The GapVis project
and automated textual geoparsing

Google Ancient Places team,
Edinburgh Language Technology Group

GeoBib Conference, Giessen, 4th-5th May 2015
1 Google Ancient Places and GapVis
   - The GAP year
   - How GapVis Works
   - Evaluation

2 Language Technology Group
   - Recent georeferencing projects
   - Problems, issues, lessons learned
Google Ancient Places and GapVis
- The GAP year
- How GapVis Works
- Evaluation

Language Technology Group
- Recent georeferencing projects
- Problems, issues, lessons learned
The GAP year – Google Ancient Places

- Google Digital Humanities award, 2010-11
- https://googleancientplaces.wordpress.com/
- Cross-disciplinary and multi-national team
  - humanities, classics, archaeology, natural language processing, graphical interface
  - England, Scotland, California
- Produced the GapVis utility for reading Classics
The Genuine Works of Flavius Josephus, the Jewish historian

By Flavius Josephus

Published 1872 • View on Google Books

A war of that consequence, as to occasion the removal or destruction of six or seven nations of the posterity of Meneas, with their cities, which Josephus would not have said, if he had not had ancient records to justify his assertions, though those records be now lost. Gen. i. 21. t Gen. x. 21. l. That the Jews were called Hebrews from their progenitor Heber, our author Josephus here rightly affirms, and not from Abram the Hebrew, or passenger over Eu- phrates, as many of the moderns suppose. Shem is also called the Father of all the Children of Heber; or of all the city Nineve, and named his subjects Assyrians, who became the most fortunate nation beyond others. Arphaxad named the Arphaconites, who are now called Chaldeans. Aran had the Aramites, whom the Greeks call Syrians, as Lud founded the Ludites, who are now called Lydians. Of the four sons of Aram, Us founded Trachonitis and Damascus; this country lies between Palestine and Cæsarea. Ul founded Armenia; and Gather the Babaeans; and Moses the Menschen; it is now called Charax Spesiana. Salu was the son of Arphaxad: and his son was Heber, from whom they originally called the Jews Hebrews. Heber begat Jocam and Phaleg. He was called Phaleg because he was born] at the dispersion of the nations to their several countries, for Phaleg among the Hebrews signifies division. Now Jocam, one of the sons of Heber had these sons, Elmon, Dad, Seleph, Asenamoth, Iera, Adoram, Azel, Deila, Abal, Abimeal, Sabue, Opheir,
Pleiades – gazetteer of the Classical World

http://pleiades.stoa.org/
Alignment with Geonames (Pleiades+)

More precise lat/long locations. Support for multiple names (incl. modern) for same place, e.g. Halicarnassus → Bodrum.

At run time (Pleiades+)

Modern names often used in translations, e.g. Egypt – not in Pleiades. Find alternative names for Egypt in Geonames, e.g. Aegyptus; search for these in Pleiades.

Pleiades team's plans

Seeking funding from NEH to build robust API, with daily download option – cf Geonames.
Alignment with Geonames (Pleiades+)

More precise lat/long locations. Support for multiple names (incl. modern) for same place, eg *Halicarnassus* → *Bodrum*.

Modern names often used in translations, eg *Egypt* – not in Pleiades. Find alternative names for *Egypt* in Geonames, eg *Aegyptus*; search for these in Pleiades.

Pleiades team’s plans

Seeking funding from NEH to build robust API, with daily download option – cf Geonames.
Alignment with Geonames (Pleiades+)

More precise lat/long locations. Support for multiple names (incl. modern) for same place, eg Halicarnassus → Bodrum.

At run time (Pleiades++)

Modern names often used in translations, eg Egypt – not in Pleiades. Find alternative names for Egypt in Geonames, eg Aegyptus; search for these in Pleiades.
Alignment with Geonames (Pleiades+)  
More precise lat/long locations. Support for multiple names (incl. modern) for same place, eg Halicarnassus → Bodrum.

At run time (Pleiades++)  
Modern names often used in translations, eg Egypt – not in Pleiades. Find alternative names for Egypt in Geonames, eg Aegyptus; search for these in Pleiades.

Pleiades team’s plans  
Seeking funding from NEH to build robust API, with daily download option – cf Geonames.
Geoparsing engine

- Automatic detection of placenames in free text (English).
Geoparsing engine

- Automatic detection of placenames in free text (English).
- Online RESTful API: http://edina.ac.uk/unlock/texts/
- Stand-alone toolkit from LTG coming soon
  https://wp.ltg.ed.ac.uk/software/geoparser/
Step 1: Geotagging

- Named Entity Recognition – **location** mentions
- Also finds **person**, **date/time** expressions, **organisation**
- Rule-based, uses syntactic and semantic clues in context
Step 2: Georesolution

- Ground location mentions against one or more gazetteers
- Prefer “populated places”
- Prefer locations closer to others mentioned in document.
Step 2: Georesolution

- Ground location mentions against one or more gazetteers
- Prefer “populated places”
- Prefer locations closer to others mentioned in document.

Document is about USA. Prefer Paris, Texas to Paris, France.
Putting it all together: How GapVis Works

classical texts from Google Books
Putting it all together: How GapVis Works

classical texts from Google Books

Edinburgh Geoparser
Putting it all together: How GapVis Works

classical texts from Google Books

Pleiades gazetteer of ancient places

Edinburgh Geoparser
Putting it all together: How GapVis Works

classical texts from Google Books

Pleiades gazetteer of ancient places

database of toponym URIs –
tied back to text snippets
Putting it all together: 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: 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
Pleides gazetteer of ancient places
Edinburgh Geoparser
georeference
geotagging
georouting
geosimulation
georesolution

Edinburgh Geoparser
georeference
geotagging
georouting
geosimulation
georesolution
GapVis is still evolving – Forks and Re-use

- GAP2 – the Geographic Annotation Platform
- Hestia2:
  - [http://enridaga.github.io/gapvis/gap2](http://enridaga.github.io/gapvis/gap2)
  - Used in undergraduate teaching
  - Includes Greek original
- Related projects: Pelagios, Perseus, Perseids
  - People network as well as places
  - Hand-annotated by students in coursework
Perseids Journey of the Hero

Mermion (Μέμου), a son of Tithonus and Eos, and brother of Emathion. In the Odyssey and Hesiod he is described as the handsome son of Eos, who assisted Priam with his Ethiopians against the Greeks. He slew Antiochus, the son of Nestor, at Troy. (Hes. Th. 984, &c.; Hom. Od. 4.188, 11. 522, Apollod. 3.12 § 4.) Some writers called his mother a Cassian woman (Kroos), from the Persian province of Cissia. (Strab. p. 728; Hdt. 5.49, 52.) As Eos is sometimes identical with Hemera, Mermion's mother is also called Hemera. (Eos.) Homer makes only passing allusions to Mermion, and he is essentially a postmemic hero. According to these later traditions, he was a prince of the Ethiopians, and accordingly black. (Ov. Amor. 1.8.4, Epist. ex Pont. 3.3.96; Paus. 10.31.2.) He came to the assistance of his uncle Priam, for Tithonus and Priam were step-brothers, being both sons of Laomedon by different mothers. (Tzet. ad Lyce. 18.) Respecting his expedition to Troy there are different legends. According to some Mermion the Ethiopian first went to Egypt, thence to Susa, and thence to Troy. (Paus. 1.42.2.) At Susa, which had been founded by Tithonus, Mermion built the acropolis which was called after him the Mermionum. (Hdt. 5.53, 7.151; Strab. p. 728; Paus. 4.31.5.) According to some Tithonus was the governor of a Persian province, and the favourite of Teseutinus; and Mermion obtained the command of a large host of Ethiopians and Susans to succor Priam. (Diod. 2.22, 4.75; Paus. 10.31.2.) A third tradition states that Tithonus sent his son to Priam, because Priam had made him a present of a golden vine. (Serv. ad Aen. 1.493.) Dictys Cretensis (4.4) makes Mermion lead an army of Ethiopians and Indians from the heights of Mount Caucasus to Troy. In the fight against the Greeks he was slain by Achilles. The principal points connected with his exploits at Troy are, his victory over Antiochus, his contest with Achilles, and lastly, his death and the removal of his body by his mother. With regard to the first, we are told that Antiochus, the dearest friend of Achilles after the fall of Patroclus, hastened to the assistance of his father, Nestor, who was hard pressed by Paris. Mermion attacked Antiochus, and slew him. (Pind. P. 6.30, &c.) According to others, Mermion was fighting with Ajax; and before his Ethiopians could come to his assistance, Achilles came up, and killed Mermion. (Dict. Cret. 4.6) The same accounts represent Antiochus as having been conquered by Hector, (Ov. Ep. 1.15, Hygin. Sab. 113.) According to the common account, however, Achilles avenged the death of Antiochus upon Mermion, of whose fate Achilles had been informed by his mother, Thetis. While both were fighting Zeus weighed the fate of the two heroes, and the whole containing that of Mermion sank. (Pind. D. 2.148, Nym. 3.110, 6.83; Quint. Smyrn. 2.224, &c.; Philostr. Icon. 2.7; Plut. De Aud. Pict. 2.) According to Diodorus (2.22) Mermion was not killed in an open contest, but fell into an ambush in which the Thessalians lay in wait for him. Eos prayed to Zeus to grant her son immortality, and removed his body from the field of battle. She wept for him every morning; and the dew-drops which appear in the morning are the tears of Eos. (Serv. ad Aen. 1.493, Ov. Met. 13.622.) Philostratus (Her. 3.4) distinguishes between a Trojan and an Ethiopian Mermion, and
Evaluation

We like it, but is it good?

- Traditional NLP evaluation – against Gold Standard
- Qualitative evaluation – use in coursework
Using a hand-annotated Gold Standard

- Geotagging
  - Precision: were tokens we tagged actually places?
  - Recall: how many did we miss?
  - F-score: harmonic mean of P and R

- Georesolution: placenames correctly located?

- Varies considerably across domains

Using a hand-annotated Gold Standard

- Geotagging
  - Precision: were tokens we tagged actually places?
  - Recall: how many did we miss?
  - F-score: harmonic mean of P and R
- Georesolution: placenames correctly located?
- Varies considerably across domains

Geotagging: 59-82 F-score; Georesolution: 69-92%

Evaluation in Coursework

GapVis used in 2 undergraduate Classics courses.

- Visualisations of time and space help
- You can ask different questions
- Digital doesn’t necessarily mean easier

Adam Rabinowitz, University of Texas, Austin
1 Google Ancient Places and GapVis
- The GAP year
- How GapVis Works
- Evaluation

2 Language Technology Group
- Recent georeferencing projects
- Problems, issues, lessons learned
Recent LTG georeferencing projects

- DEEP: Digital Exposure of English Placenames
- Trading Consequences: 19th Century commodity trading
- Palimpsest: an Edinburgh literature landscape
Recent LTG georeferencing projects

- DEEP: Digital Exposure of English Placenames
- Trading Consequences: 19th Century commodity trading
- Palimpsest: an Edinburgh literature landscape

Natural Language Processing over historical and cultural documents, with Humanities partners
Palimpsest

- Edinburgh-related fiction from British Library, Hathi Trust, Project Gutenberg, NLS, etc – 550 books
- Created a fine-grained gazetteer of Edinburgh
- Visualise geo-located snippets of text
- [http://palimpsest.blogs.edina.ac.uk/](http://palimpsest.blogs.edina.ac.uk/)
Palimpsest online interface: *LitLong*

http://litlong.edina.ac.uk/
Problems, issues, lessons learned

- New domains, new problems – newswire, classics, fiction...
- OCR issues!!
- TEI does not enforce well-formed XML
- No standard document structure
- Only as good as the gazetteer
back, and received the direct force of each concussion as he trod along, he was very careful, in accordance with his master's advice, to

After two hours the guide stopped the elephant for rest, during which K quenched his thirst at a neighbouring spring by devouring the branches and shrubs round. Neither Sir Francis nor Mr. Fogg regretted the thorny descent with a feeling of relief. "Why, of iron!" exclaimed the general, gazing at the elephant. 

"Of Torgal iron," replied Passepartout, as I was preparing a hasty breakfast.

At noon the l'irisee gave the signal of depai country soon presented a very savage aspect, dates and dwarf-palms succeeded the dense fc vast, dry j,l,.ins, dotted with scanty shrubs, an( K'-^ai Ur>tk-i of \342\200\236yl\342\200\236ito. All this portion of I ^i'^'-ii \302\253^ litli.: frequented by travellers, is ink
Dublin is the capital of Ireland.

Dublin is in the province of Leinster on Ireland's east coast, at the mouth of the River Liffey.

Founded as a Viking settlement, the Kingdom of Dublin became Ireland's principal city following the Norman invasion.

The city expanded rapidly from the 17th century and was briefly the second largest city in the British Empire before the Act of Union in 1800.
**Dublin** is the capital of **Ireland**.

**Dublin** is in the province of Leinster on **Ireland**'s east coast, at the mouth of the **River Liffey**.

Founded as a Viking settlement, the Kingdom of **Dublin** became **Ireland**'s principal city following the Norman invasion.

The city expanded rapidly from the 17th century and was briefly the second largest city in the **British Empire** before the Act of Union in 1800.

Following the partition of **Ireland** in 1922, **Dublin** became the capital of the **Irish Free State** and later the **Republic of Ireland**.
Dublin (p-A117)

Type = pplc (capital of a country)
Lat = 53.33°, Long = -6.25°

Press a button to select this as the correct location for Dublin.

Select for this mention  Select for all mentions

Baile Átha Cliath (p-A107)

Type = civil (administrative division)
Lat = 53.38°, Long = -6.27°

Press a button to select this as the correct location for Dublin.

Select for this mention  Select for all mentions
Dublin is the capital of Ireland.

Dublin is in the province of Leinster on Ireland's east coast, at the mouth of the River Liffey.

Founded as a Viking settlement, the Kingdom of Dublin became Ireland's principal city following the Norman invasion.

The city expanded rapidly from the 17th century and was briefly the second largest city in the British Empire before the Act of Union in 1800.

Following the partition of Ireland in 1922, Dublin became the capital of the Irish Free State and later the Republic of Ireland.
Conclusions

- Visualisations of space and time give new perspectives
Conclusions

- Visualisations of space and time give new perspectives
- Tailoring for each new domain is required
Conclusions

- Visualisations of space and time give new perspectives
- Tailoring for each new domain is required
- Automated NLP is fast, but not perfect
Conclusions

- Visualisations of space and time give new perspectives
- Tailoring for each new domain is required
- Automated NLP is fast, but not perfect
- Quality of copyright-free material often poor
Conclusions

• Visualisations of space and time give new perspectives
• Tailoring for each new domain is required
• Automated NLP is fast, but not perfect
• Quality of copyright-free material often poor
• Often need specialist gazetteers
Conclusions

- Visualisations of space and time give new perspectives
- Tailoring for each new domain is required
- Automated NLP is fast, but not perfect
- Quality of copyright-free material often poor
- Often need specialist gazetteers
- Collaboration between different disciplines essential
Thank You