Geoparsing and Spatial Network Analysis in the GAP Projects

Elton Barker, The Open University
Kate Byrne, University of Edinburgh
Leif Isaksen, University of Southampton
Eric Kansa, University of California at Berkeley
Nick Rabinowitz, nickrabinowitz.com

Hestia2 Workshop, 18th July 2013
1 Geoparsing ancient texts
   - adapting the Edinburgh Geoparser for GAP
   - online “Unlock Text” service and other projects

2 Visualising spatial networks
   - GapVis
   - spatial connection by textual co-occurrence
The GAP projects

- Google Digital Humanities awards, 2010-11 and 2011-12
- Cross-disciplinary and multi-national team
  - humanities, classics, archaeology, natural language processing, graphical interface
- GAP – Google Ancient Places
- GAP2 – the Geographic Annotation Platform
- GapVis utility – nrabinowitz.github.com/gapvis
- Prototype “GAP in a box”: your text into GapVis
Edinburgh Geoparser
(Claire Grover and Richard Tobin, University of Edinburgh)

1. Geotagging
NLP pipeline to identify placename mentions in text.

2. Georesolution
Select candidate places from suitable gazetteer and assign probability ranking.

Originally designed for modern texts – adapted for GAP.
As the first minister, Alex Salmond, looks south and campaigns for an independent Scotland, leaders in Orkney, Shetland and the Western Isles to his north have quietly begun talks among themselves about their own “home mile”.

The three leaders, who run the three largest island groups in the British isles, will meet in Shetland on Monday 25 March to discuss a joint project on whether they should demand a split from the Scottish and UK governments after the Scottish independence referendum in 2014 – the date of which is expected to be announced in Holyrood on Thursday.

Malcolm Bell, the convener of Shetland Islands council, said the independence referendum offered an opportunity for the islands to carve out a new political settlement. “There’s no point in Westminster devolving powers to Edinburgh if they are going to stop in Edinburgh. When you’re 300 miles from Edinburgh, or 700 from London, at those kind of distances, Edinburgh
Today is 196th anniversary of the death of Jane Austen, who spent part of her life in Southampton. 16th Dec 1775 – 18th July 1817.
Pleiades: http://pleiades.stoa.org/

Graph database of ancient places, from Barrington Atlas

Pleiades+: GAP version linked to Geonames
Adaptations to make Pleiades+

Adding alternative names from Geonames

1. Match Pleiades against Geonames by hand.
2. For matched places, add valid alternative names and extra metadata from Geonames.

eg For *Autricum*, we add *Chartrez*, *Chartres* and *Shartr*
Adaptations to make Pleiades+

Adding alternative names from Geonames
1. Match Pleiades against Geonames by hand.
2. For matched places, add valid alternative names and extra metadata from Geonames.
   eg For Autricum, we add Chartrez, Chartres and Shartr

Reverse lookup for alternative names at runtime
If geotagger finds candidate placename that’s not in Pleiades, we check for alternative names in Geonames.
   eg Egypt is not in Pleiades; Geonames supplies Aegyptus as an alternative, and this is in Pleiades.
Distinguishing Places from People

- Geoparser uses various lexicons to help distinguish placenames from other entity categories
  - lexicons are used only at the geotagging step; lexicon \( \neq \) gazetteer
- Ancient texts have different “common” names – *Paris*, *Priam*, *Medea*
- Add lexicon for ancient personal names and drop modern name lexicons
Unlock the potential in your data with our simple web services

Unlock Places

Use Unlock Places to search for data about place-names. Cross-search different gazetteers and retrieve data in different machine-readable forms for use in web-based applications.

The Gazetteers used for searching include Geonames and those available from Ordnance Survey's Open Data.

- Get started
- API documentation

Unlock Text

Use Unlock Text to extract place-names from documents and find their locations. Unlock Text uses the Places gazetteer cross-search to make ranked best guesses as to the location of place names.

For use with plain text documents, XML metadata or HTML web pages. The Unlock Text web service works best with short modern texts.

- Get started
- API documentation
Collaboration with Edina – *Unlock Text*

- *Unlock Text* service: [http://unlock.edina.ac.uk/texts/](http://unlock.edina.ac.uk/texts/)
- Ongoing enhancements from GAP and other projects
- Online API –
  1. supply URLs for input text
  2. collect bundle of output files including full text, tokenised and tagged
  3. use output files as desired, eg in graphical interface

- Online documentation not yet fully up to date (see [http://googleancientplaces.wordpress.com/2012/12/21/unlocking-text/](http://googleancientplaces.wordpress.com/2012/12/21/unlocking-text/))
Other Adaptations of the Geoparser

- Trading Consequences
  http://tradingconsequences.blogs.edina.ac.uk
- Digital Exposure of English Place-names (DEEP)
  http://englishplacenames.cerch.kcl.ac.uk
- GeoDigRef
  http://edina.ac.uk/projectGeoDigRef_summary.html
- Botanical records
  http://scargill.inf.ed.ac.uk/clare/botanic.html
- SYNC3  http://www.sync3.eu

Our conclusion
Transfers across genres well but requires adaptation each time.
DEEP – creating historical gazetteer for England

- All 86 volumes of *Survey of English Place-Names*
- Edinburgh’s role:
  - convert OCR output to structured data
  - assign lat/long based on EPNS text
  - link to Geonames and to Unlock gazetteer (of UK)
- Gazetteer will be available through *Unlock* services
DEEP – creating historical gazetteer for England

- All 86 volumes of *Survey of English Place-Names*
- Edinburgh’s role:
  - convert OCR output to structured data
  - assign lat/long based on EPNS text
  - link to Geonames and to Unlock gazetteer (of UK)
- Gazetteer will be available through *Unlock* services

**Application already using it...**

Spatial Humanities [http://www.lancaster.ac.uk/spatialhum/](http://www.lancaster.ac.uk/spatialhum/) – building Lake District lexicon (modern and historic places) for Edinburgh Geoparser, that will tie to DEEP gazetteer.
1 Geoparsing ancient texts
   • adapting the Edinburgh Geoparser for GAP
   • online “Unlock Text” service and other projects

2 Visualising spatial networks
   • GapVis
   • spatial connection by textual co-occurrence
User interfaces

- Some users want raw files to use in own systems
- ...but many would prefer graphical interface
- For use in classics teaching, interface is essential
  – clear, usable presentation of content is the goal
The Genuine Works of Flavius Josephus, the Jewish historian

By Flavius Josephus

Published 1822 - View on Google Books

http://nrabinowitz.github.com/gapvis/#book/15/read/12/678106
How GapVis Works

classical texts from Google Books
How GapVis Works

classical texts from Google Books

Edinburgh Geoparser
How GapVis Works

classical texts from Google Books

Pleiades gazetteer of ancient places

Edinburgh Geoparser
How GapVis Works

- Classical texts from Google Books
- Pleiades gazetteer of ancient places
- Edinburgh Geoparser
- Database of toponym URIs – tied back to text snippets
How GapVis Works

classical texts from Google Books

Pleiades gazetteer of ancient places

database of toponym URIs – tied back to text snippets

Edinburgh Geoparser

finding relationships between places
How GapVis Works

classical texts from Google Books
Pleiades gazetteer of ancient places
database of toponym URIs − tied back to text snippets
Edinburgh Geoparser
finding relationships between places
putting it all together − GapVis
GapVis “views”

- **Book Summary** – overview of places mentioned and where they appear in the narrative
- **Reading View** – for reading the text; with map of places mentioned on current page and “timeline” of places mentioned earlier and later in the book
- **Place Detail** – focus on a particular place, with network map of related places
The History of the Peloponnesian War

By Thucydides

Published 1880 - View on Google Books

The Google Books version of "The History of the Peloponnesian War", by Thucydides, was published in 1880. It references 213 identified ancient places. The place most frequently referenced is Athens, followed by Hellas, Peloponnese, and Lacordailer/Laconia.
The History of the Peloponnesian War

By Thucydides

Published 1818  View on Google Books

E. I. INTRODUCTION. 7 It is farther my opinion, that the assemblage of that ar- mament by Agamemnon was not owing so much to the at-tendance of the suitors of Helen in pursuance of the oaths they had sworn to Tyndarus, as to his own superior power. It is related by those who received from their ancestors the most certain memorials of the Peloponnesian affairs, that Pelop, arriving there from Asia with abundance of wealth, soon gained so great an influence over those needy people, that, though a foreigner, he had the honour to have the Country called after his own name, and that the power thus gained by him was successively enlarged by his posterity. Eurystheus indeed, whose mother was the sister of Acestes, perished in Attica by means of the Heraclidae; and Eurystheus, when he departed on that expedition, left the govern- ment of Mycenae and his kingdom, because of his affinity, in the care of Acestes, who then resided with him, having fled from his father upon the murder of Chrysoclus. When therefore the return of Eurystheus was prevented by death, and the Mycenaeans...
Spatial connection by textual co-occurrence

- Frequency counts across entire text for each placename
- Distribution chart of occurrences through text
- Co-occurrence within page (or 500-word window)
- Network is of highest frequency co-mentions
Spatial connection without spatial proximity

- Sometimes nearby places mentioned together – expected?
- Does co-mention with remote locations suggest non-proximity connection – political, military, trading?
- Is systematic analysis feasible?
- “Noise” levels for high/low frequency places?
Low frequency – Thermopylae

Thermopylae

3 references
Report a problem with this record

External Resources
- Place page on Ptolomy
- Books referencing Thermopylae
- Pelagios Graph Explorer

Top Related Places
- Athenae (2)
- Hellas (2)
- Cosinthus (1)
- Sicilia (1)
- Strymon fl. (1)
- Naxactus (1)
- Tharsus (1)
- Diphyz M. (1)
High frequency – Athens

External Resources
- Place page on Piaiades
- Rocks referencing Athenae
- Religion Graph Explorer

Top Related Places
- Hellas (54)
- Peloponnesus (25)
- Thera (22)
- Attica (22)
- Lacedaemon/Laconia (21)
- Carinthia (13)
- Naupactus (12)
- Cervara (12)

Photos from Flickr: [View on Flickr]
Research issues

- Testing interfaces like GapVis in course-work
- Users see interface rather than content
- Difficult to do normative evaluations
- Can we obtain a “gold standard”?  
  - places that are independently known to be related in ancient texts like Thucydides?
Questions?