GIR'16 Workshop Report
10th ACM SIGSPATIAL Workshop on
Geographic Information Retrieval
San Francisco, USA 4th November 2016

Christopher B. Jones[1]

[1] Cardiff University, United Kingdom

c.b.jones@cs.cf.ac.uk

Ross S. Purves[2]

[2]University of Zurich, Switzerland

ross.purves@geo.uzh.ch

(Workshop Co-chairs)

The field of Geographic Information Retrieval (GIR) is concerned with the problems of gaining access to documents and to information within documents that relate to geographical locations. The methods employed derive from those of Information Retrieval, with its emphasis upon unstructured documents and natural language processing, and of Geographical Information Systems which, while oriented to structured data, are the source of many spatial analytical and data access methods that are relevant to GIR. This workshop is the tenth of a series of workshops that have addressed research challenges for GIR, including those relating to recognising and disambiguating references to place names in text (geoparsing); determining the geographic scope of documents; developing gazetteers and ontologies to maintain knowledge of toponyms and geographic concepts; spatio-textual indexing methods that combine inverted file methods with those of spatial database indexing; managing vagueness and uncertainty in geographic terminology; extracting geo-spatial facts and events from documents; and evaluating the performance of geo-information retrieval systems.

The 10th GIR workshop was held on 31st October 2016 at the ACM SIGSPATIAL conference in San Francisco, USA. Previous workshops have been held either in combination with the SIGIR and CIKM conferences or as stand alone events in cooperation with ACM SIGSPATIAL.

At GIR'16 there were 9 presentations, of which 6 were full papers and 3 were short papers. The workshop was organized around three sessions relating to common themes of the submitted papers and a fourth session that was a discussion of ideas for shared case studies in GIR.

The first session focused on extracting information on events, the function of geographic places and the presence of people at places that they refer to in Tweets. Spitz et al presented a method for imprecise extraction of events that relate to an actor, a time and a location. They maintain a knowledge base of graphs of entities (potential actors), time and location that record relations (such as part-of and similar to) between the entities to allow searching for coarser or finer granularity instances of the elements of a query. Tardy et al presented a paper that explored the use of social media, in particular Flickr, to detect both geographic features and their function or use (as for a building). They looked at an urban context and used the Geonames gazetteer to detect geographic features along with a word sense disambiguation tool (BabelFly) to detect the sense of words. Sparks et al described a machine learning approach to detecting whether Twitter users are present at specific types of geographic facility (restaurant, airport, stadium). The intention was to exploit very widely used social media (Twitter) to supplement data on people's behaviour patterns available from less widely used social media that record location check-ins explicitly.

The second session was concerned with geo-data disambiguation and integration. Blank and Henrich used graph search and approximate string matching methods (with the Geonames gazetteer) to extract itineraries

from historic route descriptions with a depth-first branch-and-bound algorithm. Yu et al addressed the problem of place matching using multiple similarity metrics in combination with an adaption of a method referred to as Naïve Descending Extraction. Their methods improved on a baseline semantic alignment system. Golubovic et al presented a design for a system intended to alert farmers to imminent threats, related for example to weather or pests, based on automatically identifying, analysing and integrating data from multiple sources including news articles and social media.

The topic of the third session was geoparsing and evaluation. Brando et el presented the results of a comparison between several named entity recognition systems, some of which (in particular Stanford NER) use a machine learning approach and can be trained with data specific to a particular domain. Their experiments were notable for distinguishing between the standard task of recognising gazetteer names and the more challenging task of recognising various informal references to places that can include generic place types as part or all of the name. Morteza highlighted the limitations of some existing approaches to measuring precision and recall in geoparsing and proposed an approach to toponym resolution that combines a measure of confidence in the resulting geo-coded place with geographic distance between the resolved location and the gold standard location. Cai and Ye Tian studied processes employed by human annotators when resolving toponyms and noted the difference between heuristics they employed, such as hierarchical ontological relations between place references in a single article, and an assumption that places in the same article are close together. They presented a geo-referencing workbench that progressively learns a gazetteer based on manual intervention.

The final session was a lively and constructive discussion of ideas for possible shared tasks and case studies in GIR. These related to quite a variety of topics including the proposal of studies of the forms of place names and the improvement of gazetteers, the creation of shared resources for performing geo-parsing, the generation of benchmark systems and of an auto-evaluation framework for new GIR methods (analogous to some existing such frameworks in other areas of information retrieval and NLP).