Named Entity Recognition for a Text-Based Catalog of Ancient Greek Authors and Works

This paper presents a project for linguistic extraction, analysis, and annotation of the language of ancient Greek bibliographic references. The goal is the production of a structured knowledge resource about ancient Greek authors and works (i.e., names of authors and titles/descriptions of works with variants and ambiguities in the original contextual language). The project makes use of methods and technologies from NLP and computational linguistics. In particular, it applies NER for semi-automatic extraction and annotation of entities related to bibliographic references to ancient Greek authors and works.

The first part of the presentation will focus on the current state of the art of NER for ancient Greek and on the data produced in the first phase of the project, which is constituted by a full database of lemmatized and annotated named entities in the text of the Deipnosophsists of Athenaeus of Naucratis (ed. Kaibel): http://www.digitalathenaeus.org. The text of Athenaeus has been chosen because it is a reference resource for exploring language and patterns of ancient Greek bibliographic references. Moreover, its citations cover about 50% of the total number of Greek authors for the period of time between the 8th century BC and the 3rd century CE.

The second part of the presentation will show methods and tools for named entity disambiguation, linking, and coreference resolution in order to extract and annotate real entities corresponding to author names (e.g., Ἤστρος ὁ Καλλιµάχειος) and work titles (e.g., ἐν τῇ Λακεδαµονίαι Πολιτείᾳ). This work will produce a Knowledge Base for linking entity mentions to a structured vocabulary for ancient Greek authors and works that can be used to annotate other significant texts, as for example the Lexicon of the Ten Orators of Harpocration and the Suda lexicon.

The last part of the presentation will show how annotations allow to visualize references in their original context offering new and dynamic text-based tools that are not available in existing indices and catalogs of ancient Greek authors and works. This work offers new results about the language of ancient Greek bibliographic citations by documenting them with an immediate, full, and complete contextual analysis of their occurrences.

Bibliographic References