

# The World of Thucydides: From Texts to Artefacts and Back

## Matteo Romanello

King's College London, UK and German Archaeological Institute, Germany.  
*matteo.romanello@kcl.ac.uk*

## Agnes Thomas

German Archaeological Institute and University of Cologne, Germany.  
*agnes.thomas@uni-koeln.de*

### Abstract:

*The ongoing work presented in this paper is related to the Hellespont project, an NEH-DFG founded project aimed at joining together the digital collections of Perseus and Arachne. We shall discuss the design of a Virtual Research Environment (VRE) that combines archaeological data from Arachne and textual data drawn from Perseus with bibliographic information contained in JSTOR. We take as test-bed for our approach the so-called Pentecontaetia of the ancient Greek historian Thucydides (Thuc. 1,89-1,118).*

**Key Words:** *VRE, Classical Archaeology, Classical Philology, CIDOC-CRM, TEI*

## Introduction

The Hellespont project (Hellespont 2012), which involves the German Archaeological Institute and Tufts University from October 2010 to September 2013, is a case study that bridges two of the largest publicly available online databases in the field of Classical Studies: Arachne (Arachne 2012) and Perseus (Perseus 2012). Once the project is completed, the archaeological and philological evidence on any topic within these databases will be combined in a single and extensible environment. Regardless of the individual starting point of any research, primary and secondary evidence from each of the collections will be available via this bridge. Finally, merging Arachne and Perseus opens up new modes of interoperability between textual and archaeological data which can also be applied to other data collections.

What kinds of information are interesting to the historical sciences? Within the fields of the study of ancient Greece and Rome, the traditional approach to studying an ancient topic is usually to start from archaeological or philological sources, and at the same time to extract as much information as possible about the previous research on that topic, i.e. the secondary sources. With more and more information now becoming available, there is an increasing demand for global research on a specific topic, comparative studies, data transfer, data migration and data mining in heterogeneous sources. Therefore, within the broader context of the Hellespont project, the extensive and systematic integration of secondary sources within a common interface is a further aspect. Finally, the idea of connecting several types of sources in a virtual environment aims to enable quicker and less regulated access to all relevant data than is possible using traditional methods and tools.

The first part of the project, which started in January 2011, focuses on Thucydides' *Pentecontaetia* (Thuc. 1,89-1,118) as a case study and aims to show how the different components – namely the ancient text, archaeological evidence, historical background, and more specialised modern research literature – can be integrated into a single virtual environment. Our approach is partly manual and top-down, partly automatic and bottom-up. In this paper, we are presenting our first experiences of structuring and extracting different types of data. This includes the presentation of our preliminary results in modelling metadata using the CIDOC Conceptual Reference Model (CIDOC-CRM) (Doerr 2003), which enables us to integrate heterogeneous and independently developed data sources.

The thematic background of the project is the almost 50 years between the *Persian War* and the *Peloponnesian War* in Greece (479-431 BC), the main evidence being the text of the ancient historian Thucydides 1,89-1,118. The author starts his investigation from the question of how the *polis* Athens rose to hegemonic power over Greece during the 5<sup>th</sup> century BC. This hegemony caused political and military conflicts with its rival Sparta until the latter declared war on Athens in 431 BC. According to the author (Thuc. 1,23), the *Peloponnesian War* was the biggest and most destructive war ever known in ancient Greece until that moment. In only 30 chapters, Thucydides analyses the course of the political conflict between Athens and Sparta during the *Pentecontaetia* and refers to manifold and complex connections between historical persons, organisations, ancient topographies (places, cities and buildings) and historical events and activities. Because of the challenging and instructive character of the text, it is worthwhile to find a new methodological approach to the question of how to bridge different types of related information. To understand the history of this period by structuring and connecting all digitally available literary and archaeological

sources, as well as modern research literature, opens up a new scientific understanding of the topic regardless of traditional patterns of interpretation in the modern historical sciences.

The potential of integrating these two aspects into a single Virtual Research Environment should by now have become clear. Reading Thucydides' account of the *Peloponnesian War*, the user can see on the one hand the events, places and individuals that have been identified in the text, and access relevant information in the Arachne archaeological database. On the other hand, the user is provided with automatically extracted references to journal articles in JSTOR that are relevant to the text passage being read.

### **Related Work**

One aspect of the Hellespont Project, as discussed further below, is that the text of Thucydides will be structured by manual semantic annotation in order to identify and mark up the most important information in two steps. Firstly, named entities such as persons, political organisations, geographical places and built spaces are annotated and referenced with identifiers drawn from the Arachne database or from a reference source such as Smith's Dictionaries (Perseus 2012, e. g. Smith 1890). Secondly, the most important historical events and activities are annotated in order to include the historical context as described in the primary source. The result of this phase – which is described in the next section – is the production of manually extracted semantic information relating to the literary and archaeological evidence. The manual annotation, often expressed as XML markup, is a well-known and established approach to augmenting text-based resources by manually adding semantic information. In computational archaeology, the first attempts were made as early as the introduction of the XML standard (Crescioli et al. 2002), benefiting from the flexibility of the new technical solution as opposed to relational

databases, and this was continued in more recent times with the introduction of conceptual models and ontologies as a means to support to the annotation process (McAuley and Carswell 2007; Ore and Eide 2009). More recently, in the Hestia project the references to geographical places within the text of Herodotus were tagged manually and used to produce a visualisation of the text focussed on spatial relationships (Barker et al. 2010).

The other aspect of the Hellespont Project will be to mine the JSTOR archive to identify journal articles that cite the text of Thucydides as outlined in the section Extracting Related Journal Papers. Such an approach consisting of the application of Natural Language Processing (NLP) techniques to extract semantic information from unstructured texts has already been explored in relation to archaeology. Paijmans and Wubben (2007), for example, perform automatic chronological indexing of archaeological reports by applying Memory Based Learning to extract numeric and geographic features. In our specific case, finding citations of Thucydides systematically within JSTOR is possible because classical scholars over the centuries have developed a reasonably standardised way of referring to ancient texts. Such references are called canonical citations, and are essential pieces of information to be extracted because they allow us to determine which articles referred to, and possibly discussed, a specific passage of a primary source, in our case the text of Thucydides. We apply a sequence labelling algorithm drawn from Natural Language Processing in order to identify those canonical citations within the papers contained in JSTOR as described by (Romanello et al. 2009). As a result we obtain a citation network consisting both of articles citing other articles or monographs, and articles citing passages of Thucydides' work. By exploiting such a citation network it becomes possible to display relevant journal articles to a user reading a specific passage of a primary source.

### Tagging Thucydides' Text

We do not yet have an automatic tool capable of capturing passages of Thucydides' *Pentecontaetia* that are of importance to our historical and archaeological knowledge of Athens and Greece during the Classical period. Therefore we still rely on manual scholarly work on philological and archaeological sources for the identification of such 'links'. There have been previous attempts, such as (Paijmans and Wubben 2007), to automatically extract semantic information from texts, although the materials they worked on consisted exclusively of archaeological documentation. Nevertheless, the research on the automatic extraction of such information from primary sources will be worth considering at a later stage after the data model for the mapping has been developed.

Starting from the online text sources available in Perseus (Perseus xml file 2012), we use the standard of the Text Encoding Initiative (TEI 2012) to enrich the text with semantic information. We manually annotate within the text of Thucydides' *Pentecontaetia* those entities that represent categories in the archaeological and philological evidence (e.g. built spaces, geography, individual persons, populations and other organisations) i.e. all named entities. In the following overview, we will describe the TEI annotations we performed in the first testing phase of the project which ran from January to April 2011. For the moment, we have been working only with inline annotations. This is still open to changes, but should be enough as a first demonstration of what kind of information we will integrate, and in what sense this structure will be connected with a CIDOC-CRM mapping of parts of the text. An example from our TEI is shown in figure 1, taken from the paragraph Thuc. 1,89,3 in which we are informed that in the very beginning of the *Pentecontaetia* the Athenians are preparing to rebuild their destroyed city and the city walls, in the context of their return to the city from the island of Salamis and elsewhere at the end of

```
<seg xml:id="event_1-89-3-1-2" type="event" n="e25">
  καὶ τῆν
  <name xml:id="entity_1-89-3-1-2_1" key="smith:athenae-geo03 zenon:Athen allgemein"
  ref="http://arachne.uni-koeln.de/item/topographie/8002106" ana="city"
  type="topography">
    πόλιν
  </name>
  <name xml:id="entity_1-89-3-1-2_2" sameAs="#entity_1-89-3-1-1_2" type="notNoun">
    ἀνοικοδομεῖν
  </name>
  <name xml:id="entity_1-89-3-1-2_3" sameAs="#entity_1-89-3-1-1_2" type="notNoun">
    παρεσκευάζοντο
  </name>
</seg>
<seg xml:id="event_1-89-3-1-3" type="event" n="e27">
  καὶ τὰ
  <name xml:id="entity_1-89-3-1-3_1" key="athenae-geo03 zenon:Stadtbefestigungen"
  ref="http://arachne.uni-koeln.de/item/topographie/8002430" ana="city walls"
  type="topography">
    τείχην
  </name>
</seg>
```

Figure 1. Simplified TEI encoding of Thuc. 1,89,3: the Athenians prepared to rebuild their city and the city walls.

the *Persian War*. The text in bold corresponds with the annotated text of figures 1 and 3:

*Meanwhile the Athenian people, after the departure of the barbarian from their country, at once proceeded to carry over their children and wives, and such property as they had left, from the places where they had deposited them, and **prepared to rebuild their city and their walls.*** (Perseus English 2012)

Beside the names, any keywords that are related to an entity — in the given example the verbs (the Greek ἀνοικοδομεῖν παρεσκευάζοντο, “they prepared to rebuild”) as referring to a known subject from the context (“the Athenians”) — are marked up through manual co-reference resolution. Finally, in our understanding of the text, each entity lies within one or more ‘human activities’, or events, which are each described by the ancient author in a word string

(“prepared to rebuild the city” taken as the first, “and [prepared to rebuild] the city walls” taken as the second event in this example). The graph in figure 2 shows a possible and more detailed visualisation of relations between entities and events after Thucydides.

The fact that the content of the text can be segmented into single events will enable us to map the most central parts using the event-based CIDOC-CRM (see below). Therefore, in addition to named entities and further references to entities by other words, the events themselves are annotated as word strings in TEI in order to be able to refer back to the text passage from the following CRM modelling. For the purpose of this connection between the CRM and the TEI structures of the text, it is particularly important to provide each tag in TEI with its own unambiguous identifier during this part of the work.

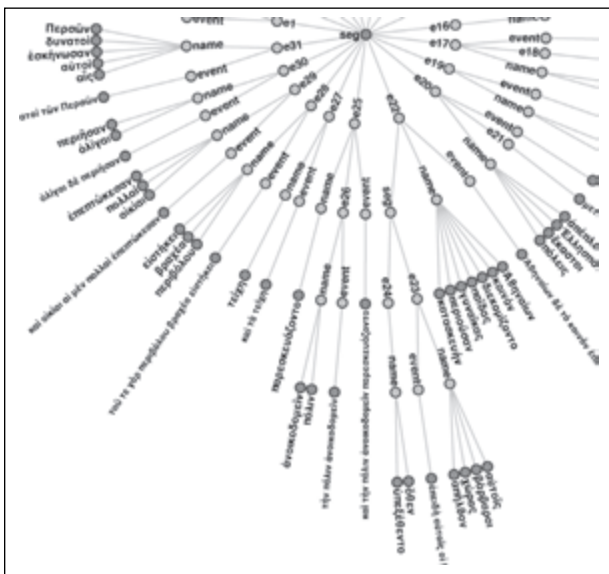


Figure 2. Possible visualisation of the semantic structure in *Thuc. 1,89* (diagram by K. Schwane).

After establishing the more formal structure of the text, meta-information is integrated with the TEI. During the first weeks of the project we tested the integration of definitions from Smith's Dictionaries as well as the related category in Zenon, the biggest online archaeological bibliography (Zenon DAI 2012), and the related entry in Arachne.

At this point, we gain a lot from a parallel project of the Arachne database. Arachne provides one of the biggest datasets of archaeological objects in a highly semantical structure with integrated meta-information. The whole database of Arachne is being mapped using the CIDOC-CRM in an on-going but almost completed project of the Cologne Digital Archaeology Laboratory (Arachne CIDOC-CRM browser 2012).

To bridge the gap between ancient text and object, the event-based CIDOC-CRM can be used to map events and logical entities like conceptual objects that we know of from Thucydides' text (Fig. 3). The CIDOC-CRM also provides via its classes a mechanism to distinguish within a text the very fact from exhibiting a fact, which is known as the

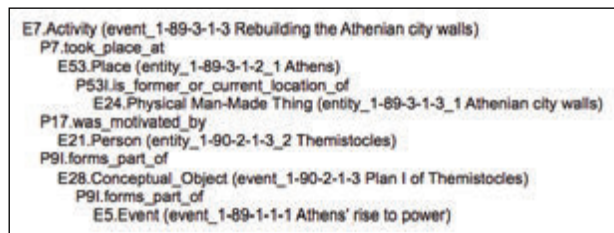


Figure 3. CIDOC-CRM example of *Thuc. 1,89*.

exhibition problem (Eide 2008). In order to integrate both annotations of the text, TEI and CIDOC-CRM, we transfer the identifiers from the TEI markup for each event and entity. In this way, each term can be easily identified within the primary text source. For a similar approach see (Ore and Eide 2009, p. 163). Furthermore, the CIDOC-CRM is able to usefully integrate even more types of information, as described in the following section of this paper.

Thus, with the manual linking of entities with Arachne entries, we are already connecting two kinds of sources, literary and archaeological. Through the mapping of the archaeological object database and the content of the literary source, using the same CIDOC-CRM we are aiming to create a common interface for searches and visualisation of these two connected areas of historical evidence. This element of the Hellespont project, as well as the integration of the secondary literature in the same interface, is still open to development.

### Extracting Related Journal Papers

#### Rationale

Devising a fully automated workflow for the tagging of Thucydides' text, as described in the previous section, is not yet an achievable goal. What can be automated, however, at least to some extent and with some degree of accuracy, is the extraction of journal papers related to a given ancient text, which is the focus of this section. One distinction commonly used in Classics is the one that exists between primary

and secondary sources, where the former are the ancient texts, and the latter are monographs, commentaries and journal articles that are written about those ancient texts.

The starting assumption here is that citations of primary sources contained within secondary sources are extremely important pieces of information for scholars, and thus deserve to be captured and exploited. Their importance lies mainly in the fact that such references allow us to determine the text passages being mentioned, and possibly discussed, in the secondary literature. Citations of primary sources can be combined with references to modern bibliographic materials to allow a network of citations between ancient and modern texts to emerge, an aspect which has not yet been considered in research on citation networks.

What we obtain after having extracted the semantics of references is a knowledge base containing simple statements such as “X cites Y”. Not only can we see when a journal article cites other articles or monographs, but we can also track the citations from journal articles to passages of ancient texts, e.g. “X cites book 8, chapter 25, paragraph 1 of Thucydides’ *Historiae*”. Such knowledge is then formalised by means of an ontology (Romanello and Pasin 2011) thus enabling some logical reasoning related, for instance, to the topology of citations (“Thuc. 8,25-8,28” includes “Thuc. 8,26” and “Thuc. 8,27”, etc.).

Once such a citation network is extracted we are able to ask our system questions such as “which journal papers mention Thucydides’ *Pentecontaetia*?” or “What are the books and articles cited by resources that mention the *Pentecontaetia*?”. In other words, extracting the journal articles related to a given passage of an ancient text allows us to add a bibliographic ‘dimension’ to our VRE. In a typical use case scenario, a user of our VRE reading a passage of the *Pentecontaetia* – which becomes the

reading context – will be shown a list of journal articles that are related to that very text passage.

### *Extracting and parsing references*

Although discussing in detail the extraction and parsing of bibliographic and canonical references is beyond the scope of this paper, let us pause for a moment to examine its basic working principles.

First of all, we shall distinguish between the two tasks of extraction and parsing. *Extracting* a reference means determining which sequence in a stream of ‘tokens’ constitutes a reference (either bibliographic or canonical), whereas *parsing* a reference implies extracting its semantics and reconstructing its internal structure. In the case of a modern bibliographic citation, the parsing leads to the identification of fields such as author, title, date of publication and so forth. In the case of parsing a canonical reference such as “Hom. *Il.* 1,1-1,10; 9,100”, the parsing allows us to determine that two distinct citations are contained, the first pointing to line 1-10 of book 1 and the second to book 9, line 100.

Given that we are using the corpus of JSTOR via the DfR API (JSTOR 2012), the phase of extracting references is already performed on the data provider’s side. However, given that such an output is often ‘noisy’ for our purposes, that is, it includes sequences that are not references (or not complete references), we still need to refine the output obtained from the JSTOR’s API by filtering out bibliographic and canonical citations from the noise caused by sequences mistakenly classified as references.

For the extraction of canonical citations we are developing a dedicated software tool-kit (CRefEx 2012) which performs this task with an F-score of 0.87, as preliminary results have shown (Romanello et al. 2009). This tool-kit is written in the programming language Python

```
<reference>Deinarchus 1.71</reference>
<reference year="1952">
  C. Hignett, A History of the Athenian Constitution [Oxford 1952] 191 and n. 7, 224 and n. 10
</reference>
<reference>Thuc. 8.25.1</reference>
<reference>Thuc. 8.47.2</reference>
<reference>Thuc. 8.63.3</reference>
<reference>Thuc. 8.73.4-6</reference>
<reference>Thuc. 8.75.2</reference>
<reference>Thuc. 8.75</reference>
<reference>Thuc. 8.76.2</reference>
<reference>Thuc. 8.82.1</reference>
<reference refid="10.2307/628231">
  A. Andrewes ("The Generals in the Hellespont, 410-407 B.C.," JHS73 [1953] 4)
</reference>
<reference year="1904">
  G. Busolt (Griechische Geschichte III2 [Gotha 1904] 1515 [hereafter Busolt, G.G.])
</reference>
<reference>Thuc. 8.99-100</reference>
<reference>Thuc. 8.101-03</reference>
<reference>Thuc. 8.104-5</reference>
<reference>Diod. 13.39.4</reference>
<reference>Diod. 13.40.4</reference>
<reference>Diod. 13.40.6</reference>
<reference>Thuc. 8.106.2, 5</reference>
<reference>Thuc. 8.107</reference>
<reference>campaign (13.40.6)</reference>
```

Figure 4. XML snippet containing the references as returned by a call to the DfR API.

and uses a machine learning approach which enables it to be trained to work on different corpora including JSTOR. For the parsing of modern bibliographic references, instead, we are using ParsCit (ParsCit 2012), another machine-learning-based open source software that has been used to build CiteSeerX, the bibliographic search engine for Computer Science literature (Councill and Kan 2008).

### The Corpus

We decided to use as a corpus of secondary sources the journal articles available in JSTOR via the Data for Research API (DfR) (Burns et al. 2009). However there are advantages and disadvantages to this approach. DfR does not provide access to the full text of the papers, but to data extracted from them. At the time of writing, the data DfR provides researchers access to are: word frequency, bigrams, trigrams, quadgrams and references (DfR documentation 2012). Being able to access for instance the references contained in a paper without being able to put them back in their context makes the DfR something of a black box. On the one hand there is no control over the algorithms used to derive those pieces of

information. But on the other hand we can easily access a corpus of 60,000 papers related to Classics, covering some 150 years, for our own purposes. Another drawback of using DfR is that this service is still in beta. There is no guarantee that data will not change, and this compelled us to devise a completely automatic and repeatable workflow in order to be able to reproduce the results whenever the input data change.

### System architecture

From a technical standpoint, the architecture of our system is built upon the Python framework Django (Django 2012), which allows us to manipulate data as Python objects and to store them persistently in a database back-end. The choice of Django is due to several factors. Being written in Python, the framework integrated nicely with the other software components that we are using for data processing. Moreover, the framework provides a set of features that eases the design of a Graphical User Interface (GUI) on top of a Django application, to create rapidly and with minimal effort a GUI-based environment to enable the manual correction of automatically annotated data. The backend of our choice is a MySQL database hosted by the UK National Grid Services (NGS 2012), thus providing our system with an infrastructure that can cope with scalability issues.

The workflow we defined consists of the following steps:

1. JSTOR data are dumped from the DfR API and stored as Django objects.
2. The data are then further processed: for example, references are split into sets of single references, and each reference is split into 'tokens'.
3. These data are then processed to extract and parse bibliographic references: the additional

information obtained at this stage is also stored as Django objects.

### Conclusions and Further Work

In this paper we have presented the preliminary results of the Hellespont project, a NEH-DFG funded project ending in 2013. Although we are still at an early stage of development, we believe that the approach we are proposing will have a revolutionary impact from a methodological point of view on both classical philology and archaeology. Artefacts, ancient texts and bibliographic information can be integrated, partly manually and partly automatically, into a single Virtual Research Environment, thus allowing us virtually to overcome the boundaries and limited perspectives often imposed by the disciplinary divisions that characterise modern scholarship.

### Acknowledgements

This work was supported by a grant from the National Endowment for the Humanities (NEH) and the Deutsche Forschungsgemeinschaft (DFG) under the bilateral Digital Humanities programme *Enriching Digital Collections*. We also want to gratefully thank the reviewers of this paper for their useful comments, as well as Karen Schwane (CoDArchLab, Cologne) and Charlotte Tupman (King's College, London).

### Bibliography

Arachne. "iDAI.Images Arachne." Accessed January 15, 2012. <http://www.arachne.uni-koeln.de>

Arachne CIDOC-CRM browser. "Arachne CIDOC-CRM browser." Accessed January 15, 2012. <http://codarchlab.uni-koeln.de>

Barker, E., Bouzarovski, S., Pelling, C., and Isaksen, L. 2010. "Mapping an ancient historian in a digital age: the Herodotus Encoded Space-Text-Image Archive (HESTIA)." *Leeds International Classical Studies* 9 (1).

Burns, J., Brenner, A., Kiser, K., Krot, M., Llewellyn, C., and Snyder, R. 2009. "JSTOR - Data for Research." In *Research and Advanced Technology for Digital Libraries*, edited by M. Agosti, J. Borbinha, S. Kapidakis, C. Papatheodorou, and G. Tsakonas, 416-419. Berlin/Heidelberg: Springer.

Councill, C. L. G. I., and Kan, M.-Y. 2008. "ParsCit: an Open-source CRF Reference String Parsing Package." In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08). Marrakech, Morocco: European Language Resources Association (ELRA)*.

Crescioli, M., D'Andrea, A., and Niccolucci, F. 2002. "XML Encoding of Archaeological Unstructured Data." In *Archaeological informatics : pushing the envelope CAA 2001 : Computer Applications and Quantitative Methods in Archaeology, proceedings of the 29th conference, Gotland, April 2001*, edited by G. Burenhult and J. Arvidsson, 267-275. Oxford: Archaeopress.

CRefEx. "GitHub. CRefEx." Accessed January 15, 2012. <https://github.com/mromanello/CRefEx>

DfR documentation. "JSTOR. Data for Research. Documentation." Accessed January 15, 2012. [http://dfr.jstor.org/??view=text&&helpview=about\\_api](http://dfr.jstor.org/??view=text&&helpview=about_api)

Django. "Django." Accessed January 15, 2012. <https://www.djangoproject.com>

Doerr, M. 2003. "The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata." *AI magazine* 24 (3).

Eide, Ø. 2008. "The Exhibition Problem. A Real-life Example with a Suggested Solution." *Literary and Linguistic Computing* 23 (1):27-37.

Hellespont. "The Hellespont Project." and "DAI. Arachne. Hellespont-Projekt: Integration von Arachne und Perseus." Accessed January 15, 2012. <http://www.arachne.uni-koeln.de/drupal/?q=de/node/230> and [http://www.dainst.org/index\\_04b6084e91a114c63430001c3253dc21\\_en.html](http://www.dainst.org/index_04b6084e91a114c63430001c3253dc21_en.html)



- JSTOR. "JSTOR. Data for Research." Accessed January 15, 2012. <http://dfr.jstor.org>
- McAuley, J., and Carswell, J. 2007. "An open Approach to Contextualising Heterogeneous Cultural Heritage Datasets." In *Proceedings of the 35th Annual Conference on Computer Applications and Quantitative Methods in Archaeology (CAA2007)*, 1-6. Berlin.
- NGS . "NGS. Connecting Infrastructure, Connecting Research." Accessed January 15, 2012. <http://www.ngs.ac.uk>
- Ore, C.-E., and Eide, Ø. 2009. "TEI and cultural heritage ontologies: Exchange of information?" *Literary and Linguistic Computing* 24 (2):161 -172.
- Pajmans, H., and Wubben, S. 2007. "Preparing Archaeological Reports for Intelligent Retrieval." In *Proceedings of the 35th Annual Conference on Computer Applications and Quantitative Methods in Archaeology (CAA2007)*, 212-217. Berlin.
- ParsCit. "ParsCit." Accessed January 15, 2012. <https://github.com/knmnyn/parscit>
- Perseus. "Perseus Digital Library." Accessed January 15, 2012. <http://www.perseus.tufts.edu/hopper/>
- Perseus, e. g. Smith 1890. "Smith, W. 1890. A Dictionary of Greek and Roman Antiquities." Accessed January 15, 2012. <http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3atext%3a1999.04.0063>
- Perseus English. "Thucydides 1,89 English." Accessed January 15, 2012. <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.01.0200:book=1:chapter=89>
- Perseus xml file. "Thucydides 1,89 xml." Accessed January 15, 2012. <http://www.perseus.tufts.edu/hopper/xmlchunk?doc=Perseus:text:1999.01.0199:book=1:chapter=89>
- Romanello, M., and Pasin, M. 2011. "An Ontological View of Canonical Citations." <http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-143.xml>.
- Romanello, M., Boschetti, F., and Crane, G. 2009. "Citations in the digital library of classics: extracting canonical references by using conditional random fields." In *NLPIR4DL '09: Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*, 80-87. Morristown: Association for Computational Linguistics.
- TEI. "Text Encoding Initiative." Accessed January 15, 2012. <http://www.tei-c.org>
- Zenon DAI. "DAI. Zenon." Accessed January 15, 2012. <http://opac.dainst.org>