Formalizing and Querying a Diachronic Termino-Ontological Resource: the *CLAVIUS* Case Study

Silvia Piccini, Andrea Bellandi, Giulia Benotto
Istituto di Linguistica Computazionale "Antonio Zampolli",
Via Giuseppe Moruzzi 1, 56124 Pisa, Italy
name.surname@ilc.cnr.it

Abstract

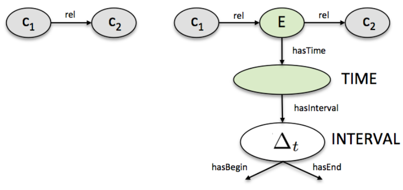
In this work, we describe the modelling of a diachronic termino-ontological resource, named *CLAVIUS*, representing the evolution of astronomical concepts and theories from antiquity until the dawn of the modern age. The resource was built by means of existing tools allowing the scholars to formalize knowledge even though they are not familiar with the models and the languages underlying the representation. More specifically, *Protégé*, a free open-source ontology editor, which supports OWL (and OWL 2) and *Chronos*, a plug-in for *Protégé* to manage temporal aspect, were used. A raw evaluation of the resource is provided by means of a controlled natural language interface, which enables scholars to answer a set of salient queries defined by our domain expert.

1 Introduction

Concepts, as well as the words used to evoke them, are subject to the inexorable law of change. This is particularly the case for the history of science. Over the centuries scholars have built different theoretical models in response to the continuous innovation that emerged from observation sometimes producing a real scientific revolution in the world view. Needless to say, a change in conceptual level often corresponds to a change in terminology: new terms can be introduced to express the new system of concepts while old terms can be dismissed when the concept becomes obsolete or can refer to new concepts.
Technology can come to the aid of scholars in their endeavors with textual hermeneutics and information retrieval. The key point is that the concepts evoked within the text, as well as the terms representing these concepts, need to have a structured organization, and to be explicitly and univocally defined through the relationships that unite them.[1] In other words, the knowledge conveyed by a text needs to be represented in a termino–ontological resource where the ontology is connected to a lexical component.
In the age of the Semantic Web it is important that these resources are built according to the technologies of Semantic Web and Linked Open Data, in order to enable interoperability so that they can be shared, and reused across scientific communities (Ciotti, 2014).
In this work we present the modelling of a diachronic termino-ontological resource, named *CLAVIUS*, devoted to the astronomy domain in a time span ranging from antiquity until the dawn of the modern age. *CLAVIUS* was built within the Project Clavius on the Web[2] and got its name from Cristophorus Clavius, a German Jesuit mathematician and astronomer, who was one of the most respected and influential scholars of his time. The texts he wrote were widely used within Europe to teach astronomy and were studied even by scholars such as René Descartes, Marin Mersenne and Johannes Kepler (Lattis, 1994).
The resource was built on the basis of Clavius' commentary on Sacrobosco's *De sphaera mundi*, both of which are considered highly influential works of pre-Copernican astronomy in Europe. In his 500-hundred page tome the Jesuit father describes and comments on the tenets of ancient and medieval astronomy, most often upholding the traditional standpoint. In the commentary, the history of scientific thought is described from Platonic-Aristotelian theories to the early *novitates astronomicae*, that pave

[1]Although the conceptual and the terminological layers are intimately linked, the theoretical necessity of distinguishing between them led to the development of new paradigms (Roche, 2007), and strategies (Reymonet et al., 2007).
[2]*http://claviusontheweb.it/*

Relation	Temporalization
$c_1\ rel\ c_2$	$\exists\Delta_t.\ c_1\ rel\ E, E\ rel\ c_2$



Property	Inferences
Reflexive	$\exists\Delta_t.\ E\ rel\ E, c_1\ rel\ c_1, c_2\ rel\ c_2$
Symmetric	$\exists\Delta_t.\ E\ rel\ c_1, c_2\ rel\ E$
Transitive	$\forall\Delta_t.\ c_1\ rel\ c_2$
Inverse	$\exists\Delta_t.\ E\ rel^{-1}\ c_1, c_2\ rel^{-1}\ E$

Figure 1: Examples of inferences about a relation rel by means of n-ary model.

the way for the modern era and signal the end of the medieval world view.

The remainder of this paper is organised as follows: in Section 2, we will present the main approaches for representing temporal aspects in ontologies; in Section 3 we will describe the *CLAVIUS* resource, and in Section 4, we will present a controlled natural language interface able to answer a set of salient queries defined by our domain expert. Finally, in Section 5 some conclusions will be drawn.

2 Approaches to Represent Temporal Aspects

In his commentary on Sacrobosco's *De Sphaera*, Clavius illustrates the evolution of the authors' conceptualization from the antiquity until the dawn of the modern age. In order for scholars to access the semantic content of the text through "sophisticated" queries, the evolution of astronomical concepts - as well as of the terms adopted to evoke them - need to be formalized into a dynamic and temporal ontology. In literature, the problem of representing the dynamically evolving information in ontologies has been addressed by adopting several different approaches (Flouris et al., 2008). One very simple solution is to create a version of the ontology for each temporal event as described in (Grandi and Scalas, 2009) (ontology versioning). For this, an ad-hoc versioning algorithm is developed in order to access the different temporal variants. Other solutions are the reification approach, the n-ary model, and the 4D-fluent approach. The first suffers from data redundancy (Batsakis and Petra, 2011), and offers limited OWL reasoning capabilities (Welty et al., 2006). The n-ary model represents a relation as two properties, each one related with a new object. These two objects are linked to each other with an n-ary relation. This approach requires only one additional object for every temporal interval, maintains property semantics but suffers from data redundancy in the case of inverse and symmetric properties (Noy and Rector, 2006). Concerning the 4D-fluent approach, described in (Welty et al., 2006), concepts in time are represented as 4-dimensional objects with the 4^{th} dimension being the time. However, all the proposed approaches lead to a massive proliferation of objects, making reasoning and querying unnecessarily complex, expensive, and error-prone. This is due to the underlying data structure, the RDF triple (Krieger, 2010). Consequently, each model presents advantages and disadvantages. The choice of a model is linked to the specific needs for representation, querying and reasoning. In the following, we will show the model adopted in *CLAVIUS* and we will give reasons for our choice.

3 Formalizing the Astronomical Domain in *CLAVIUS*

The two key points determined our approach to modelling the static and dynamic knowledge conveyed in Clavius' texts have been i) the use of standard Semantic Web and Linked Data technologies, and ii) the possibility to use existing tools, which allow the scholars to formalise knowledge even though they are not familiar with the models and the languages underlying the representation.

Concerning i), *CLAVIUS* was coded in OWL (Web Ontology Language), which is the family of knowledge representation languages for authoring ontologies. As regards ii), the resource was built in Protégé, the most well-known free and open source editor of ontologies supporting OWL natively. The diachronic aspect was modelled using $Chronos$ (Preventis et al., 2012), a plug-in for Protégé aimed at managing temporal ontologies. It can be downloaded from the Web, and is based on n-ary relation model.

In *CLAVIUS* ontology concepts and terms were represented both as OWL classes in order to ensure

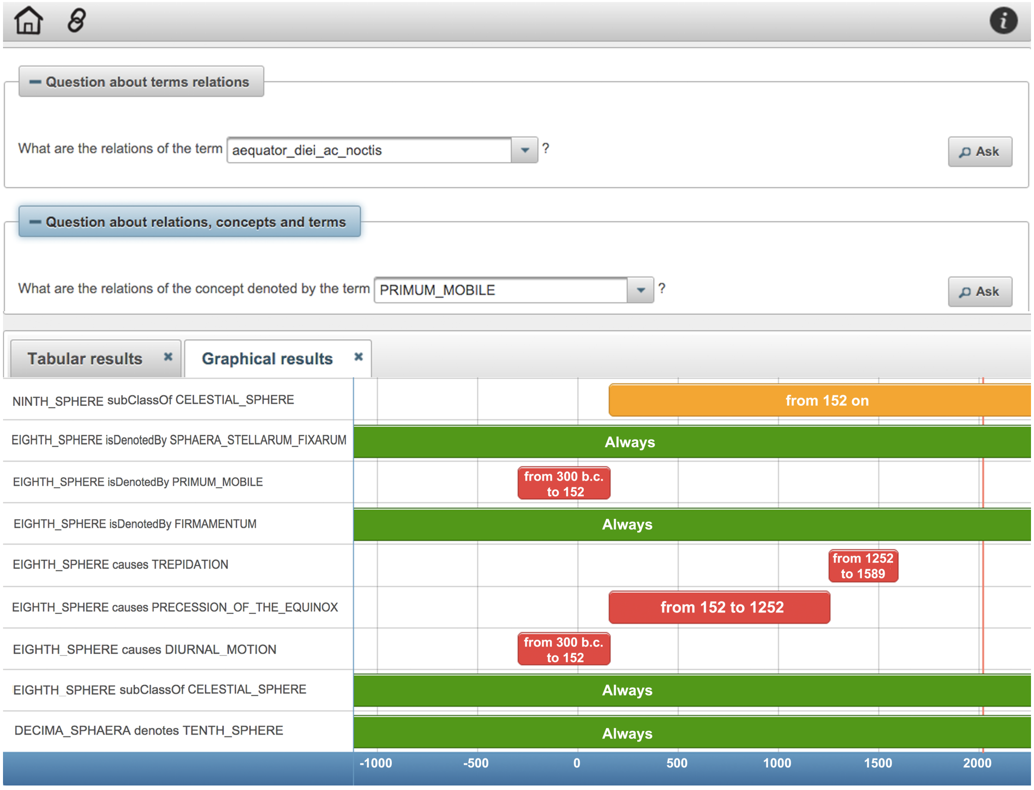


Figure 2: An example of answer to the query "*What are the relations of the concept denoted by the term Primum Mobile*" is presented. A demo version is available at: *http://146.48.93.19:8080/clavius*

autonomy of the terminological and the conceptual level. The top level of the ontology is represented by the two OWL disjoint classes CONCEPT and TERM, which subsume, respectively, all astronomical concepts and all astronomical terms. In *CLAVIUS* the conceptual level is expressed in English while the terminological level is made up of Latin words which are linked to the concepts they evoke through the relation *denotes* (and its inverse *isDenotedBy*). The ontology currently consists of 106 classes organised into four hierarchical levels, a set of 10 DataType Properties and 18 Object Properties, which make it possible to give a precise representation of the concepts and the terms. There are three basic types of relations: **lexical**, which express the paradigmatic relations among terms (hypernymy, hyponymy, meronymy, holonymy, synonymy and antonymy); **inter-level** which link the sense of a term to the concept it evokes; **conceptual** which describe the relations holding between the concepts. Among these there are **domain specific** relations, introduced to better formalize the characteristics of the astronomical domain (*isNear*, *revolvesAround*, etc.). Finally, when opportune, the relations were easily temporalized by means of the $Chronos$ editor. However, as described in Section 2, adopting the n-ary model brings about issues of reasoning. In fact, uncorrected or redundant triples are inferred, w.r.t. the temporalized properties, in particular when these properties are defined as symmetric, transitive or they have an inverse. In Figure 1 an example of inferences referred to a relation rel is shown. To solve such problems, it is necessary to filter out from the SPARQL queries results all the triples that were erroneously inferred or redundant. For example, as Figure 1 depicts, if rel holds in a particular time frame and is transitive, a reasoner infers that rel always holds in time.

4 Querying *CLAVIUS*

In order to give a raw evaluation of the formalised resource as well as to facilitate access for scholars, a controlled natural language interface (Schwitter, 2010) was developed to query the ontology. As illustrated in Figure 2, query templates were created, and each of them is made up of a fixed part that typifies a specific querying model and a variable part that allows the user to select an element of the ontology from the drop-down list. Question templates are processed by the software into SPARQL queries. Queries can be made in controlled natural language by taking into account the lexical level (*"What are the relations of a specific term?"*), the ontological level (*"What are the relations of a concept denoted by a specific term?"*) or both. The questions themselves could involve diachronic aspects, such as in *"What relation exists between two concepts in a specific temporal interval?"*.

In Figure 2 an example of query is provided. The term *primum mobile* is highly ambiguous as over time it has denoted different concepts. In the Aristotelian view it was the eighth sphere, the outermost sphere containing the fixed stars, also called *sphaera stellarum fixarum* or *firmamentum*. With Ptolemy, *primum mobile* became the ninth sphere, introduced to explain the precession of the equinoxes observed by Hipparchus. In the Alfonsine Tables *primum mobile* denoted the tenth sphere, which was added to explain the trepidation motion noted by the astronomer Thabit Ibn Qurra Arab [3]. In Figure 2 it is possible to see how the relations between the term *primum mobile* and the concepts it denotes have changed over time.

5 Conclusions

Formalizing the evolving knowledge conveyed by Clavius' work raised interesting challenges about knowledge representation. From a theoretical standpoint, many models have been proposed to formalize the diachronic evolution of information. Nevertheless, either these models are not supported by scholar-oriented tools or the available tools are based on approaches which lead to a massive proliferation of objects, making reasoning and querying complex and error-prone. This paper is intended as a springboard to discussion within the scientific community, in particular Digital Humanities, which increasingly feels the need to adopt Semantic Web technologies in order to create resources that can be shared, reused and built on by scholars.

References

Batsakis S., Petrakis, E.G.M. 2011. Representing Temporal Knowledge in the Semantic Web: The Extended 4d Fluents Approach. *Combinations of Intelligent Methods and Applications*, pp. 55–69, Springer, Heidelberg.

Ciotti, F. 2014. Digital Literary and Cultural Studies: State of the Art and Perspectives. *Between*, vol. 4, no. 8.

Flouris, G., Manakanatas D., Kondylakis H., Plexousakis D., Antoniou G. 2008. Ontology Change: Classification and Survey. *The Knowledge Engineering Review*, vol. 23, pp. 117–52, Cambridge University Press.

Grandi, F., Scalas M.R. 2009. The Valid Ontology: A Simple OWL Temporal Versioning Framework. 3^{th} *International Conference on Advances in Semantic Processing*, pp. 98–102, IEEE Press, New York.

Krieger, H.U. 2010. A General Methodology for Equipping Ontologies with Time. *Proceedings of LREC 2010*.

Lattis J. 1994. Between Copernicus and Galileo: Christoph Clavius and the Collapse of Ptolemaic Cosmology. *University of Chicago Press*, Chicago.

Noy N., Rector B. 2006. Defining N-ary Relations on the Semantic Web. *W3C Working Group Note*, http://www.w3.org/TR/swbpn-aryRelations

Preventis A., Marki P., Petrakis E.G.M., Batsakis S. 2012. Chronos: A Tool for Handling Temporal Ontologies in Protégé. 24^{th} *International Conference on Tools with Artificial Intelligence*, Athens, Greece.

Reymonet A., Thomas J., Aussenac-Gilles N. 2007. Modelling ontological and terminological resources in OWL DL. *The Lexicon/Ontology Interface, Workshop*, pp. 415–425, Busan (South Korea).

Roche C. 2007. Le terme et le concept: fondements d'une ontoterminologie. *Actes de la première conférence TOTh*, pp.1–22, Annecy.

Schwitter R. 2010. Controlled Natural Languages for Knowledge Representation. 23^{rd} *International Conference on Computational Linguistics*, pp. 1113–1121.

Welty, C., Fikes R., Makarios S. 2006. A Reusable Ontology for Fluents in OWL. 4^{th} *International Conference FOIS*, pp. 98–102, IEEE Press, New York.

[3]C. Clavius, *In Sphaeram Comm.*: *"Quare cum corpus simplex vnico tantum motu ferri sit aptum, ut uolunt philosophi, non potest nonum coelum esse primum mobile, sed supra ipsum erit decimum statuendum coelum, quod sit primum mobile"*.