

Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names

Jochen L. Leidner

Abstract

The problem of automatic toponym resolution, or computing the mapping from occurrences of names for places as found in a text to an unambiguous spatial footprint of the location referred to, such as a geographic latitude/longitude centroid is difficult to automate due to insufficient and error-prone geographic databases, and a large degree of place name ambiguity: common words need to be distinguished from proper names (geo/non-geo ambiguity), and the mapping between names and locations is ambiguous (London can refer to the capital of the UK or to London, Ontario, Canada, or to about forty other Londons on earth).

This thesis investigates how referentially ambiguous spatial named entities can be grounded, or resolved, with respect to an extensional coordinate model robustly on open-domain news text by collecting a repertoire of linguistic heuristics and extra-linguistic knowledge sources such as population. I then investigate how to combine these sources of evidence to obtain a superior method. Noise effects introduced by the named entity tagging that toponym resolution relies on are also studied. While few attempts have been made to solve toponym resolution, these were either not evaluated, or evaluation was done by manual inspection of system output instead of creating a re-usable reference corpus. A systematic comparison leads to an inventory of heuristics and other sources of evidence. In order to carry out a comparative evaluation procedure, an evaluation resource is required, so a reference gazetteer and an associated novel reference corpus with human-labelled referent annotation were created for this thesis, to be used to benchmark a selection of the reconstructed algorithms and a novel re-combination of the heuristics catalogued in the inventory. Performance of the same resolution algorithms is compared under different conditions, namely applying it to the output of human named entity annotation and automatic annotation using an existing Maximum Entropy sequence tagging model.

Toponym Resolution
in Text

Annotation, Evaluation and Applications
of Spatial Grounding of Place Names



Jochen L. Leidner

ORDER FORM

Toponym Resolution in Text: Annotation, Evaluation and Applications of
Spatial Grounding of Place Names
by Jochen L. Leidner

ISBN: 1-58112-384-1; paperback; 289 pages, US \$25.95

<http://www.dissertation.com> – 2008 Preview ebook on the internet.

Credit cards accepted on-line at internet address above.

Qty	Description	Price	Total
	Toponym Resolution in Text	\$25.95	
Florida residents please add 6% sales tax.		SalesTax	
SHIPPING CHARGES			
Add \$4.90 shipping and handling for the first copy and \$1.95 for each additional copy in the U.S./Can/Mex. Other countries: \$9.90 for the first copy and \$6.95 for each additional copy.		Shipping	
		Total	

SHIP TO

Name	
Apt, Suite	
Street	
City, State	
Zip, Country	
Phone/Email	

Send check in US dollars drawn on US bank to:

DISSERTATION.COM
23331 Water Circle
Boca Raton, FL 33486-8540 • USA
fax (561) 750-6797