The *Geographic Annotation Platform* – a Framework for Unlocking the Places in Free-text Corpora

Elton Barker, Kate Byrne, Leif Isaksen, Eric Kansa and Nick Rabinowitz

Introduction

The original GAP (Google Ancient Places[1]) project was funded by a Google Digital Humanities Award and resulted in the creation of the *GapVis* tool.[2] GapVis provides an online interface to classical literature, with maps and data visualisations to allow the reader to follow the narrative timeline through the text and see the spatial journey through the ancient landscape at the same time.

Thanks to follow-on funding from Google, the GAP team have now embarked on a new project, to turn the tool designed for classical texts into a generic framework to use on any free-text corpus: the *Geographic Annotation Platform*. We aim to produce a toolkit that anyone can use to process a text, on any subject matter as long as there is mention of real-world places, and produce for themselves an interactive display in the style of GapVis.

Tools and Methods

We showed in the original GAP project how it is possible to identify and spatially locate ancient placenames, given a suitable gazetteer. This is the "geoparsing" step, consisting of:

1. Geotagging – finding toponym mentions in free text using Named Entity Recognition.
2. Georesolution – selecting the most likely latitude/longitude position given a set of candidates matching the toponym string.

The Edinburgh Geoparser was adapted to work with ancient placenames instead of modern, using the Pleiades[3] gazetteer. This was augmented with links to modern names in Geonames, to become Pleiades+, which allows us to aggregate data about the same place even if it occurs in the literature under different appellations, as in its 'original' form (Athenae, Roma) or later translations (Athens, Rome).

For the present project we aim to offer the user flexibility over classical or modern place mentions, by parameterising the choice of gazetteer and the supporting entity recognition.

Once the input text has been tagged with toponyms, the next stage is to transform the names into URIs, and to create a database to manage the references for web visualisation. Working with Google Books, we have generated URIs pointing directly to text snippets within the online page images of OCR-ed texts. Using techniques originally developed in the HESTIA project , networks of connections between places can be generated, allowing one to visualise the relationships between, and relative importance of, different places mentioned in the text.

The visualisation stage integrates these components and adds a timeline widget, adapted in *GapVis* to allow the user to scroll through the text, page by page, seeing the places described in the text coming in and out of focus as the narrative progresses. The database of toponym URIs permits links to be embedded throughout, allowing the user to drill down for more information on a particular place at will, or to see the connections between it and other places mentioned in the text.

Geographic Annotation Platform

Our current work revolves around taking the prototype tools developed in the original GAP project and turning them into a generic framework that can be used to process any text with real-world spatial content.

The framework will include these components:

1. The Edina *Unlock* service (see below).
2. A template database to be populated with toponyms translated to URIs.
3. A template for creating a website with the GapVis visualisation tools.

The GAP team has collaborated with Edina[4] to incorporate the adapted Geoparser pipeline into their *Unlock* services[5] for toponym exploration. We aim to incorporate the Edina APIs into our framework, so that the *Geographic Annotation Platform* will have access to a robust and persistently hosted service.

References

[1] Leif Isaksen, Elton Barker, Eric C. Kansa and Kate Byrne. (2011) Googling Ancient Places. *Digital Humanities 2011 (DH2011)*, Stanford, CA, June 2011.

[2] Claire Grover, Richard Tobin, Kate Byrne, Matthhew Woollard, James Reid, Stuart Dunn and Julian Ball (2010) "Use of the Edinburgh Geoparser for georeferencing digitised historical collections" *Philosophical Transactions of the Royal Society A*, vol 368, no. 1925, pp3875-3889, Aug 2010. ISSN:1364-503X.

[3] Elton Barker, Stefan Bouzarovski, Christopher Pelling and Leif Isaksen (2010) "Mapping an Ancient Historian in a Digital Age: the Herodotus Encoded Space-Text-Image Archive (HESTIA)," *Leeds International Classical Journal 9* , pp. 1-24.

[1] http://googleancientplaces.wordpress.com/

[2] http://nrabinowitz.github.com/gapvis/

[3] http://pleiades.stoa.org/

[4] http://edina.ac.uk/

[5] http://unlock.edina.ac.uk/