

Chapter 30

Using Event Semantics for Toponym Disambiguation

Kirk Roberts

University of Texas at Dallas, USA

Cosmin Adrian Bejan

University of Southern California, USA

Sanda Harabagiu

University of Texas at Dallas, USA

ABSTRACT

This chapter discusses a method for improving the disambiguation of location names using limited event semantics. Location names are often ambiguous, as the same name may refer to locations in different states, countries, or continents. Ambiguous location names, also known as toponyms, need to be disambiguated (or grounded) when resolving many spatial relations expressed in textual documents. Previous methods for disambiguating toponyms have utilized simple heuristics, statistical ranking, and ontological methods in order to resolve a location reference. However, since toponyms are used in documents that refer to events, semantic knowledge characterizing events can be used to ground location names. We propose an ontology-based method with a technique that considers the participants in events such as people, organizations, and other locations. Event semantics are integrated into an ontology that is used to distinguish geographical names through a probabilistic approach based on logistic regression. Our experimental results on the SpatialML corpus (Mani et al., 2008) indicate that using event structures improves the quality of disambiguated toponyms.

INTRODUCTION

Toponym disambiguation is the task of grounding ambiguous spatial locations in text (toponyms) by normalizing them to some structured representation (e.g., geo-coordinates, database entry,

or location within a geographic ontology). This task proves to be quite difficult for some highly ambiguous locations. For example, there are over one thousand places named Springfield. To be able to assess the level of location name ambiguity, we have performed a two-tiered study. In the study, we used two gazetteers to count the number of locations with ambiguous names:

DOI: 10.4018/978-1-60960-741-8.ch030

- Geographic Names Information System (GNIS)¹ provided by the U.S. Geological Survey (USGS), which contains over two million locations and facilities within the United States.
- GEOnet Names Server (GNS)² provided by the U.S. National Geospatial-Intelligence Agency (NGA), which contains over six million locations world-wide, excluding the United States.

We first determined that of more than two million unique location names in the two gazetteers, less than 20% are ambiguous. Then, to test the ambiguity of spatial locations in common use, we analyzed the SpatialML corpus (Mani et al., 2008), which contains 428 documents and 715 unique, manually-annotated location names. For these names, more than 80% are ambiguous. For example, our database contains 44 matches for “*America*,” including *América/Mexico*, *America/Guinea-Bissau*, and the *United States of America*;

24 matches for “*Palestine*,” including *Palestine/Texas*, and the *Occupied Palestinian Territory*; and 7 matches for “*Baghdad*,” including both the Iraqi city of Baghdad and the Iraqi Governate of Baghdad that contains the city. For more details on this case study, see Table 1.

Since location name ambiguity is such a pervasive problem, we have developed a toponym disambiguation method that uses the hypothesis that locations are groundings of events. Therefore event semantics impacts the quality of toponym disambiguation. In our approach, toponyms are disambiguated by (1) taking into account all possible spatial relations induced from event structures in which they participate, and (2) scoring all paths in the ontology between toponyms and event participants. Scoring is performed by logistic regression.

Algorithm 1 illustrates a generalized method for toponym disambiguation used by many of the systems discussed later. First, the location mentions in the document are extracted (EXTRACT-

Table 1. Case study on ambiguous names. (a) Globally ambiguous names collected using USGS and NGA gazetteers. (b) Ambiguous names in corpus collected on 715 unique names in 428 documents from SpatialML (Mani et al., 2008).

Duplicates	Entries	Percent
(a)		
1	2,150,855	80.2%
2+	531,550	19.8%
5+	86,493	3.2%
10+	30,759	1.1%
50+	2,294	0.086%
100+	617	0.023%
1000+	5	0.0002%
(b)		
1	119	16.6%
2+	596	83.4%
5+	438	61.3%
10+	310	43.4%
50+	83	11.6%
100+	16	2.2%

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/using-event-semantics-toponym-disambiguation/61068

Related Content

Machine Translation within Commercial Companies

Tomáš Hudík (2015). *Modern Computational Models of Semantic Discovery in Natural Language* (pp. 256-272).

www.irma-international.org/chapter/machine-translation-within-commercial-companies/133882

Shape Grammar Approaches in Architecture

(2020). *Grammatical and Syntactical Approaches in Architecture: Emerging Research and Opportunities* (pp. 19-55).

www.irma-international.org/chapter/shape-grammar-approaches-in-architecture/245859

Integrating Chinese Community into Canadian Society: Podcasts, Technology Apprehension, and Language Learning

Yuping Mao, Martin Guardado and Kevin R. Meyer (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 1550-1568).

www.irma-international.org/chapter/integrating-chinese-community-into-canadian-society/108793

Robustness Against DA/AD Conversion: Concepts, Challenges, and Examples

Martin Steinebach (2008). *Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks* (pp. 248-259).

www.irma-international.org/chapter/robustness-against-conversion/8335

Laying the Ground for Online English as a Second or Foreign Language (ESL/EFL) Composition Courses and University Internationalization: The Case of a U.S.-China Partnership

Estela Ene (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 521-539).

www.irma-international.org/chapter/laying-the-ground-for-online-english-as-a-second-or-foreign-language-eslefl-composition-courses-and-university-internationalization/108736