint Conference on Digital Libraries - JDCL-07. 2007. Vancouver, British Columbia, Canada.

13. Doerr, M., The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI Magazine archive, 2003. 24(3): p. 75-92.

14. AKT. AKT Reference Ontology v.2 - AKTive Portal Ontology v.2. 2002.

15. Gangemi, A., et al., Task Taxonomies for Knowledge Content - D07 in Deliverable of the EU FP6 project Metokis. 2005.

16. Gangemi, A., et al. Sweetening Ontologies with DOLCE. in 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02). 2002.

17. IFLA, Functional requirements for bibliographic records : final report, K.G. Saur, Editor. 1998: München.

18. Cornish, K., The Jew of Linz 1998: Century Hutchinson.

19. Pasin, M. and E. Motta. Supporting Philosophers' Work through the Semantic Web: Ontological Issues. in Fifth International Workshop on Ontologies and Semantic Web for E-Learning (SWEL-07). 2007. Marina Del Rey, California, USA.

Email: m.pasin@open.ac.uk