

Leveraging a Narrative Ontology to Query a Literary Text

Anas Fahad Khan¹, Andrea Bellandi², Giulia Benotto²,
Francesca Frontini², Emiliano Giovannetti², and Marianne Reboul⁶

- 1 Dipartimento di Studi Umanistici – Università di Ca' Foscari, Venezia, Italy
and Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italy
fahad.khan@unive.it
- 2 Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italy
andrea.bellandi@ilc.cnr.it
- 3 Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italy
giulia.benotto@ilc.cnr.it
- 4 Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italy
francesca.frontini@ilc.cnr.it
- 5 Istituto di Linguistica Computazionale “Antonio Zampolli”, Pisa, Italy
emiliano.giovannetti@ilc.cnr.it
- 6 Labex OBVIL, Sorbonne Paris-IV, Paris, France
marianne.reboul@free.fr

Abstract

In this work we propose a model for the representation of the narrative of a literary text. The model is structured in an ontology and a lexicon constituting a knowledge base that can be queried by a system. This *narrative* ontology, as well as describing the actors, locations, situations found in the text, provides an explicit formal representation of the timeline of the story. We will focus on a specific case study, that of the representation of a selected portion of Homer's *Odyssey*, in particular of the knowledge required to answer a selection of salient queries, formulated by a literary scholar. This work is being carried out within the framework of the Semantic Web by adopting models and standards such as RDF, OWL, SPARQL, and lemon among others.

1998 ACM Subject Classification I.2.4 Knowledge Representation Formalisms and Methods

Keywords and phrases Ontology, Lexicon, Computational Narratology, Time, Semantic Web

Digital Object Identifier 10.4230/OASISs.CMN.2016.10

1 Introduction

In this article we present a model for the computational representation of narrative using ontologies encoded in the Web Ontology Language (OWL) with the aim of querying literary texts on a semantic basis. We have chosen to focus on Homer's *Odyssey* as a test case both because of its importance as one of the foundational texts of Western literature but also because of the great deal of research that has been already carried out on the identification of possible narrative structures in Homer's epic poem (for example see the well known study: [11]). In general most of the published literature on the *Odyssey* makes at best limited use of computational techniques and there seems to be a lack of research into how one might make such data more accessible and more useful from a computational viewpoint. We have attempted to fill this gap by bringing together methods, techniques, and tools from fields such as ontology engineering and linked data in addition to the more traditional scholarly



© Anas Fahad Khan, Andrea Bellandi, Giulia Benotto, Francesca Frontini,
Emiliano Giovannetti, and Marianne Reboul;
licensed under Creative Commons License CC-BY

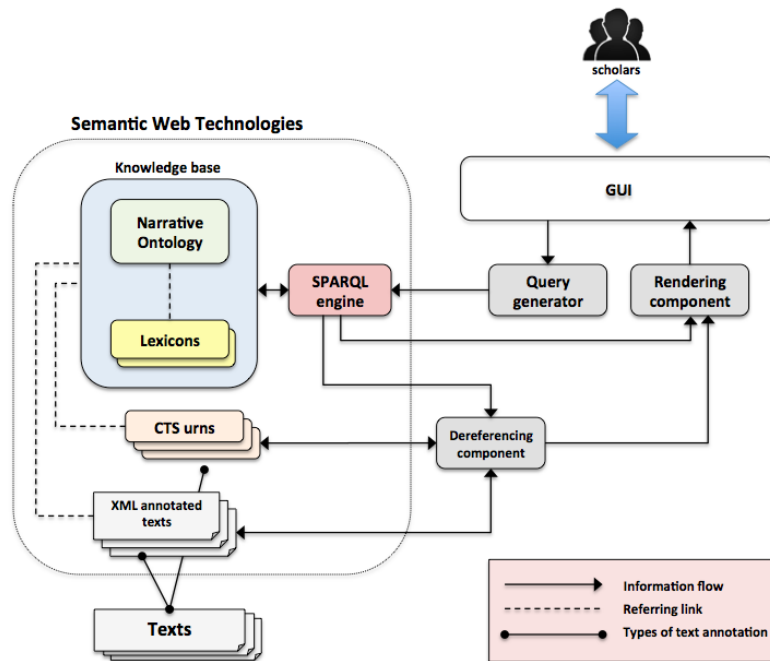
7th Workshop on Computational Models of Narrative (CMN 2016).

Editors: Ben Miller, Antonio Lieto, Rémi Ronfard, Stephen G. Ware, and Mark A. Finlayson; Article No.10;
pp. 10:1–10:10



Open Access Series in Informatics

OASIS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



■ **Figure 1** A sketch of the overall architecture of the system.

methodologies applied to the Homeric text, to create a resource that will assist and enable scholarly research on the text of the Odyssey.

Within the wider context of the Digital Humanities, machine-readable representations of literary texts are already regarded as essential for a number of purposes, amongst which those of literary and narratological research [23, 22] as well as automatic storytelling [3, 8, 20, 18]. In order to pursue these aims, many aspects of narrative have been formally modelled such as the descriptions of full narratives as logically- and temporally-ordered streams of formalized elementary events [21, 17], the notions of characters and narrative world [23, 1], stories and actions [3].

In Section 2, we will detail the construction of our Homeric “narrative” ontology and describe how it models the content of the Odyssey by including formal descriptions of events, characters, places, etc, as well as the timeline (or more accurately timelines) of the story. This ontology is only one component in a more general Semantic Web based system for studying literary texts, in this case one specific text and its translations. We chose to work with Semantic Web based technologies because they make it far easier to publish and share such scholarly work, as well as to link different datasets and resources together. The overall system can be seen in schematic form in Figure 1, and is essentially a Web application (the use of which, thus, just requires a browser) implementing an advanced querying system. In our system the ontology is linked to one or more RDF lexicons in Greek and in a number of other languages. These datasets are themselves linked to a corpus of TEI-XML files representing different editions and/or translations of the same literary work; references to the corpus are also provided as CTS-URNs¹ [16] addressing specific parts of the texts. The ontology and

¹ CTS stands for Canonical Text Services, the protocol of requests and responses that allow one to work with texts over networks.

the lexicons together constitute a knowledge base to be queried using the SPARQL Query Language. We use a graphical interface that makes use of a controlled natural language to allow users unfamiliar with the SPARQL query language to interact with the knowledge base and retrieve facts about both the narrative structure and the linguistic features of the text.

One of the core motivations behind this work was to enable the study of a literary text in its different versions and translations from several different points of view, including, but not limited to, the (inter) textual, the linguistic, as well as the semantic and narratological; as mentioned above the use of RDF-based technologies in particular facilitates the linking and comparison of different versions/translations of the same text. The system will be evaluated based on: i) how efficiently a user is able to set the kinds of queries that will help him/her answer questions relating to the different descriptions of the storyline in the text, ii) how different aspects of the story or of the protagonists are represented at various points in the text, and iii) whether the system responds correctly to the given queries. Since an evaluation of the functions of a system like this (usually including precision and recall) can be particularly challenging, we will initially ask to a group of users (students and scholars) to fill in a survey to state their opinion about system's performance.

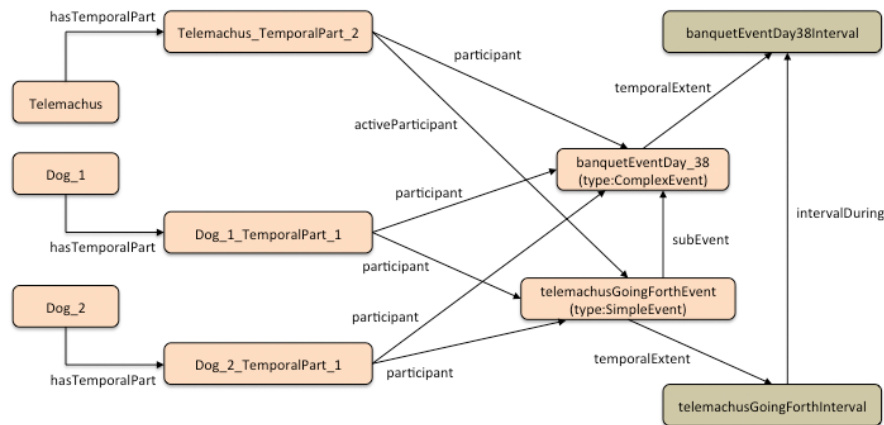
2 Case Study: Homer's Odyssey

For this case study we will only focus on a specific part of the Odyssey: that part called the Telemachy (or Telemacheia), that is to say books I to IV, and books XVI to XXIV. At the beginning of the Odyssey, Telemachus is portrayed as a young prince who is not yet capable of ruling Ithaca; and yet at the end of the poem not only has he acquired "regal" characteristics, but he has begun to prove himself by acting like a true ruler.

2.1 The ODY-ONT Knowledge Base

The technical challenges of representing dynamically evolving information in the Web Ontology Language (OWL), which being based on the RDF model permits only unary and binary relations and thus makes it difficult to add temporal arguments, led us to adopt the perdurant based approach developed in [19] in order to model the situations and changes described in the narrative (we rejected other approaches to add temporal information, such as ontology versioning [7], reification, n-ary relations [12] as being less intuitive and harder to work with). In our perdurant based approach, it is a timeslice of each protagonist that participates in an event or process unfolding over time; these timeslices can be seen as bundles of properties that are stable over a given period of time (the temporal extent of each timeslice), at least in our model of the narrative. Using this approach we are able to represent protagonists in the plot of the Odyssey as possessing a given attribute at a certain interval of time and not at other times. We used the vocabularies OWL-TIME and TIME-PLUS as a basis for the temporal parts of ODY-ONT ontology; we used the Proton ontology (<http://ontotext.com/products/proton/>) as an upper level for ODY-ONT. Proton is a basic upper-level ontology providing coverage of the general concepts necessary for a wide range of tasks, including semantic annotation, indexing, and retrieval of documents. It contains about 300 classes and 100 properties and it is formalised in OWL-Lite.

The diagram in Figure 2 represents the event of Telemachus' passing through the corridor ("striding forth") purposefully. Notice that it is a temporal part of Telemachus that participates in the event. ODY-ONT captures all the events (up to a given granularity) described in the Telemachy as well as the main protagonists in the text and some necessary background knowledge not explicitly stated in the text (e.g., Ithaca is a place). The first stage in the



■ **Figure 2** The perdurantist approach. Example from Book XVII of the Odyssey.

creation of ODY-ONT was the identification of the pertinent events in the text. Afterwards, when we had created an ontology individual for each event in the narrative we were able to tag an XML copy of the original Greek text and add the ontology ID for each event referred to in the text, we were also able to do the same for the French translations too. The granularity of the events represented was determined in large part by the kinds of queries that we determined would be interesting and useful both for researchers from a literary or classics background as well as for students and educators. We represented a simple event as an event which is identified by means of exactly one verb or verbal phrase; a complex event is a sequence of many simple events. It is important to point out that events in a story may be ordered in terms of the story timeline, i.e., where they appear in the story described the narration, as well as the narrative timeline, i.e., where they appear in the narration; for the Telemachy these two tend to coincide on the whole.

Figure 3 presents an excerpt from the annotated texts, first in Greek, then in French. The passage in question, from Book I of the Odyssey, describes a first meeting between Telemachus and Athena and is essential in the Odyssey’s narrative as it is the first time in the Odyssey a god actually meets a mortal, who in this case will be there explicitly to guide him throughout his adventures. Athena is traditionally the goddess who kept a close watch on Odysseus in the Iliad, and is now about to do the same for his son: the audience immediately understands that Telemachus is about to “grow up” and become the real ruler that he was destined to become. The use of the verb $\tau\delta\epsilon$ shows how Telemachus is first and foremost a resigned character, confined to passive expectations, and that he needs to gain control, and learn how to master his own house and country. We link entities such as characters and places to both DBpedia and Wikidata entries (when available). Telemachus of Figure 2, for example, is linked to URIs <http://dbpedia.org/page/Telemachus> and <https://www.wikidata.org/wiki/Q192482>. Furthermore, we link items in the ontology with the fragments of text mentioning those items. Currently we use both CTS URN’s to address specific portions of the Greek text of the Odyssey as well as X-Pointer to refer to TEI-XML versions of some selected French translations and the original Greek annotated with elements from the ontology [2]. Note that unlike some other similar works in textual narratology [5] we have chosen to (temporarily) put aside all the issues related to discourse analysis and, instead, we chose to annotate the events and predicates in the text and to focus on the modeling of a much richer ontological

<pre> <ody-Event> τιν δὲ πρῶτον <ody-pred>ἴδε</ody-pred> Τηλέμαχος <ody-attribute>θεοειδής</ody-attribute> </ody-Event> , <ody-Event> <ody-pred>ἦστο</ody-pred> γὰρ ἐν μνηστῆρσι <ody-Event> φίλον <ody-pred>εἰρημένον</ody-pred> ἦτορ </ody-Event> ... </pre>	<pre> <ody-Event> Et , le premier de tous , le <ody-attribute>divin</ody-attribute> Télémachos <ody-pred>vit</ody-pred> Athènes </ody-Event> . <ody-Event> Et il <ody-pred>était assis</ody-pred> parmi les Prétendants , <ody-Event> le cœur triste </ody-Event> ... </pre>
--	--

■ **Figure 3** Custom XML tags used for annotating the text.

level that includes information on the structure of the narrative and the individual events described in the text. For instance, we codified the temporal sequence of the events in the ontology instead of annotating the text with, for example, TimeML tags. Similarly, in the present version of the work we were not interested in techniques for the automatic or semi-automatic annotation of narrative, as discussed in [6, 14].

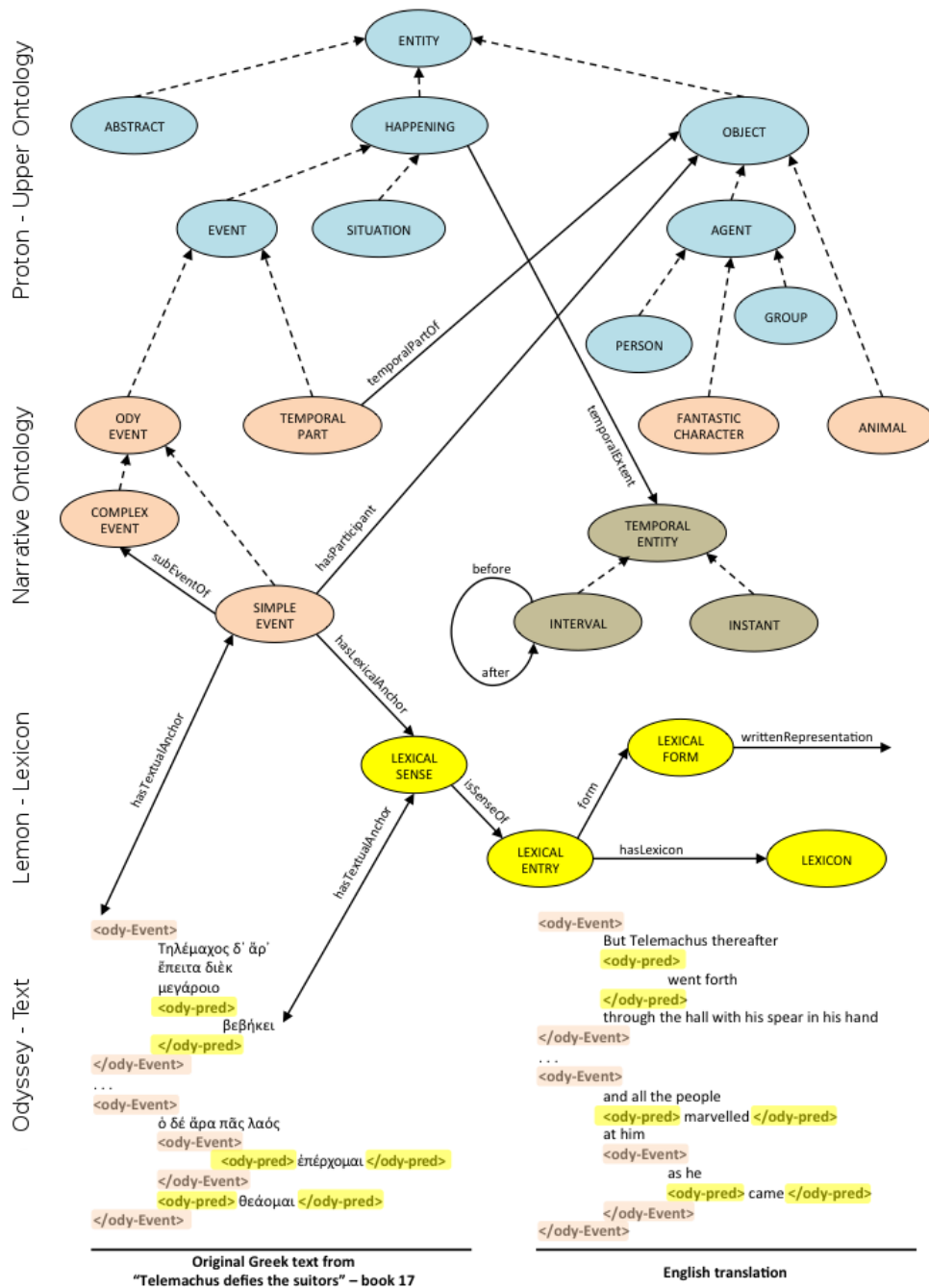
In order to add linguistic information to our system we have also included a lexicon incorporating information from an abridged version of the well known Liddell-Scott Greek lexicon in the dataset: we have converted this lexicon into RDF using the lemon model (<http://lemon-model.net/>). Lemon is a model for representing lexico-semantic data on the Semantic Web. It is heavily influenced by previous computational lexicographic models such as LMF [10]. It has already been used to model a number of other important lexical resources such as the Princeton Wordnet [9], Framenet and Verbnet. We link Greek words tagged in the texts to the relevant senses in the lexicon. We also plan to include RDF French lexicons in the near future. The overall representation model is depicted in Figure 4.

2.2 Querying the ODY-ONT

We designed our system in collaboration with a scholar in the field to ensure that it could answer at least three pertinent types of queries, each differing with respect to the type of knowledge involved and each corresponding to a different typology of research question. The first set of questions involve the events in the Telemachy without reference to their temporal sequence. A possible example relating to our test topic would be to search for all interactions between humans and gods, this for example, would enable us discover the character who interacts the most with divine entities, apart from Odysseus is Telemachus himself.

The corresponding query deals both with the ontological layer as well as with the text. In particular all instances of characters that take part in a dialogue (formalised as a type of event) with a divinity are picked out, together with the instances of the dialogue itself: for each one the system should return the relevant sections of text using the CTS-URN scheme. Indeed, it should be possible to reach the relevant part of the text from each event through the CTS-URN process. For example, in order to access the global section where the action defined by the verb ἴδε (cf. Figure 3) is described, the following CTS-URN can be used: <http://data.perseus.org/citations/urn:cts:greekLit:tlg0012.tlg002.perseus-grc1:1.80-1.124>.

10:6 Leveraging a Narrative Ontology to Query a Literary Text



■ **Figure 4** The proposed representation model of ODY-ONT. An example of representation of a simple event. Only the main classes and properties are depicted.

The second set of questions pertain to the temporal aspects of the story and the evolution of the plot and characters. One such query might look into Telemachus' role in the Odyssey, his status as an actor in the story, and in particular which verbs are used to refer to his actions. Our system allows researchers to easily search out a variety of verbs, and to therefore learn about Telemachus' actions. Users can then track the changes, within the narrative

■ **Listing 1** SPARQL code

```
SELECT ?event ?strippedForm ?writtenRepresentation ?cts_urn
WHERE { ?event ody-ont:hasActiveParticipant ?p .
        ?p ody-ont:temporalPartOf ody-ont:Telemachus .
        ?event ody-ont:hasLexicalAnchor ?sense .
        ?event ody_ont:hasTextualAnchor ?cts_urn .
        ?sense praclex:strippedForm ?strippedForm .
        ?sense lemon:isSenseOf ?le .
        ?le lemon:form ?lf .
        ?lf lemon:writtenRep ?writtenRepresentation }
```

structure, in the use of verbs, and the evolution in Telemachus' characteristics throughout the story.

Finally, a specialist in traductology may be interested in seeing the evolution of translation motifs throughout the centuries as well as for instance the evolution of the characters. He/She may therefore want to examine the differences between translations of the verbs in the above query group for our Telemachus case study, and if the translators of the text were able to identify and reproduce the changes in Telemachus' behaviour. The scholar can then measure both an internal evolution, within one translator's text, and an external evolution, within a wider chronological scale, comparing different translators with each other. This kind of query would exploit all the resources of our system, by involving the ontology, the Greek lexicon, the French lexicons, as well as the source text and its translations. As a matter of fact, our system is equipped with an interface to help scholars produce complex queries to answer to complex queries (as, for example, the latter) in a more efficient and effective way.

2.3 Query Interface

Accessing structured data in the form of ontologies requires training and learning formal query languages which poses significant difficulties for non-expert users. In Listing 1, we show a SPARQL query related to the second question presented above; for reasons of lack of space we omit the other queries. The query foresees the interrogation of events that deal with actions performed by Telemachus, i.e., those actions for which Telemachus is the agent (or active participant). This type of query may involve both the Greek text and the Greek lexicon, in order to retrieve the verbs used to describe the actions and to visualize them as a list, which is ordered on the basis of the story or the narration.

One way of lowering the learning overhead and making ontology queries more straightforward to formulate is through a Natural Language Interface (NLI) [4]. In order to make our resource easily accessible to scholars, we are developing a controlled natural language interface [15] for querying the ontology. Research on the text can be performed by taking into account the lexical level, the ontological level or both. The solution we provide in this work refers to the creation of query templates [13]. By means of these, different types of queries can be carried out, using the menus inside each template. The end user can then put together a natural language question, guided by the chosen model. Figure 5 shows an example of our interface related to the query above presented. As illustrated, each template is made of a fixed part that typifies a specific querying model and a variable part that allows the user to choose a specific element of the ODY-ONT knowledge base from the drop-down list. Question templates are processed by the software and converted into SPARQL queries aimed at interrogating the ontology, providing the user with the desired answers. Concerning

The screenshot shows a web interface for querying a narrative ontology. At the top, there are three tabs: "Question about events" (selected), "Question about lexicon", and "Question about translations". The main query area contains the text: "What are the **simple** events in which **Telemachus** is **active** participant and their related **verbs**?" Below this is an example: "Example: What are the simple events in which Telemachus is active participant and their related verbs?". A dropdown menu is open over the word "Telemachus", listing names: Athena, Circe, Calypso, Penelope, Nausicaa, Alcinous, Tiresias, Polyphemus, and Aeolus. Below the query area are three tabs: "Question about events", "Question about lexicon", and "Question about translations".

EVENT NAME	GREEK TEXT	VERB - Lemma	VERB - Stripped Form	TEMPORAL INFO
telemachusThreateningEvent	Τηλέμαχος δ' ἐτέρωθεν ἀπειλήσας	γέγωνα	epic perf. with pres. signif., pluperf. used as impf., to call out so as to be heard, ὄσσον τε γέγωνε βοήσασ	Day 39 - after seeingEvent, after telemachusThreateningEvent
telemachusGoingForthEvent	Τηλέμαχος δ' ἄρ' ἔπειτα διέκ' μεγάροιο βεβήκει	βαίνω	intr. to walk, step, properly of motion on foot, ποσσὶ or ποσὶ βαίνειν Hom., etc.; c. inf. in Hom., βῆ ἴεσαι, βῆ ἴεμεν set out to go, went his way, II.; βῆ θέειν started to run, id=II.; βῆ δ' ἔλααν, id=II., etc.:‐c. acc.	Day 38 - after seeingEvent, before telemachusThreateningEvent
seeingEvent	τὴν δὲ πολὺ πρῶτος ἶδε Τηλέμαχος θεοειδής	εἶδον	to see, perceive, behold, Hom., etc.; after a Noun, θαῦμα ἰδεσθαι a marvel to behold, II.; οἰκτρόσ ἰδεῖν Aesch.	Day 1 - before telemachusGoingForthEvent, before telemachusThreateningEvent

Below the table is a timeline visualization with three rows corresponding to the events. Green squares indicate temporal relations: seeingEvent is before telemachusGoingForthEvent, and telemachusGoingForthEvent is before telemachusThreateningEvent.

■ **Figure 5** Final user GUI. Running example of the query “What are the simple events in which Telemachus is agent and their related verbs”?

the query of Figure 5, the system retrieves three simple events and for each event it returns the lemma of the verb representing the action, the related text snippet accessed by means of the appropriate CTS-URN, the stripped form of the verb, and the available temporal information. The inference engine uses this information to compute the temporal relations among the retrieved events, and the system draws the appropriate timeline.

3 Conclusions

We feel that the computational representation of the narrative of the text has not yet been adequately addressed in the context of the Semantic Web, despite the fact that linked data technologies are now mature and well-established. In this paper we have proposed a system for the querying of a literary text that integrates existing semantic web technologies. The open issues are still numerous and we intend to address them (in future publications) by carrying on the construction of the ODY-ONT resource (currently being implemented using Stanford’s Protégé application), developing the relative lexicons and by carrying out the implementation of the user interface.

References

- 1 Fabio Ciotti. Toward a formal ontology for narrative. *Matlit: Revista do Programa de Doutorado em Materialidades da Literatura*, 4:29–44, 2016. doi:10.14195/2182-8830_4-1_2.

- 2 Fabio Ciotti, Maurizio Lana, and Francesca Tomasi. TEI, ontologies, linked open data: Geolat and beyond. *Journal of the Text Encoding Initiative*, 8, 2014. doi:10.4000/jtei.1365.
- 3 Rossana Damiano and Antonio Lieto. Ontological Representations of Narratives: a Case Study on Stories and Actions. In Mark A. Finlayson, Bernhard Fisseni, Benedikt Löwe, and Jan Christoph Meister, editors, *2013 Workshop on Computational Models of Narrative*, volume 32 of *OpenAccess Series in Informatics (OASICs)*, pages 76–93, Dagstuhl, Germany, 2013. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/OASICs.CMN.2013.76.
- 4 Danica Damljanovic, Valentin Tablan, and Kalina Bontcheva. A text-based query interface to OWL ontologies. In *The 6th International Conference on Language Resources and Evaluation (LREC)*, pages 205–212, 2008.
- 5 David K. Elson. *Modeling Narrative Discourse*. PhD thesis, Columbia University, 2012.
- 6 Mark A. Finlayson. *Learning narrative structure from annotated folktales*. PhD thesis, Massachusetts Institute of Technology, 2011.
- 7 Fabio Grandi and Maria R. Scalas. The valid ontology: A simple OWL temporal versioning framework. In *Advances in Semantic Processing, 2009. SEMAPRO'09. Third International Conference on*, pages 98–102, 2009. doi:10.1109/SEMAPRO.2009.12.
- 8 Michael O. Jewell, K. Faith Lawrence, Mischa M. Tuffield, Adam Prugel-Bennett, David E. Millard, Mark S. Nixon, Nigel R. Shadbolt, et al. Ontomedia: An ontology for the representation of heterogeneous media. In *In Proceedings of SIGIR workshop on Multimedia Information Retrieval*, 2005.
- 9 John P. McCrae, Christiane Fellbaum, and Philipp Cimiano. Publishing and Linking WordNet using lemon and RDF. In *Proceedings of the 3rd Workshop on Linked Data in Linguistics*, 2014. URL: <https://github.com/jmccrae/wn-rdf-paper/raw/master/rdf-wordnet.pdf>.
- 10 John P. McCrae, Dennis Spohr, and Philipp Cimiano. Linking Lexical Resources and Ontologies on the Semantic Web with lemon. In *Proceedings of the 8th Extended Semantic Web Conference*, pages 245–249, 2011.
- 11 Glenn W. Most. The structure and function of Odysseus' Apologoi. *Transactions of the American Philological Association (1974-)*, 119:15–30, 1989.
- 12 Natalya Fridman Noy and Alan Rector. Defining N-ary relations on the semantic web, 2006.
- 13 Borislav Popov, Atanas Kiryakov, Damyan Ognyanoff, Dimitar Manov, and Angel Kirilov. KIM—a semantic platform for information extraction and retrieval. *Natural Language Engineering*, 10(3-4):375–392, 2004. doi:10.1017/S135132490400347X.
- 14 Elahe Rahimtoroghi, Thomas Corcoran, Reid Swanson, Marilyn A. Walker, Kenji Sagae, and Andrew Gordon. Minimal narrative annotation schemes and their applications. In *Seventh Intelligent Narrative Technologies Workshop*, 2014.
- 15 Rolf Schwitter. Controlled natural languages for knowledge representation. In *23rd International Conference on Computational Linguistics*, pages 1113–1121. Association for Computational Linguistics, 2010.
- 16 Neel D. Smith and Christopher W. Blackwell. Four URLs, limitless apps: separation of concerns in the Homer multitext architecture. *Donum natalicium digitaliter confectum Gregorio Nagy septuagenario a discipulis collegis familiaribus oblatum: A Virtual Birthday Gift Presented to Gregory Nagy on Turning Seventy by His Students, Colleagues, and Friends*, 2012.
- 17 Mischa M. Tuffield, Dave E. Millard, and Nigel R. Shadbolt. Ontological approaches to modelling narrative, 2006. URL: <http://eprints.soton.ac.uk/261962/>.

- 18 Jasper R. R. Uijlings. Designing a virtual environment for story generation. Master's thesis, University of Amsterdam, 2006.
- 19 Chris Welty and Richard Fikes. A reusable ontology for fluents in OWL. In *Proceedings of the 2006 Conference on Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006)*, pages 226–236, Amsterdam, The Netherlands, 2006. IOS Press.
- 20 Dov Winer. *Language, Culture, Computation. Computing of the Humanities, Law, and Narratives: Essays Dedicated to Yaacov Choueka on the Occasion of His 75th Birthday, Part II*, chapter Review of Ontology Based Storytelling Devices, pages 394–405. Springer Berlin Heidelberg, 2014. doi:10.1007/978-3-642-45324-3_12.
- 21 Gian Piero Zarri. Representation and management of complex 'narrative' information. In *Part II of Essays Dedicated to Yaacov Choueka on Language, Culture, Computation. Computing of the Humanities, Law, and Narratives – Volume 8002*, pages 118–137. Springer-Verlag New York, Inc., 2014. doi:10.1007/978-3-642-45324-3_8.
- 22 Amélie Zöllner-Weber. *Noctua literaria : a computer-aided approach for the formal description of literary characters using an ontology*. PhD thesis, Bielefeld University, 2008.
- 23 Amélie Zöllner-Weber and Andreas Witt. Ontology for a formal description of literary characters. In *Proceedings of Digital Humanities 2006*, 2006.