Perseids and Plokamos: Weaving pedagogy, data models and tools for social network annotation

Marie-Claire Beaulieu
marie-claire.beaulieu@tufts.edu
Tufts University, United States of America

Frederik Baumgardt
frederik.baumgardt@tufts.edu
Tufts University, United States of America

Bridget Almas
bridget.almas@tufts.edu
Tufts University, United States of America

The Perseids Project is developing a platform on which students and scholars engage in collaborative acts of scholarship and research on ancient texts (Almas & Beaulieu, 2016). A core value of the project is the focus on pedagogy and the development of undergraduates as researchers. This is complemented by an emphasis on reuse and sharing of tools, data and resources. We keep these values in mind as we develop infrastructure to support complex workflows for the production of new forms of digital publications that are both machine-actionable and human-understandable. In this paper we describe one specific research activity undergraduate students have been conducting on Perseids, the annotation of the social networks of mythological characters. We discuss how opportunities and challenges, both pedagogical and technical, have presented themselves throughout multiple iterations of this effort, and how we evolved the architecture, information structures, and pedagogical workflows in response. We will use our findings to guide future decisions on when to build or reuse tools, and to formulate empirically founded recipes and approaches for specific user scenarios and data types.

The social network annotation project was motivated by an interest in teaching how to produce interpretations of mythological figures and texts. As explained by Schacht, annotation is an activity that is well known to produce deep engagement with a text in the form of close reading while promoting collaboration and conversation among students (Schacht 2016). In this case, we needed to produce interpretations that would be anchored in the primary materials and allow for a representation at the conceptual level. We decided to annotate *Smith's Dictionary of Greek and Roman Biography and Mythology* (Smith's), which offers both a complete narrative for each figure and references to the primary sources on which the narrative is based. This allowed for a double learning outcome. For instance, students would observe that Scylla is directly connected only to first and second generation Titans who represent monstrous or rebellious aspects of nature such as Typhon (volcanoes) and Charybdis (whirlpools). In addition, by following and researching the references to the primary sources, students would note that ancient texts characterize Scylla with words such as "rabid", "aggressive", and "deadly". In this way, students learned that mythological genealogies and social connections are the links which the Greeks made between different concepts represented by the mythological figures. By studying the words with which ancient texts characterize mythological figures, the students understood the value (positive or negative) associated with these concepts in Greek culture.

As we always look to reuse rather than build from scratch when possible, we developed an annotation workflow for this activity using Hypothes.is, an existing open source annotation tool (on the use of Hypothes.is in the classroom, see Kennedy 2016) . We also selected the Standards for Networking Ancient Prosopographies (SNAP) ontology for representation of the social network in the annotations, and the Open Annotation (OA) data model for serialization of the data (Sanderson et al. 2013a) Hypothes.is lacked support for controlled vocabularies, but offered free-form text entry as well as tags, worked on any website, and provided an API for retrieval of the annotations. We prepared explicit tagging instructions for the students with rules that would enable us to apply the controlled terms and data model to the annotations. Students submitted lists of their annotation URIs to the Perseids platform for ingest, review and publication of the data. Perseids software retrieved the students' data from the Hypothes.is API, and upon ingest, applied a transformation, producing OA-compliant annotations using the SNAP ontology. Once the annotations were approved by reviewers in Perseids, we exported the data for final publication via the GapVis interface, to which we added a social

network visualization and support for Canonical Text Services data sources.

We completed two full annotation rounds with separate student groups using this workflow. A key finding from a review of the data from the first round was that the lack of ability to visualize the networks at the time of annotation left too much room for error in the directionality of the annotations (On the efficacy of visualization in computer-assisted learning, see Baek and Lane 1988). Despite having explicit instructions on how to identify the subject and object of the annotations, it was difficult for both the students and the reviewers to appreciate their importance without being able to see how their choices impacted the final data. We ended up, for example, with annotations which identified a mother as the son of her child. We tried to address this in the second round by providing even more precise instructions, but the same mistakes were made. Our instructions and transformation rules also became more complex because, having identified the pedagogical significance of the characterizations, we asked students to annotate them as well as the social network connections. Through this process it became clear that we were trying to use the Hypothes.is tool in a way which was very different from the use cases it was developed to support. As a result, we had a workflow which required too much focus on following complex written instructions. This detracted from the pedagogical effectiveness of the activity as well as the overall quality of the resulting annotation data.
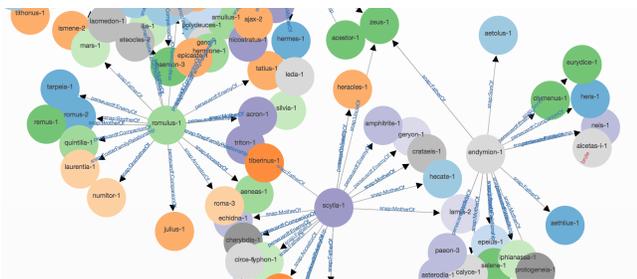
Figure 1. Plokamos network visualization based on students' annotations

At the end of this experimentation phase, we undertook a process of surgical development to address these concerns. With a much clearer understanding of our requirements, and the importance of immediate, visual feedback to the annotation and review process, we developed a targeted interface for semantic annotation which would work on any source text and allow for the data network to be visualized during the annotation process (see Fig. 1). The tool we developed - Plokamos, which is Greek for "something woven" - is a Javascript application backed by an RDF-based triple store. The Plokamos interface is also designed for reuse in other workflows and by other teams. It can be embedded into an existing application and can be extended to support other ontologies and rdf-based annotation bodies. At all times, the data itself remains separate from the tool and available for export and reuse. The configuration is also externalized from the code, and managed, along with the data, as RDF graphs. In our current deployment of Plokamos we reuse Perseids' user model, the Nemo Citable Text Services text browsing interface, and the Apache Marmotta triple store, and we continue use of the OA data model and the SNAP ontology.

We can also see the evolution and objectives of the project reflected in the underlying data structures of the annotations themselves. The annotations consist of a frame with metadata pertaining to their type, provenance and the targeted data source; linked from the frame is the annotation body containing the actual semantics of the annotation. We examine these structures at two architectural levels and from two usage perspectives.
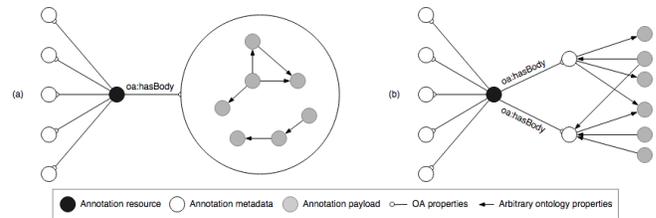

Figure 2. Graph- and Resource-based anchoring of annotation bodies

In designing the body we considered different topologies (of the connection between body and frame; and of the body itself -- structural multiplicity, see Sanderson et al, 2013b) and the compromises they represent between clarity of the annotation body and ease of traversal between annotation frame and body. An annotation body can be embedded into a distinct and uniquely named graph which is referenced by the annotation frame (see Fig. 2 (a)); or it can be anchored through one or more identifying resources which are referenced as the annotation body (see Fig. 2 (b)). The former approach enables quick and easy delineation of individual annotations and allows for complex topologies with multiple graph components. The latter approach offers less flexibility in the structure and complexity of the individual annotations but linking the payload with intermediate resources provides

easier pathways to their reuse in other graphs, queries and analyses.
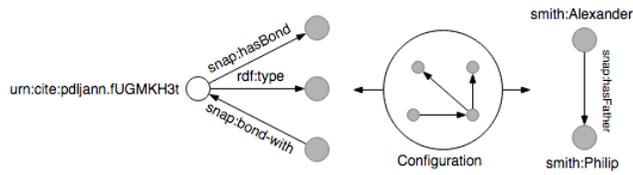
Figure 3. Transformation between machine-actionable and human-readable topologies

The need for the resulting annotation bodies to be understandable by humans as well as algorithmic processing is another factor impacting the data model. Both user groups have different requirements for the topology of the annotation data. Humans may prefer a more direct representation of the data in which object-relational structures are left implicit, while algorithms are not only indifferent to indirect construction but benefit from a more explicit and formal description of the underlying data.
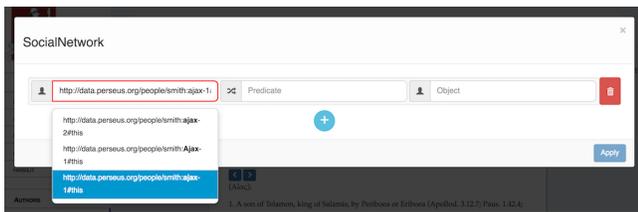
Figure 4. Annotation interface for entry of social network data

We have used the design of the Plokamos' user interface to help us mediate between these different perspectives. The interface guides the users through the annotation process with a simplified representation of entities and relations in the form of unadorned subject-predicate-object triples (Fig. 4), offering pre-configured choices to to help ensure data integrity, and we use a graph-based system of configuration to transform and expand to the more complex structures in the final annotation data.

Through this iterative approach to supporting the social network annotation activity, putting our core values of pedagogy and reuse front-and-center, we have been able to explore the pedagogical effectiveness of annotation as a learning method with a fairly low initial investment of resources. This allowed us to validate the importance of supporting this activity and refine our understanding of the architecture and data models that would be best suited to it. We were then able to approach the development

of custom tools more efficiently, while still designing for maximum extensibility and reuse. The resulting web interface with its RDF-based data source and configuration can be used on a wide variety of existing classroom resources, and expanded upon to support new use cases with varying annotation body, target types, and visual representations.

Bibliography

Almas, B., and Beaulieu, M.-C. (2016) "Scholarship for all!". *Classics Outside the Echo-Chamber: Teaching, Collaboration, Outreach and Public Engagemen*t, Gabriel Bodard & Matteo Romanello eds. Ubiquity Press. http://www.ubiquitypress.com/site/books/detail/21/ digital-classics-outside-the- echo-chamber/

Apache Marmotta (n.d.) - Home,http://marmotta.apache.org/ (accessed 10.25.16).

Baek, Y. K. and Layne, B. H. (1988). Color, graphics and animation in a computer-assisted learning tutorial lesson. *Journal of Computer Based Instruction*, 15(4), 131–135.

Kennedy, M. (2016). Open annotation and close reading the victorian text: Using hypothes.is with students. *Journal of Victorian Culture,* pages 1-9. http://www.tandfonline.com/doi/full/10.1080/13555 502.2016.1233905

Nemo. (n.d.) Nemo documentation [WWW Document], URL http://flask-capitains-nemo.readthedocs.io/en/latest/ (accessed 10.25.16).

Perseids Project: Plokamos. (n. d.) perseids-project/plokamos: RDF-based annotations for Perseids, https://github.com/perseids-project/plokamos (accessed 10.25.16).

Rabinowitz, N. (2011). GapVis: Visual Interface for Reading Ancient Texts. Available at http://perseids.org/sites/joth/#index (accessed 9.29.16).

Smith, W. *A Dictionary of Greek and Roman biography and mythology*. London. 1873.

Sanderson, R., Ciccarese, P., Van de Sompel, H. (2013a) W3C. Open Annotation Data Model. Community Draft. 08 February 2013. http://www.openannotation.org/spec/core/

Sanderson, R., Ciccarese, P., and Van de Sompel, H. (2013b) "Designing the W3C Open Annotation Data Model." ACM Press, 2013. 366–375.

Schacht, P. (2016) "Annotation". *Digital Pedagogy in the Humanities*, MLA Commons.

The Hypothesis Project. (n.d.) https://hypothes.is/ (accessed 10.25.16).

SNAP:DRGN. (n.d.) Standards for Networking Ancient Prosopographies. https://snapdrgn.net/ontology (accessed 9.29.16).