Michele Pasin, Simon Buckingham Shum, Enrico Motta

Knowledge Media Institute, Open University

Milton Keynes, United Kingdom

m.pasin@open.ac.uk

Formalizing 'philosophical' narratives: the tension between form and content

Abstract: The philosophical scholarship is undoubtedly based on processes such as the *interpretation* of a view (or a text), and the *argumentation* for supporting or contrasting such interpretations. How can the most recent computer technologies support these activities, and in particular, how could the latest web developments bring the philosophical experience to a new level? In order to better understand and foster further discussion on such topics we will present some relevant work we have been doing at the Knowledge Media Institute, Open University. In particular, the tools we are describing attempt to tackle these problems by using *argument mapping* techniques, for properly identifying and visualizing the various elements which compose an argument, and *semantic technologies*, for being able to encode at least part of the meaning of the philosophical resources we are dealing with, so to have machines perform a better job in tasks such as search or navigation. What these two techniques have in common is the fact that by *formalizing* a specific domain of knowledge we can use computers in a more effective way, e.g. for mimicking behaviors which are normally associated to competent humans. Nonetheless, by presenting these works we also want to convey to the reader a central idea, which originates in a *tension* between the attempts to formalize the *form* of an argument, as opposed to those ones formalizing the *contents* of the argument. Our insight is that the way forward, especially in the area of distributed semantic e-learning applications, must involve the resolution of this dichotomy. In other words, we believe that the way human discourse takes shape within the languages of the different disciplines is a phenomenon which can be better simulated and supported with computers only thanks to an hybrid approach.

1. Introduction

What does it mean for a student to come to an understanding of a philosophical standpoint and can the explosion of resources now available on the web support this process? The philosophical scholarship is undoubtedly based on processes such as the *interpretation* of a view (or a text), and the *argumentation* for supporting or contrasting such interpretations [1]. How can the most recent computer technologies support these activities, and in particular, how could the latest web developments bring the philosophical experience to a new level?

In order to better understand and foster further discussion on such topics we will present some relevant work we have been doing at the Knowledge Media Institute, Open University. In particular, the tools we are describing attempt to tackle these problems by using *argument mapping* [2] techniques, for properly identifying and visualizing the various elements which compose an argument, and *semantic technologies* [3], for being able to encode at least part of the meaning of the philosophical resources we are dealing with, so to have machines perform a better job in tasks such as search or navigation.

What these two techniques have in common is the fact that by *formalizing* a specific domain of knowledge we can use computers in a more effective way, e.g. for mimicking behaviors which are normally associated to competent humans.

For example, in the context of the Iraq debate [4], we could codify the *knowledge* needed in order to query a database containing a discussion forum, by using questions such as “what are the claims backing the arguments pro-war, which have not been challenged yet?”. Instead, in the context of a Wittgensteinian debate [5] we might want to provide a data model capable of providing sensible answers to more content-oriented ques-

tions, such as “which other authors have used a picture-inspired theory, outside the philosophy of language?” or “did any of Wittgenstein’s contemporaries use the concept of *sense* in a similar fashion?”.

In the following sections we will describe two systems attempting to provide such functionalities. Firstly we will describe Cohere, a web-based environment for generating networks of ideas and allowing asynchronous discussions around them. Secondly, we will present PhiloSURFical, a system which intends to facilitate the study and understanding of a philosophical text, by letting users browse it through a map of the ideas related to it.

These applications, although already operative and used by learners in informal contexts, are still to be considered as prototypes. Since they deal with the formalization of *abstract* domains of knowledge (i.e. ideas, philosophies, etc.), which are notoriously hard-to-formalize by using logical languages, their development will require a thorough trial-and-error process (to this end, we are currently running various evaluation experiments so to gather feedback and address the weak points of the systems).

Nonetheless, by presenting these works we want to convey to the reader a central idea, which originates in a *tension* between form and content. As we will see, the *formalization techniques* we just mentioned tend to focus either on modeling the <u>form</u> of an argument (i.e. the various possible relations among the claims of a debate, such as the ones of *agreement* or *disagreement*), or on modeling the <u>content</u> of the resources we may refer to in an argument (i.e. the *ontological status* of an object, with respect to a chosen set of descriptors: for example, the fact that a resource is *about a philosopher,* or that it *deals with the creation of a concept*).

Our insight is that the way forward, especially in the area of distributed semantic e-learning applications, must involve the resolution of this dichotomy. We believe that the

way we proceed in forging our debates (particularly, the philosophical ones), and similarly, the way human discourse takes shape within the languages of the different disciplines, are phenomena which can be better simulated and supported with computers only thanks to an hybrid approach. That is, an approach capable of putting form and content into the same formal representation, as opposed to the traditional approach which tends to isolate them. In the last section of this paper, we will spend more words on this idea, also picturing how it could be embodied in our future work.

2. Modeling arguments: the Cohere system

Cohere [6], the most recent tool we've created, attempts to incorporate several principles emerged with the Web 2.0 in a web-based argumentation environment. Cohere aims to be semantically and technically open, provide an engaging user experience and social network, but provide enough structure to support argument analysis and visualization. By letting users import their data from a large number of data formats, such as web feeds, DELICIOUS tags [7] or even more complex ontologies, Cohere invites them to make connections between ideas. This broader framing aims to meet the need of many *sense-making* communities to express how ideas or resources are related (whether or not this is argumentative) in a way that goes beyond plain text blog postings, wikis or discussion forums. Moreover, we believe that Cohere will help forge links not only between Ideas, but between the people publishing them.

A fundamental aspect of the theoretical approach behind the Cohere system has been defined elsewhere as characterized by the concept of 'principled disagreement' [8]: users have total freedom on the choice of the *contents* (ideas) they disagree on, but they are forced to use certain *forms* (relations, structures) of disagreement. This allows the system to meaningfully derive some consequences from their choices of relations. For example, Cohere can highlight all the ideas *supporting* another idea, or all the ones *challenging* it. These are among the simplest examples: we envision the creation of more complex results from this sort of queries, once the system's database will reach an adequate size through users' interaction.

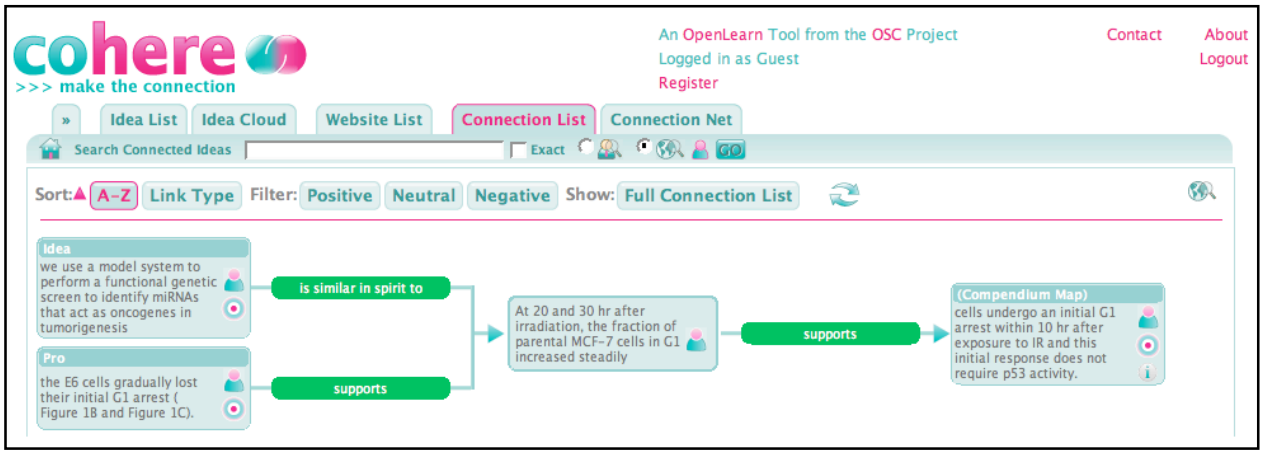

Fig. 1 Screenshot of Cohere

It is important to mention that in Cohere users can also *tag* the ideas they create, if not for other reasons, just for more easily organizing and sharing them. Although this activity is often considered to be equivalent to the one of giving a 'hint' about the semantics of the object we are tagging, we should not be fooled by this analogy. The *tag* that describes the contents of an idea is quite different from any formal specification of its semantics. This could be instead achieved in other software applications, for example, by 'describing' the same idea using a concept from an ontology or a taxonomy. In other words, *folksonomies* and *ontologies* have distinct roles [9]: tags work well for some purposes, but in order to allow more complex *reasoning* tasks we definitely need some more structured metadata. It

is fair to say, therefore, that the overall 'system semantics' in Cohere (i.e. the degree of formal specification of structured information) must be intended as limited to the *types of relations* that can be employed to construct connections among ideas.

3. Modeling philosophical entities: the PhiloSURFical system

PhiloSURFical [10] is an application based on the idea of semantically-informed *narrative pathways*, which can be used to guide students in navigating information about the philosophical domain. Our approach exploits the rapidly growing amount of semantic markup on the web deriving from world-wide initiatives such as the Semantic Web [11] and the Linking Open Data project [12]. In particular, we attempt to provide mechanisms for automatically integrating semantic information available from distributed resources, and to present it to the user in the form of a map of related entities.

PhiloSURFical has been prototyped with Wittgentein's Tractatus Logico-Philosophicus [13] and it relies on an ontology [14] created to describe Wittgenstein's world and the philosophical domain in general at various levels of abstraction. Arguably the ontology is at present the most extensive in its kind.

Learners can start by freely navigating the text (which is available in the original German version and the two main English translations), then move on to browsing it using a *smart index* of the related ideas which is updated depending on the text fragment they are focusing on. The more advanced functionality of the tool is instead the *pathways* facility: by selecting a content of interest it is possible to find other related materials on the web (also with respect to a wider context than just the Tractatus' one). This exploration process deliberately wants to be *semi-structured*: students are free to find new resources (which can or cannot be relevant to their research), but they must use some *query templates* that con-

strain their search to be effective only on some specific semantic relations (which are classified as *historical*, *geographical*, *theoretical,* etc.).

Thanks to these 'query templates', the resulting data-set is smaller than what it would be by using a classic keyword-based search mechanism. Moreover, the data retrieved are also partially ordered according to their meaning (i.e. their formal description). For example, in the context of a *historical* search-template, data would be ordered chronologically; instead, by using a *theoretical* perspective data would be organized using argumentation schemas or other logical links [15].

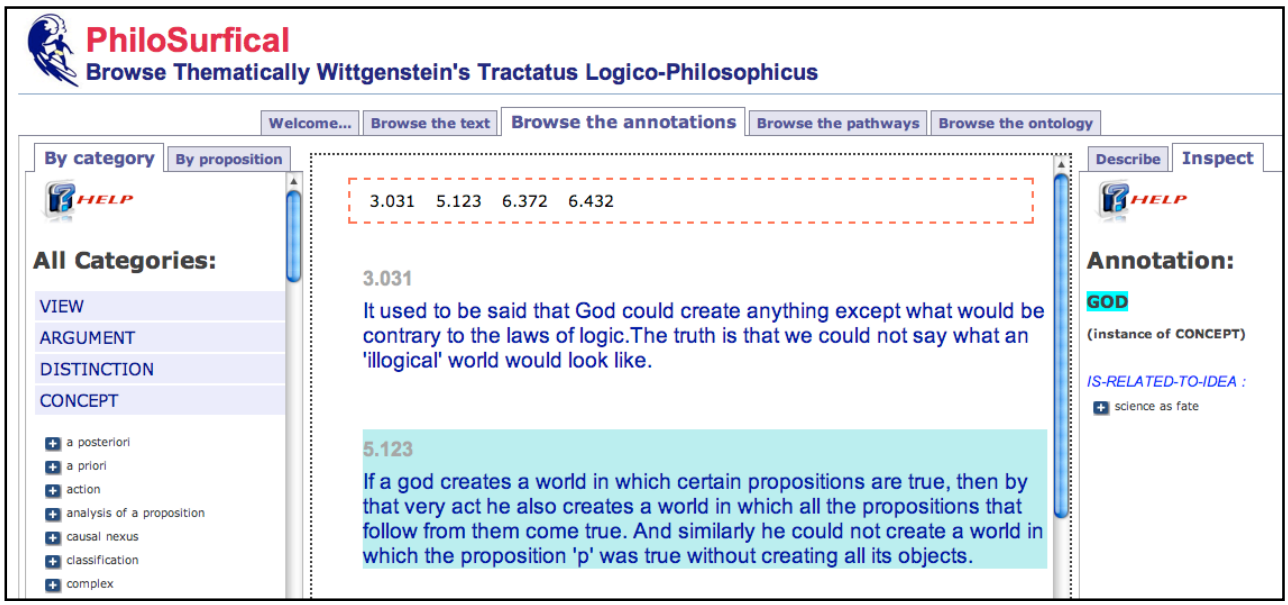

Fig. 2 Screenshot of PhiloSURFical

In general, the idea that inspired us here is that the resulting organization should attempt to mimic some of the classic *ways* the discipline employs to *narrate itself* (for this reason, the pathways are called *narrative pathways).* Obviously, in order for such a complex mechanism to work, data must be adequately structured and stored using a formal representation language. As at the moment we can only rely on a small number (but constantly growing) of resources on the 'web of data', for bootstrapping purposes we decided to implement an initial internal knowledge-base by using information extraction techniques on various free philosophy-related web resources.

The interesting point we want to highlight is that PhiloSURFical's functionalities are *entirely* based on a quite complex underlying ontology. The data model acts as the 'brain' of the system: it describes precisely the possible 'senses' (i.e. meanings) an entity can have in the context of the software application. However, soon enough we realized that although the entities contemplated by the model are many (more than 300), they would always end up being a very small subset of the entities a philosopher normally talks about.

In other words, by modeling a large set of the domain's discourse *contents*, PhiloSURFical can produce interesting navigations and probably drive users to the serendipitous discovery of interesting resources. Nonetheless, its main limitation lies in the same semantic model it is empowered by: the model is fixed, thus *meanings* must also be fixed. As a result, the overall application becomes hardly extendable, if not by using a *top-down* approach; moreover, the platform does not support the process of *agreement* or *disagreement* about its content descriptors. In fact, if users were able to change them, the whole *reasoning chain* would collapse, making the application unusable.

In conclusion, it appears that the advantages of creating suitable semantic structures for making a computer program perform *content-intensive* tasks are always counterbalanced by its weaknesses when it comes to support the *evolution, modification* and in general the *debate* around such structures.

4. Conclusion: are form and content meant to be separate?

From the two examples we have just described form and content appear to be orthogonal dimensions - one should not have influence on the other - and often their being separated, in the context of software applications, is necessary in order to limit *complexity*, if not also *computability*.

Imagine a system that should be capable of determining the relations among a series of contesting claims, based on the declared (formalized) contents of the claims. Surely, it would be harder to construct and also more prone to errors, at least for the simple fact that the combinatorial space it implements is much larger. Moreover, such system would also need much more structured information from users, in order to start being functional (i.e. to provide answers to the example-queries presented in the introduction). In fact, its 'system semantics' should support the representation of knowledge about both the *structure* and the *contents* of an imaginary philosophical debate, or of the contrasting interpretations of a philosophical viewpoint.

Despite these difficulties, we believe that the described dichotomy between form and content must be overcome in order to pave the way for a new generation of tools supporting learners' and scholars' activities. The rapid increase of computing power on one hand, and of available repositories of structured data on the other, is making more feasible the construction of systems where the computation is based, simultaneously, on information regarding both argumentation structures and resources' contents'.

In future work, we plan to investigate further in this direction by building systems that bring together the functionalities of both Cohere and PhiloSURFical. We hope that this brief description of them, together with the reflections on their limitations, will foster further discussion during the conference which will help us in refining our thinking.

1. Carusi, A., Taking Philosophical Dialogue Online. Discourse: Learning and Teaching in Philosophical and Religious Studies, 2003. 3(1): p. 95-156.
2. Kirschner, P., S.B. Shum, and C. Carr, Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making. 2003, London: Springer-Verlag.
3. Staab, S. and R. Studer, Handbook on Ontologies. 2004: Springer.
4. Shum, S.B. and A. Okada. Modelling the Iraq Debate: Mapping Argumentation in a Document Corpus. Available from: http://kmi.open.ac.uk/projects/compendium/iraq/.

5. Pasin, M. PhiloSURFical, browse thematically Wittgenstein's Tractatus Logico-Philosophicus. Available from: http://philosurfical.open.ac.uk/.
6. Shum, S.B. and M. Bacher. Cohere: make the connection. Available from: http://cohereweb.net.
7. DEL.ICIO.US - A social bookmark manager. Available from: http://del.icio.us/.
8. Buckingham-Shum, S.J., et al., Modelling Naturalistic Argumentation in Research Literatures: Representation and Interaction Design Issues. International Journal of Intelligent Systems, 2004.
9. Gruber, T., Ontology of Folksonomy: A Mash-up of Apples and Oranges. International Journal on Semantic Web & Information Systems, 2007. 3(2).
10. Pasin, M. PhiloSURFical: browse Wittgenstein's Tractatus with the Semantic Web. in 30th international Wittgenstein Symposium. 2007. Kirchberg am Wechsel, Lower Austria.
11. Berners-Lee, T., J. Hendler, and O. Lassila, The Semantic Web, in Scientific American. 2001.
12. Linking Open Data Project. Available from: http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/.
13. Wittgenstein, L., Tractatus Logico-Philosophicus. 1921: Routledge & Kegan Paul.
14. Pasin, M., E. Motta, and Z. Zdrahal. Capturing Knowledge about Philosophy in K-CAP'07. 2007. Whistler, British Columbia, Canada.
15. Pasin, M. and E. Motta. Semantic Learning Narratives. in International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL), KCAP-05. 2005. Banff, Canada.