Evaluation of Georeferencing

Richard Tobin, Claire Grover, Kate Byrne, James Reid, Jo Walsh

School of Informatics and EDINA
University of Edinburgh

18th February 2010
The Edinburgh Georeferencing System

overview
geotagging
georesolution

Evaluation Setup

gazetteers used
evaluation data
“gold standard” annotation

Evaluation Results

g geotagging scores
g georesolution scores
g end-to-end, for whole pipeline
The Georeferencing System

- Specialised NLP pipeline tuned for geospatial referencing
- Developed using general purpose Information Extraction tools
  - LT-XML2 – for parsing and navigating XML documents
  - LT-TTT2 – rule-based NLP components
  - see [http://www.ltg.ed.ac.uk/software](http://www.ltg.ed.ac.uk/software)
- Available online through EDINA’s Unlock service
  - [http://unlock.edina.ac.uk/](http://unlock.edina.ac.uk/)
- Modular construction: geotagger + georesolver
The Georeferencing System

- Specialised NLP pipeline tuned for geospatial referencing
- Developed using general purpose Information Extraction tools
  - LT-XML2 – for parsing and navigating XML documents
  - LT-TTT2 – rule-based NLP components
  - see http://www.ltg.ed.ac.uk/software
- Available online through EDINA’s Unlock service
  - http://unlock.edina.ac.uk/
- Modular construction: geotagger + georesolver
The Georeferencing System

- Specialised NLP pipeline tuned for geospatial referencing
- Developed using general purpose Information Extraction tools
  - LT-XML2 – for parsing and navigating XML documents
  - LT-TTT2 – rule-based NLP components
  - see http://www.ltg.ed.ac.uk/software
- Available online through EDINA’s Unlock service
  - http://unlock.edina.ac.uk/
- Modular construction: geotagger + georesolver
The Pipeline
Geotagging step: identifying place names in text

Rule-based NER
Person-Place ambiguity – Frances Chichester
Organisation-Place ambiguity – University of Edinburgh

Contextual clues passed to georesolution step
Containment – Ipswitch, Suffolk
Co-ordination – Chelmsford, Colchester and Ipswitch
Proximity – Manningtree near Ipswitch
Alternative name forms – Down for Co. Down
Georeferencing system

Geotagging


Select a file to upload: Browse...

Choose a gazetteer: geonames

Type is text/xml (xml), gazetteer is geonames

Please wait geotagging... doing gazetteer lookup...

545

107

In 2107, the maximum temperature at Balmoral on November 14th, and the highest maximum 292°F at Felixstowe, Halstead, Kew observatory, Tottenham, St. James's Park, and Killerton on August 19th.

Precipitation. - Taking the British Isles as a whole, the year, although by no means exceptional, was probably the driest since 1908, a result largely due to the persistent shortage of rain during the summer and autumn.

The general rainfall for the greater divisions of the country during the year was 25.3 per cent. of the normal in England, 98 in Wales, 92 in Scotland, and 90 in Ireland, and the percentage for the British Isles as a whole 96.

In December the percentage was as high as 147, and 143 in March, 124 in January and 113 in April, but only 54 in October and 55 in July.

During January the aggregate at many stations was more than double the normal, 230 per cent., for instance, at Portsmouth, 242 at Portland Bill, and 25 at Bournemouth.

In February the totals were very varied, and at Fort William the total (16 mm.) was only 8 per cent. of the normal, and at Glencarron, normally very wet, the total (28 mm.) was only 5.8.
Georesolution step: locating the places on a map

- Obtain candidate locations from gazetteer
- Pick correct one
  - collect available attributes (lat/long, feature type, etc)
  - augment with population and country if not given
  - add external knowledge – bounding box
  - apply heuristics to rank candidates
Georesolution heuristics

- Feature type – eg prefer populated place to facility
- Population – prefer bigger places
- Use contextual clues from geotagger – containment, proximity
- Prefer places within bounding box, if supplied
- Favour clustered places – lowest average distance to nearest candidates

Clustering example

Doc mentions London, Boston and Wembley Stadium – probably all UK;
Click on a lat/long to centre the map there.

<table>
<thead>
<tr>
<th>Location</th>
<th>Lat/Long</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canterbury</td>
<td>51°17'N 1°5'E</td>
</tr>
<tr>
<td>London</td>
<td>51°31'N 0°8'W</td>
</tr>
<tr>
<td>Dover</td>
<td>51°8'N 1°18'E</td>
</tr>
<tr>
<td>Ebbsfleet</td>
<td>51°19'N 1°21'E</td>
</tr>
<tr>
<td>Gravesend</td>
<td>51°26'N 0°22'E</td>
</tr>
<tr>
<td>St. Pancras</td>
<td>51°26'N 0°22'E</td>
</tr>
</tbody>
</table>
The Edinburgh Georeferencing System

overview
geotagging
georesolution

Evaluation Setup

gazetteers used
evaluation data
“gold standard” annotation

Evaluation Results

geotagging scores
georesolution scores
end-to-end, for whole pipeline
Gazetteers used

- GeoNames: http://www.geonames.org/
- GeoCrossWalk (no longer available online)
- Unlock: http://unlock.edina.ac.uk/
- GeoCrossWalk and Unlock are GB only
- – augment with world-wide places having population >200,000

Gazetteer issues
Point data inappropriate for large regions
Often contain near duplicates (Bristol: populated place, admin district)
List of candidates for given name can be very long
Granularity variable
Gazetteers used

- GeoNames: http://www.geonames.org/
- GeoCrossWalk (no longer available online)
- Unlock: http://unlock.edina.ac.uk/
- GeoCrossWalk and Unlock are GB only
- – augment with world-wide places having population $>$ 200,000

Gazetteer issues

Point data inappropriate for large regions
Often contain near duplicates (Bristol: populated place, admin district)
List of candidates for given name can be very long
Granularity variable
Data: from JISC-funded projects with EDINA

- Historical texts, scanned and OCRed
- Histpop
  - historical population reports for Britain and Ireland, 1801–1937
  - 9,329 sentences; 5,890 placename mentions
- BOPCRIS (British Official Publications...)
  - journals of the House of Lords, 1688–1854
  - 5,486 sentences; 1,181 placename mentions
- The Stormont Papers
  - parliamentary debates in Northern Irish Parliament, 1921–1972
  - 7,601 sentences; 1,216 placename mention
- Also: modern data from ACE 2005 SpatialML corpus
DIE Martis, 5° die Martis.

Dominus tam Spirituales quam Temporales presentes forenses:

Dux Cumberlaid.
L. President.
Dux Somerset.
Dux Grafton.
Dux Bolton.
Marq. Halifax.
L. Steward.
Comes Shrewsbury.
Comes Pembroke.
Comes Suffolk.
Comes Bridgewater.
Comes Devon.
Comes March.
Comes Middlesex.
Comes Rivers.
Comes Stamford.
Comes Kingston.
Comes Carnarvon.
Comes Bath.
Comes Coventry.
Comes Buckingham.
Comes Suffolk.
Comes Ferrars.
Comes Manchelesfield.
Comes Radnor.
Comes Nottingham.
Comes Rochester.
Comes Portland.
Comes Fowenberge.
Comes Malmouth.
Comes Marborough.
Comes Terrington.
Vicount Newport.
Vicount Weymouth.
Vicount Lambeth.

PRAYERS.

This Day Thomas Earl of Suffolk, James Lord Chamberlain, and Robert Lord Levington, took the Oaths, and made and subscribed the Declaration, pursuant to the Statutes.

The Earl of Bridgewater reported from the Committee of Privileges, "That the Committee is of Opinion, That the Earl of Malborough's Privilege hath been broken, by Lieutenant-Crane's prevailing and determining John Williston his Lordship's Waterman; and that the said John Williston ought to be released, and

It is ORDERED, by the Lords Spiritual and Temporal in Parliament assembled, That the said Richard Hedges be, and is hereby, required to put in his Answer thereunto, in Writing, on Tuesday the 8th Day of April next, at Ten of the Clock in the Forenoon: and the Petitioners are to cause Notice to be given to the said Richard Hedges, to the End be answered accordingly.

Upon the Petition of Edward Coke Esquire, Censor an infant, by the Lady Anne Coga Widow, his Mother, and Guardian, praying the Revival of their Appeal, whereupon Andrew Fountain Esquire is Defendant, and a Day may be appointed for hearing thereof:

It is ORDERED, by the Lords Spiritual and Temporal in Parliament assembled, That this House will hear the said Cause, by Counsel on both Sides, at the Bar, on Wednesday the Third Day of April next, at Ten of the Clock in the Forenoon: And it is further ORDERED, That the mentioned Papers and Vellum Book, or any other Writings, which were formerly produced upon any Hearings in this House by the said Andrew Fountain, shall be then produced; wherefore the Petitioners are to cause Notice to be given to the Defendants, to the End they attend with their Counsel accordingly.

Upon the Petition of Vincent Vincent Esquire: Saying, "That, since his Appeal lodged in this House, by Order of the Court of Chancery, he is taken into Custody, and committed to the Prison, for not performing the Decrees of that Court; from which he appealed to this House; where he now remains, to his great Danger, Damage, and Expense; and humbly praying Relief therein:"

It is thereupon ORDERED, by the Lords Spiritual and Temporal in Parliament assembled, That the said Petition of Vincent Vincent Esquire be, and is hereby, referred to the Lords Committees appointed to consider and inspect the Journals of this House, whether any Proceedings may be had in the Courts in Westminster Hall, in the Intervals of Parliament, after Appeals or Writs of Error are depending in this House.

Upon the Petition of Sir Robert Brook, Sir R. Brooke, and Sir R. Brooke, Baronet and Dame Martha his Wife, Relief and Advance of £10,000 to Christopher Tenison, deceased, praying the Revival of their said Appeal, and that Sir Charles Porter may be made a Party thereto:

It is thereupon ORDERED, by the Lords Spiritual and Temporal in Parliament assembled, That the said Sir Charles Porter may be made a Party thereto.
Annotating “gold standard” data

- Two stage annotation
  1. Place and Person names for NER step (geotagging)
  2. Georeferencing annotation
- Georeferenced separately for GeoCrossWalk and GeoNames
- Specially written annotation tool – user picks point from map
- All candidate place names retained, not just “correct” one
- Compare system and human choice from same candidate list
Sir EDWARD: ARCHDALE: I do not agree that the figure given by the hon. Member is the amount required.

The small farmers have nearly all got enough seed potatoes.

The larger farmers between Lisnaskea and the Lough shore are the ones who have suffered damage by flooding, and they are well able to get sureties.

Mr. HEALY: Is the right lion Gentleman prepared to act on the recommendation of his own inspector against the recommendations of two boards of guardians who know the locality and who recommend a certain course of action?

In other words, are the wishes of public representatives who are going to bear the responsibility to be turned down because you prefer to act on the recommendation of a single inspector who does not know the locality?

Mr. BEATTIE: I would like to ask the right hon. Gentleman whether he is prepared to give those seed potatoes to disabled ex-Servicemen.
The Edinburgh Georeferencing System

overview
geotagging
georesolution

Evaluation Setup

gazetteers used
evaluation data
“gold standard” annotation

Evaluation Results

geotagging scores
georesolution scores
end-to-end, for whole pipeline
Evaluation of geotagging step

- Scores on historical data:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histpop</td>
<td>82.09%</td>
<td>80.78%</td>
<td>81.43</td>
</tr>
<tr>
<td>BOPCRIS</td>
<td>55.92%</td>
<td>61.56%</td>
<td>58.61</td>
</tr>
<tr>
<td>Stormont</td>
<td>71.72%</td>
<td>74.67%</td>
<td>73.17</td>
</tr>
</tbody>
</table>

- Early text (BOPCRIS, 1688-1817) suffers from OCR errors

- Scores on newswire and usenet data:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpatialML</td>
<td>63.39%</td>
<td>75.26%</td>
<td>68.82</td>
</tr>
</tbody>
</table>
Evaluation of georesolution

- Resolver works over list of candidate places from gazetteer
- Georesolution heuristics scaled to 0-1 range
- Weighted combination to produce score for each candidate
- System is presented with exactly same candidates as annotator
  - Two evaluation metrics, over “gold” NEs:

  **Exact evaluation**
  System makes same choice as annotator; tested with and without known bounding box.

  **Proximity evaluation**
  System choice within 5km of gold position is considered correct.
Evaluation of georesolution

- Resolver works over list of candidate places from gazetteer
- Georesolution heuristics scaled to 0-1 range
- Weighted combination to produce score for each candidate
- System is presented with exactly same candidates as annotator
- Two evaluation metrics, over “gold” NEs:

<table>
<thead>
<tr>
<th>Evaluation Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact evaluation</td>
<td>System makes same choice as annotator; tested with and without known bounding box.</td>
</tr>
<tr>
<td>Proximity evaluation</td>
<td>System choice within 5km of gold position is considered correct.</td>
</tr>
</tbody>
</table>
Evaluation of georesolution

- Resolver works over list of candidate places from gazetteer
- Georesolution heuristics scaled to 0-1 range
- Weighted combination to produce score for each candidate
- System is presented with exactly same candidates as annotator
- Two evaluation metrics, over “gold” NEs:

**Exact evaluation**
System makes same choice as annotator; tested with and without known bounding box.

**Proximity evaluation**
System choice within 5km of gold position is considered correct.
## Evaluation Results

### Exact evaluation

<table>
<thead>
<tr>
<th></th>
<th>Histpop</th>
<th>GeoNames</th>
<th>GeoCrossWalk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>baseline</td>
<td>21.9%</td>
<td>44.7%</td>
</tr>
<tr>
<td></td>
<td>correct without locality</td>
<td>69.8%</td>
<td>63.9%</td>
</tr>
<tr>
<td></td>
<td>correct with locality</td>
<td>75.3%</td>
<td>63.9%</td>
</tr>
<tr>
<td></td>
<td>BOPCRIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>baseline</td>
<td>21.2%</td>
<td>35.8%</td>
</tr>
<tr>
<td></td>
<td>correct without locality</td>
<td>67.2%</td>
<td>79.2%</td>
</tr>
<tr>
<td></td>
<td>correct with locality</td>
<td>76.9%</td>
<td>79.2%</td>
</tr>
<tr>
<td></td>
<td>Stormont</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>baseline</td>
<td>48.7%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>correct without locality</td>
<td>84.9%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>correct with locality</td>
<td>90.2%</td>
<td></td>
</tr>
</tbody>
</table>
Proximity evaluation

- On historical text, using bounding boxes:

<table>
<thead>
<tr>
<th>Gazetteer</th>
<th>No. of place names (all found in gaz)</th>
<th>Correct within 5km geonames</th>
<th>Correct within 5km xwalk</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Histpop</strong></td>
<td></td>
<td>5091</td>
<td>4435</td>
</tr>
<tr>
<td></td>
<td></td>
<td>82.0%</td>
<td>92.7%</td>
</tr>
<tr>
<td><strong>BOPCRIS</strong></td>
<td></td>
<td>735</td>
<td>650</td>
</tr>
<tr>
<td></td>
<td></td>
<td>81.4%</td>
<td>91.2%</td>
</tr>
<tr>
<td><strong>Stormont</strong></td>
<td></td>
<td>985</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>92.1%</td>
<td></td>
</tr>
</tbody>
</table>

- Gazetteers contain many near duplicates
Proximity evaluation

- On SpatialML data:

<table>
<thead>
<tr>
<th></th>
<th>GeoNames</th>
<th>Unlock</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SpatialML</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no. of place names</td>
<td>3628</td>
<td>3628</td>
</tr>
<tr>
<td>no. for which gaz entries found</td>
<td>3538</td>
<td>3049</td>
</tr>
<tr>
<td>correct within 5km</td>
<td>2946</td>
<td>2143</td>
</tr>
<tr>
<td>as % of all 3628 place names</td>
<td>81.2%</td>
<td>59.0%</td>
</tr>
</tbody>
</table>

- Not all names in gazetteer at all
- SpatialML text covers world-wide area
- Bounding boxes not applicable
- Unlock gazetteer is GB (augmented with large places)
End-to-end comparison with Yahoo! Placemaker

<table>
<thead>
<tr>
<th></th>
<th>System</th>
<th>Placemaker</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SpatialML</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no. of place names</td>
<td>3628</td>
<td>3628</td>
</tr>
<tr>
<td>no. for which gaz entries found</td>
<td>2923</td>
<td>2635</td>
</tr>
<tr>
<td>correct within 5km</td>
<td>2504</td>
<td>882</td>
</tr>
<tr>
<td>as % of total</td>
<td>69.0%</td>
<td>24.3%</td>
</tr>
<tr>
<td>correct within 25km</td>
<td>2520</td>
<td>1067</td>
</tr>
<tr>
<td>as % of total</td>
<td>69.5%</td>
<td>29.5%</td>
</tr>
<tr>
<td>correct within 50km</td>
<td>2558</td>
<td>1677</td>
</tr>
<tr>
<td>as % of total</td>
<td>70.5%</td>
<td>46.2%</td>
</tr>
<tr>
<td>correct within 100km</td>
<td>2664</td>
<td>2133</td>
</tr>
<tr>
<td>as % of total</td>
<td>73.4%</td>
<td>58.8%</td>
</tr>
</tbody>
</table>
Summary

- Many variables – evaluation not straightforward
- Comparing systems hard when different gazetteers used
- Eg WoeIDs vs GeoNames/IGDB positions
- Our evaluation compares human and machine choice from list –
- – not just accuracy of lat/long point selected