

# PhiloSURFical: browse Wittgenstein's world with the Semantic Web

Michele Pasin

Knowledge Media Institute

The Open University, Walton Hall, Milton Keynes, UK, MK7 6AA

m.pasin@open.ac.uk

## Abstract

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 modeling issues which emerged as fundamental in supporting such navigations, in the emerging web of data.

## 1. 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 (Berners-Lee, Hendler, & Lassila, 2001), 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 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 centered 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 (Beavers, 2005) 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 (Halpin, 2004). 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 final section of this article, we will 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. In the next section, instead, we will outline the most important functionalities of the PhiloSURFical application (which is available online at <http://philosurfical.open.ac.uk>) showing how they could facilitate and enhance the understanding of the Tractatus.

## 2. The PhiloSURFical tool

### 2.1 Overview

PhiloSURFical is a pedagogical application which allows the contextual navigation of a semantically-enhanced version of Wittgenstein's Tractatus Logico-Philosophicus (Wittgenstein, 1921). 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. 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 (W3C, 2004b), SPARQL (W3C, 2007), OWL (W3C, 2004a)). At the time of writing, the prototype is mainly focused on browsing functionalities. However, in the next version we plan to extend it so that users can also store 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.

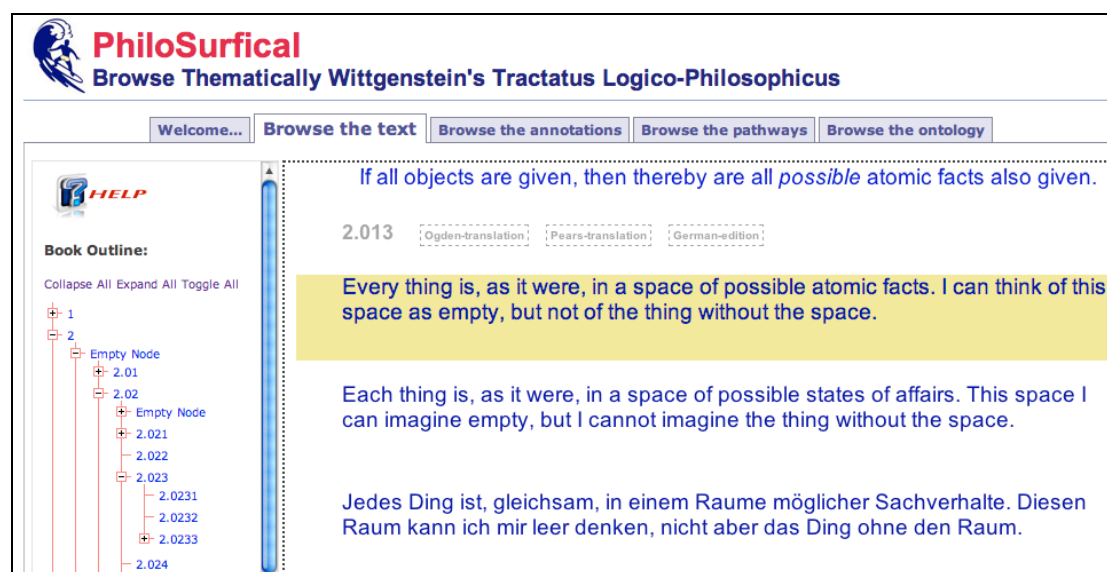


Figure 1. Screenshot of the PhiloSURFical application

The tool is organized into five main sections or *tabs*. In general, we attempted to organize the tabs' sequence according to their increasing difficulty of usage (namely, the first tab requires less 'learning effort' than the second one, the second one than the third one, etc.). By doing so, we wanted users to have a more 'gradual' encounter with the software. This becomes important especially when considering that not all Wittgenstein's scholars might be familiar with complex web-based educational tools. The five tabs can be described as follows: the *welcome* tab serves as a splash screen and provides some contextual information and links to relevant resources; the *browse the text* tab presents three versions of the Tractatus's text in a simple but highly interactive interface; the *browse the annotations* tab supports a different type of text navigation by means a smart-index of the topics associated to the Tractatus' fragments; the *browse the pathways* tab lets users select topics of interest and explore related resources by using the "pathways" metaphor; finally, the *browse the ontology* tab visualizes the tree-hierarchy of the ontological representations PhiloSURFical relies on, letting users have a glance at the underlying complexity of the model. In the next section we will now give a more detailed description of the tabs' functionalities, highlighting for each one of them what are the most salient features and functionalities.

## 2.2 Tabs' description

### Tab 1 – Welcome

This section does not provide specific functionalities for navigating the Tractatus, but it has a fundamental role in giving some information to the (possibly random) user regarding the purpose of the prototype. Moreover, it links to various external learning resources about Semantic Web related topics, which should help users in understanding both the terminology and the functioning of PhiloSURFical.

### Tab 2 – Browse the text

In this tab users can simply browse the text, which is made available in three of its 'versions' (the original German edition and the two major English translations). In order to facilitate this activity, a tree-like outline of the book on the left hand side lets them jump quickly to a specific paragraph; also, it provides a handy overview of the highly hierarchical Tractatus' structure. Another interesting feature in this tab is a javascript mechanism by which it is possible to select what Tractatus' version to visualize: when the mouse hovers one of the paragraphs, this is highlighted and a contextual menu appears above the text. By clicking on one of the available options, it is possible to view more than one translation at the same time (as shown in Figure 1, with the proposition 2.013).

### Tab 3 – Browse the annotations

This is where the ontological backbone of PhiloSURFical starts becoming more evident. At the centre of the screen users can still read the text, but now four panels, two on the right side and two on the left side, provide alternative ways to engage with the Tractatus. We can imagine this functionality as a 'smart index' of the text. That is, an index of the subjects that is dynamically updated depending on what text fragment the user is focusing on. However, in order to understand how this works we should first examine the underlying ontological representations.

Thanks to the PhiloSURFical ontology (see section 3), all the text fragments are represented as *instances* of *information-objects* (specifically, they are instances of the class *text-fragment*). Each one of them has associated one or more annotations, that is, they have been interpreted by some experts in our team as *being-about* certain topics. At the ontological level, this has been carried out thanks to another class called *information-object-interpretation*, whose instances express the connection between an information object and an idea or topic. By doing so, every text-fragment could have multiple interpretations and each one of them is encoded in a different object which remains separated from the instance expressing the text-fragment itself. So, for example (see figure 2), paragraph 2.1 has been 'linked' to the annotation object "picture". Similarly, other paragraphs have also been associated to this annotation and consequently whenever we click on the topic "picture" we can retrieve all relevant the text fragments.

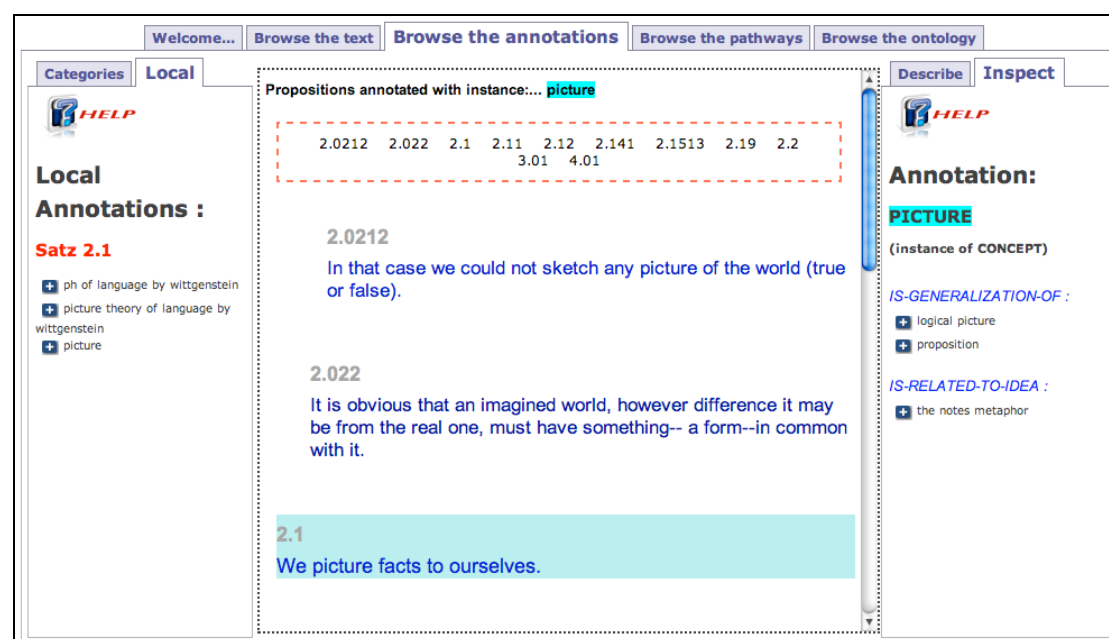


Figure 2. Browsing the text using the annotations' smart-index.

This is the first functionality this tab is offering to the user: by selecting a topic, it is possible to see only all the sections of the Tractatus that have been associated to it. Also, in the "describe" panel we can read a short description of the topic just selected. Conversely, by clicking on a paragraph, it is possible to see which are the associated contents (which appear in the "local" panel) and use them for navigating the text in a non-linear manner.

It is important to remember, at this point, that all the Tractatus' interpretations we created are clearly just *our* view of the text's meanings. In other words, they are just some possible interpretations which, although we hoped being significant and thus capable of helping learners in understanding the text, could have been done differently and are here primarily as a way to showcase the functionalities of the tool (actually, in future releases, we would like to create an interface that supports users in creating *their own* annotations and possibly also sharing them).

The second functionality provided by the third tab of PhiloSURFical is related to the fact that a certain topic (e.g. in our case, the topic “picture”) is not just a string or an unstructured tag, but is instead an instance of one of the *types of ideas* we classified in the ontology. In particular, it is an instance of the type *concept*<sup>1</sup>. Thus, being an instance, it has been created along with several properties (such as name, description, etc.) and it can also be *interpreted* in an analogous manner as the text-fragment interpretation we have seen above (only, this time, this is achieved through a *concept-interpretation* class, which serves for describing a concept-instance using various other properties). For example, the “picture” concept has been interpreted from our experts by using the properties *is-generalization-of* and *is-related-to-idea*, whose values are respectively the concepts of “logical-picture” and “proposition” for the first property, and the rhetorical-figure “the notes metaphor” for the second property. As shown in figure 2, the “inspect” panel displays these interpretations. Also here, users can click on them and thus re-load the text displayed accordingly. The “inspect” panel is updated too and now shows the interpretations of the newly selected topic. This sort of navigation facility can be compared to the usage of a “map” of the Tractatus’ topics for browsing the text’s paragraphs.

Finally, in the “categories” panel it is possible to find all the topics associated with the text, organized into the eight idea-categories defined in the ontology. This panel acts as a general index where users can find quickly the contents they are interested in, or simply browse them to see what is available.

#### Tab 4 – Browse the pathways

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 (Brown, Collins, & Duguid, 1989), 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 matter 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” (Brooks, 1996) and attempts to transpose it within the specific scenario made up of philosophical entities.

Figure 3. Example of an historical learning pathway: the PhD-advisors chain

Accordingly, in the fourth tab we attempted to create a virtual environment for building user-triggered “learning pathways”. In more practical terms, users can select a content of interest (or just use the most recently selected content, which is in focus by default) and use it as the starting point of a *semantic* navigation. Once they have selected an item, learners may click on one of the available choices appearing in the ‘pathways list’ panel. Each selectable pathway-type comes also with a brief description explaining its meaning. Once triggered, the pathway’s results are shown on the right panel as a set of interrelated entities. The relations are supposed to highlight the significant connections among the pathway’s items; moreover, when such items are clicked they are automatically put into focus for a new semantic search, so to support a recursive search process. A ‘recent items’ panel is used to keep track of all the selected items since we started, so to be able to put them back into focus simply by clicking on them.

<sup>1</sup> In total, through various analyses we discovered eight idea-types: *problem*, *problem-area*, *view*, *concept*, *argument-structure*, *method*, *distinction* and *rhetorical-figure*. For a more detailed exposition of this, please refer to (Michele Pasin, Motta, & Zdrahal, 2007)

This sort of navigation will usually produce results that go beyond the limited scope of the Tractatus: for example, starting from the *view* defined as '*Wittgenstein's picture theory of language*', a *theoretical-pathway* highlighting the network of rival theories could visualize '*Frege's theory of meaning*' and '*Augustine's theory of language*', among others. Instead, if we started from the instance of *person* named '*Frank Ramsey*' (as shown in fig.3), by using an *historical-pathway* that highlights the chain of PhD advisors we would find out that he is in the lineage originating from C.H.Weisse.

In other words, users here can benefit from a set of 'query templates' (which we call learning pathways) so to explore PhiloSURFical's knowledge-base and other web repositories in a contextual manner. The 'pathways' rely on the semantic relations formalized in the ontology and on other 'semantic mappings' that link them to other ontologies available in the Semantic Web. However, these pathways cannot be completely open-ended: if they were so, even if they are using the 'semantic' links among resources (as opposed to the more usual 'syntactic' ones), they would easily reproduce well-known phenomena such as information overload or inconclusive navigations. As a possible solution, we decided to provide users with a series of *semi-structured* ways to query the knowledge base, by constraining them only on *certain types* of semantic relations. Thus, we formalized a number of "generic" learning pathways that represent the most interesting ways to browse the ontology across one of its dimensions (or more than one simultaneously). So, for example, we can have a *theoretical* learning pathway (which focuses on the contrasting relations among ideas), a *textual* learning pathway (which attempts to retrieve related information objects), a *historical* learning pathway (which keeps results in chronological order) a *geographical* one etc. Of course, the paths can also be specialized: within the theoretical pathway, 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. At the time of writing, the pathways facility is the only functionality in PhiloSURFical which is still under development: this is for both the inherent complexity of its implementation and for the fact that it strongly relies on a large quantity of highly structured philosophical data (which is still not accessible). Nonetheless, we envision that as the emerging Semantic Web makes available a larger number of queriable resources (e.g. the DBpedia (Auer & Lehmann, 2007), a structured version of the Wikipedia), so the navigation mechanisms will develop with regards to their complexity and interestingness.

#### Tab 5 – Browse the ontology

This section aims at familiarizing PhiloSURFical's users with the underlying technology the application is using. On the left hand side, by means of a tree-like menu it is possible to navigate the hierarchical structure of the ontology used to represent the Tractatus and all the other related resources the tool is presenting. By clicking on the classes' names we can see their description on the right panel: this is composed by some information about their position in the ontology (such as what *super-classes* or *sub-classes* they have), a natural language explanation of the classes' role and significance and a series of *slots* or *properties* linking them to the other classes. In the following section we will give more details about the ontology and its characteristics.

## 2. 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" (Gruber, 1993) 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" (Niepert, Buckner, & Allen, 2007).

The specific approach used to realize the PhiloSURFical ontology has at its centre the decision to employ the CIDOC Conceptual Reference Model (Doerr, 2003) 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.

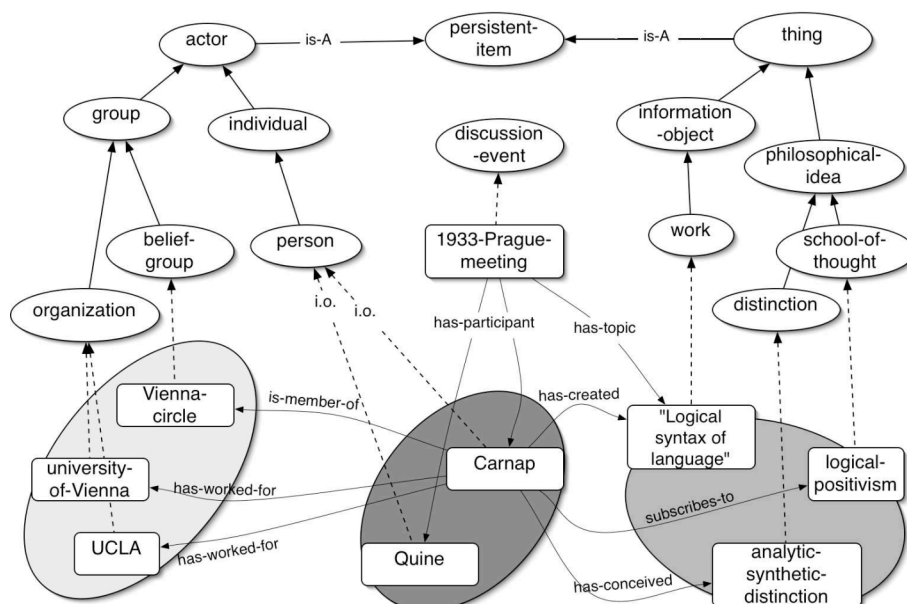


Figure 4. Example of an event-based representation

As an example, in figure 4 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 (AKT, 2002) and knowledge about information objects from the related module (Gangemi, Borgo, Catenacci, & Lehmann, 2005) of the DOLCE foundational ontology (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002). 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 (IFLA, 1998), which are a very influential standard for librarians. Finally, a large portion of the PhiloSURFical ontology is constituted by a series 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 5, we can specify that the book by Kimberley Cornish (titled "The Jew of Linz" (Cornish, 1998)) 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.

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 (Michele Pasin & Motta, 2007). 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.



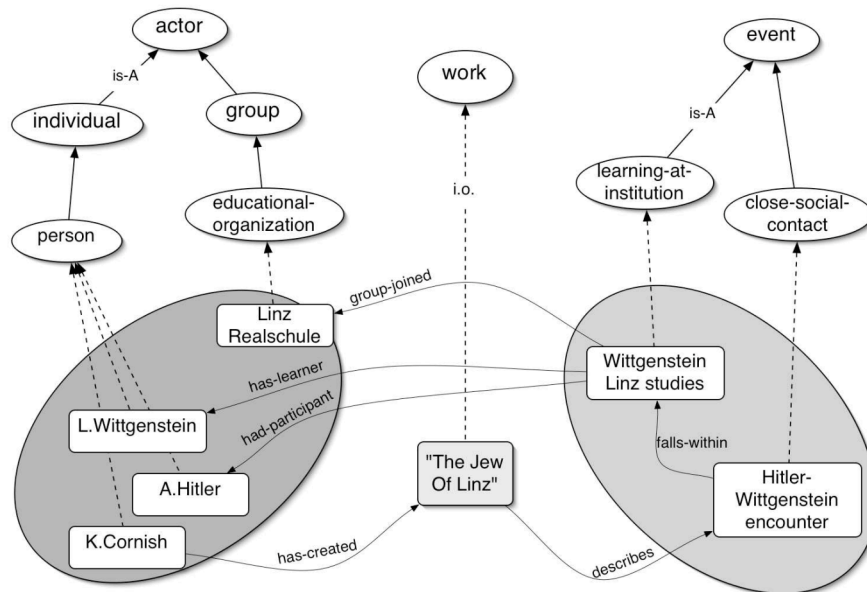


Figure 5. Representing the content of a work through events

## Conclusions

In this article we presented PhiloSURFical, a software tool that 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>.

## References

- AKT. (2002). AKT Reference Ontology v.2 - AKTive Portal Ontology v.2.
- Auer, S., & Lehmann, J. (2007). *What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content*. Paper presented at European Semantic Web Conference 2007.
- Beavers, A. (2005). Searching for Philosophy: A Review of Google Scholar and Google News. *Teaching Philosophy*, 28(4).
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*.
- Brooks, K. M. (1996). *Do Story Agents Use Rocking Chairs? The Theory and Implementation of One Model for Computational Narrative*. Paper presented at the ACM Multimedia, Boston MA, USA.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, 18(1), 32-42.
- Cornish, K. (1998). *The Jew of Linz* - Century Hutchinson.
- Doerr, M. (2003). The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI Magazine archive*, 24(3), 75-92.
- Gangemi, A., Borgo, S., Catenacci, C., & Lehmann, J. (2005). *Task Taxonomies for Knowledge Content - Deliverable 07*
- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). *Sweetening Ontologies with DOLCE*. Paper presented at the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02).

- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199-220.
- Halpin, H. (2004). *The Semantic Web: The Origins of Artificial Intelligence Redux*. Paper presented at the Third International Workshop on the History and Philosophy of Logic, Mathematics, and Computation, (HPLMC-04 2005), Donostia San Sebastian, Spain. .
- IFLA. (1998). *Functional requirements for bibliographic records : final report*. München.
- Niepert, M., Buckner, C., & Allen, C. (2007). *A dynamic ontology for a dynamic reference work*. Paper presented at the Joint Conference on Digital Libraries - JDCL-07, Vancouver, British Columbia, Canada.
- Pasin, M., & Motta, E. (2007). *Supporting Philosophers' Work through the Semantic Web: Ontological Issues*. Paper presented at the Fifth International Workshop on Ontologies and Semantic Web for E-Learning (SWEL-07), Marina Del Rey, California, USA.
- Pasin, M., Motta, E., & Zdrahal, Z. (2007). *Capturing Knowledge about Philosophy*. Paper presented at the International Knowledge Capture conference (K-CAP'07) Whistler, British Columbia, Canada.
- W3C. (2004a). OWL Web Ontology Language Overview. from <http://www.w3.org/TR/owl-features/>
- W3C. (2004b). RDF Primer. from <http://www.w3.org/TR/rdf-primer/>
- W3C. (2007). SPARQL Query Language for RDF. from <http://www.w3.org/TR/rdf-sparql-query/>
- Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus* (C. K. Odgen, G. E. Moore & F. P. Ramsey, Trans.): Routledge & Kegan Paul.