A Utility Centered Approach for Evaluating and Optimizing Geo-Tagging

Albert Weichselbraun

Vienna University of Economics and Business
Department of Information Systems and Operations
Augasse 2-6, 1090 Vienna
albert.weichselbraun@wu.ac.at

September 7, 2009
Agenda

Definition

Motivation
   Assign unique locations to resources
   Standardize evaluation sets

Method
   Idea
   User preferences
   Evaluation ontology

Demonstration

Outlook & Conclusions
Definition

Geo-tagging of phrases (deterministic)
“identify geographic references in resources and ground them to geographic entities.”

Geo-tagging of resources (non-deterministic)
“assign one geographic entity to a resource”

Difficulties
- multiple location references
  - focus algorithms
- correct focus and the impact of incorrect tags often depend
  - on the user and
  - the use case
Motivation: assign unique locations to resources

Figure: Tivoli Hotels in Madeira.
Motivation: standardize evaluation sets

- Clough and Sanderson [1] – importance of comparative evaluations → stimuli for research
  - scope
  - coverage
  - correctness
  - granularity
  - balance and richness
- Turpin and Hersh [3] – IR metrics do not necessarily correspond to user performance and satisfaction
different people (use case, user) $\leftrightarrow$ different priorities

classic economic problem

utility functions - map user preferences ($p_u$), answers ($a_i$) and solutions ($s_i$) to a utility score

$$u = f(p_u, a_i, s_i)$$

ontologies provide context information to support the mapping (e.g., Salzburg is a city in Austria, Madeira is a state of Portugal, ...)
User preferences

- basic weights \( f_{eval}(a_i) = \prod_{j=1}^{n} w_{d_j} \)
- more detailed specifications are possible but not necessary

![Diagram showing the relationship between regions in Austria]

Figure: Example: Utility Scoring.
Evaluation ontology & algorithms

- based on GeoNames; handles GeoNames instance data
- Evaluation metrics:
  - uses the evaluation ontology + instance data
  - translates movements alongside ontological dimensions to weights
  - uses heuristics to handle sparse data
Handling of sparse data

- isNeighor: restricted to instance data on the same scope (e.g. country – country)
- example heuristics for “close matches”

\[
\begin{align*}
  u_c &= u_c^h + u_c^o & (2) \\
  u_c^o &= (1 - u_c^h) \cdot f_{eval} & (3) \\
  f_{eval} &= \max(0, (1 - \frac{d}{d_e}) \prod_{i=1}^{n} w_{di}) & (4) \\
  d_e &= E(d_{random}) = \frac{1}{3} \sqrt{A_{S_1}/\pi} & (5)
\end{align*}
\]
Geo-tagger evaluation

<table>
<thead>
<tr>
<th>Comparison</th>
<th>$\cong$</th>
<th>$A \sqsupseteq B$</th>
<th>$A \sqsubseteq B$</th>
<th>$A \sqsupset B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenCalais vs. Reuters</td>
<td>20 %</td>
<td>72 %</td>
<td>31 %</td>
<td>78 %</td>
</tr>
<tr>
<td>geoLyzard vs. Reuters</td>
<td>17 %</td>
<td>62 %</td>
<td>25 %</td>
<td>75 %</td>
</tr>
<tr>
<td>OpenCalias vs. geoLyzard</td>
<td>47 %</td>
<td>51 %</td>
<td>48 %</td>
<td>62 %</td>
</tr>
</tbody>
</table>

Table: Evaluation of geo-tags created by OpenCalais and geoLyzard.

- improve the comparability of geo-tagger results
Outlook & Conclusions

Conclusions

- more fine grained notion of correctness
- user preference, evaluation ontologies and heuristics
- application of this approach to geo-taggers
- use to improve the comparability of geo-taggers

Outlook

- create a standardized geo-tagger evaluation set
- implement a test driven development methodology for use case specific geo-taggers
Paul Clough and Mark Sanderson.  
A proposal for comparative evaluation of automatic annotation for geo-referenced documents.  

Jochen L. Leidner.  
An evaluation dataset for the toponym resolution task.  

Andrew H. Turpin and Falk Scholer.  
User performance versus precision measures for simple search tasks.  
In SIGIR ’06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pages 11–18, New York, NY, USA, 2006. ACM.