

# Location Extraction from Social Media: Geoparsing, Location Disambiguation and Geotagging

Stuart E. Middleton, Electronic and Computer Science, University of Southampton, UK

Giorgos Kordopatis-Zilos, Symeon Papadopoulos, Yiannis Kompatsiaris, Information Technologies Institute, CERTH, Greece

DOI: <https://doi.org/10.1145/3202662> Contact: [sem03@soton.ac.uk](mailto:sem03@soton.ac.uk); [www.ecs.soton.ac.uk/people/sem03](http://www.ecs.soton.ac.uk/people/sem03)

## Definitions

**Location** – place name (e.g. London), synonymous with the term ‘toponym’, ‘location phrase’ and ‘location mention’.

**Geocoding** - transform well-formed textual representation of an address into a valid spatial representation (e.g. spatial coordinate).

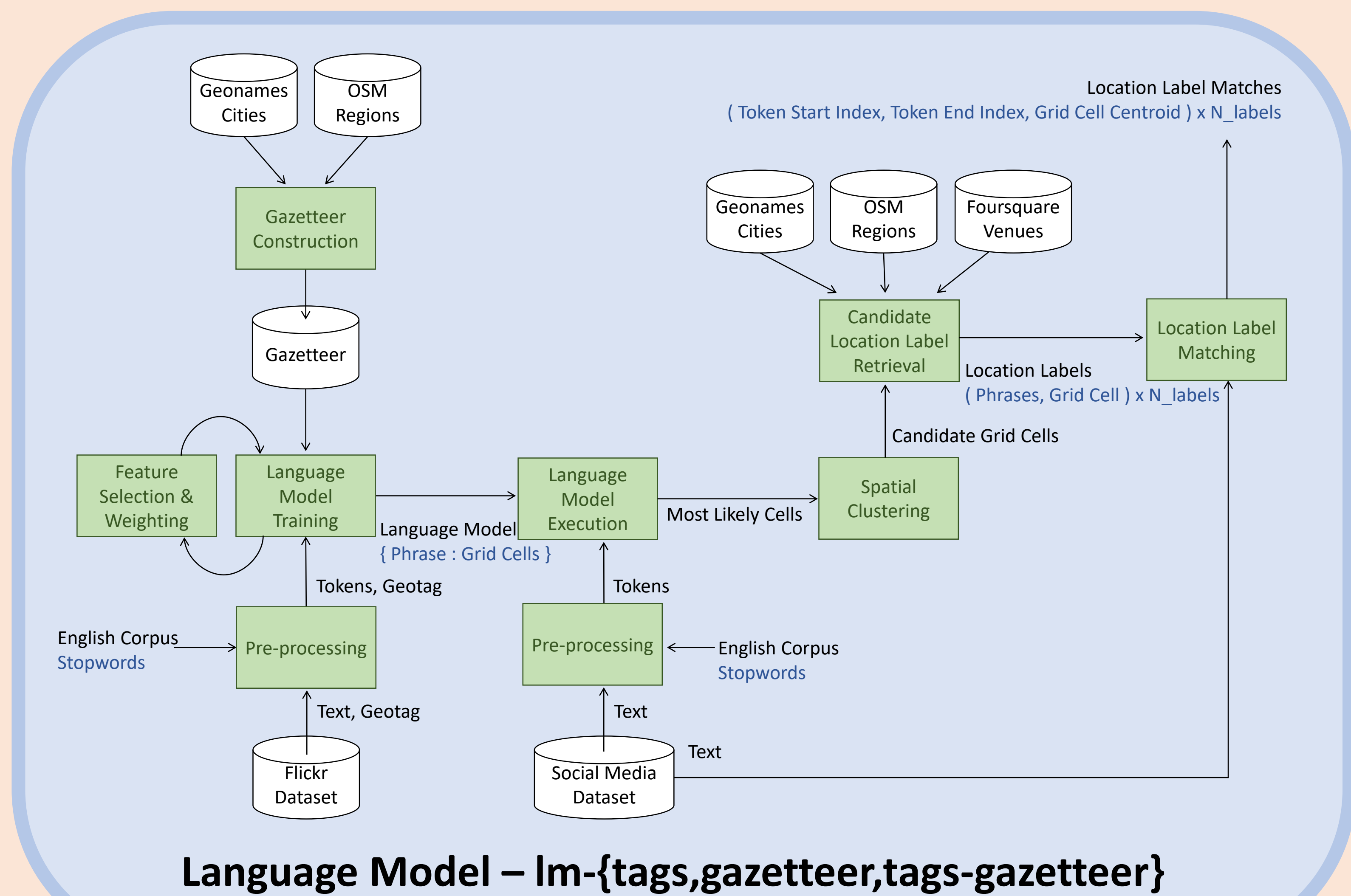
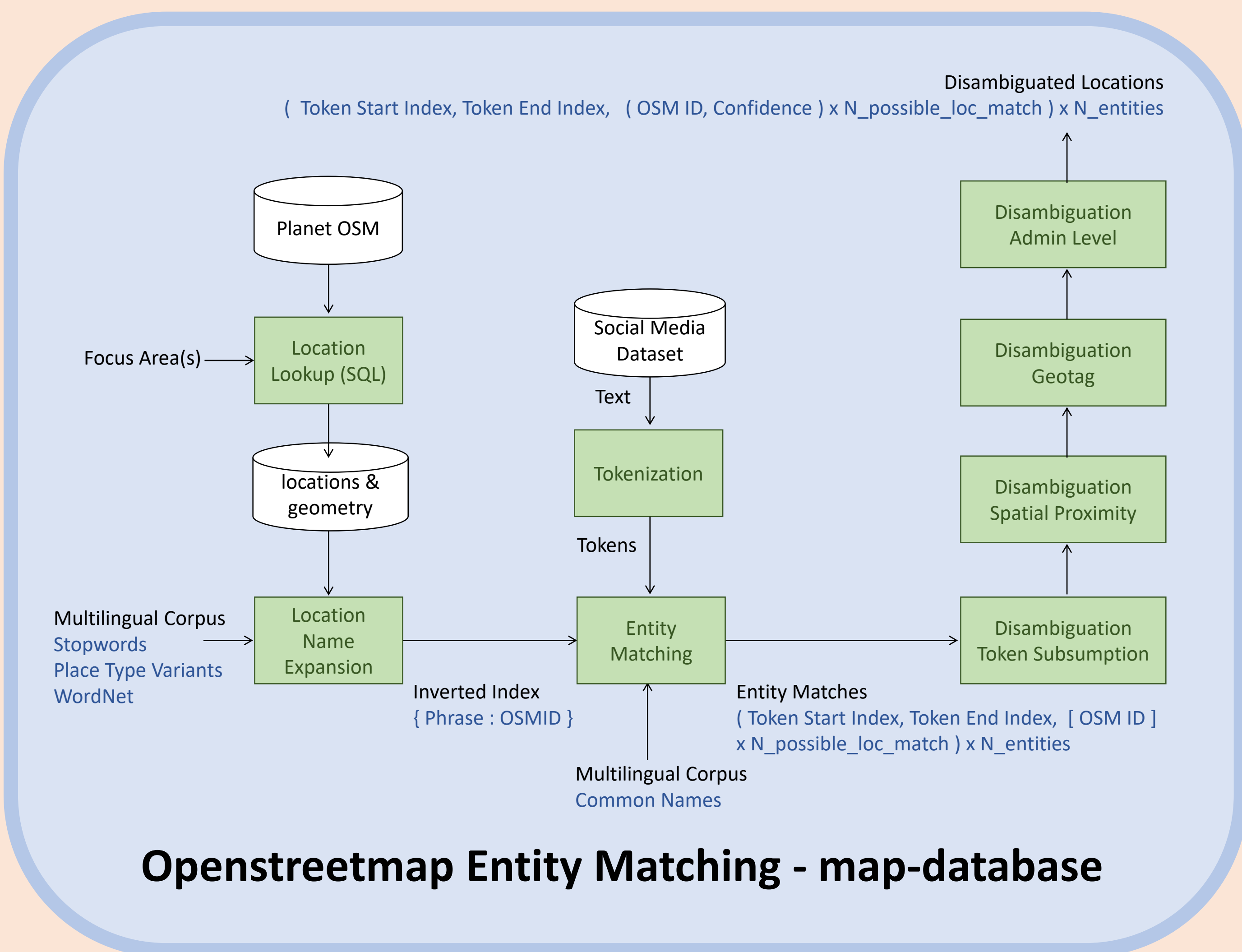
**Geoparsing** - does same for unstructured free text and involves location extraction and location disambiguation prior to geocoding.

**Geotagging** - assigns spatial coordinates to media content items.

## Approach

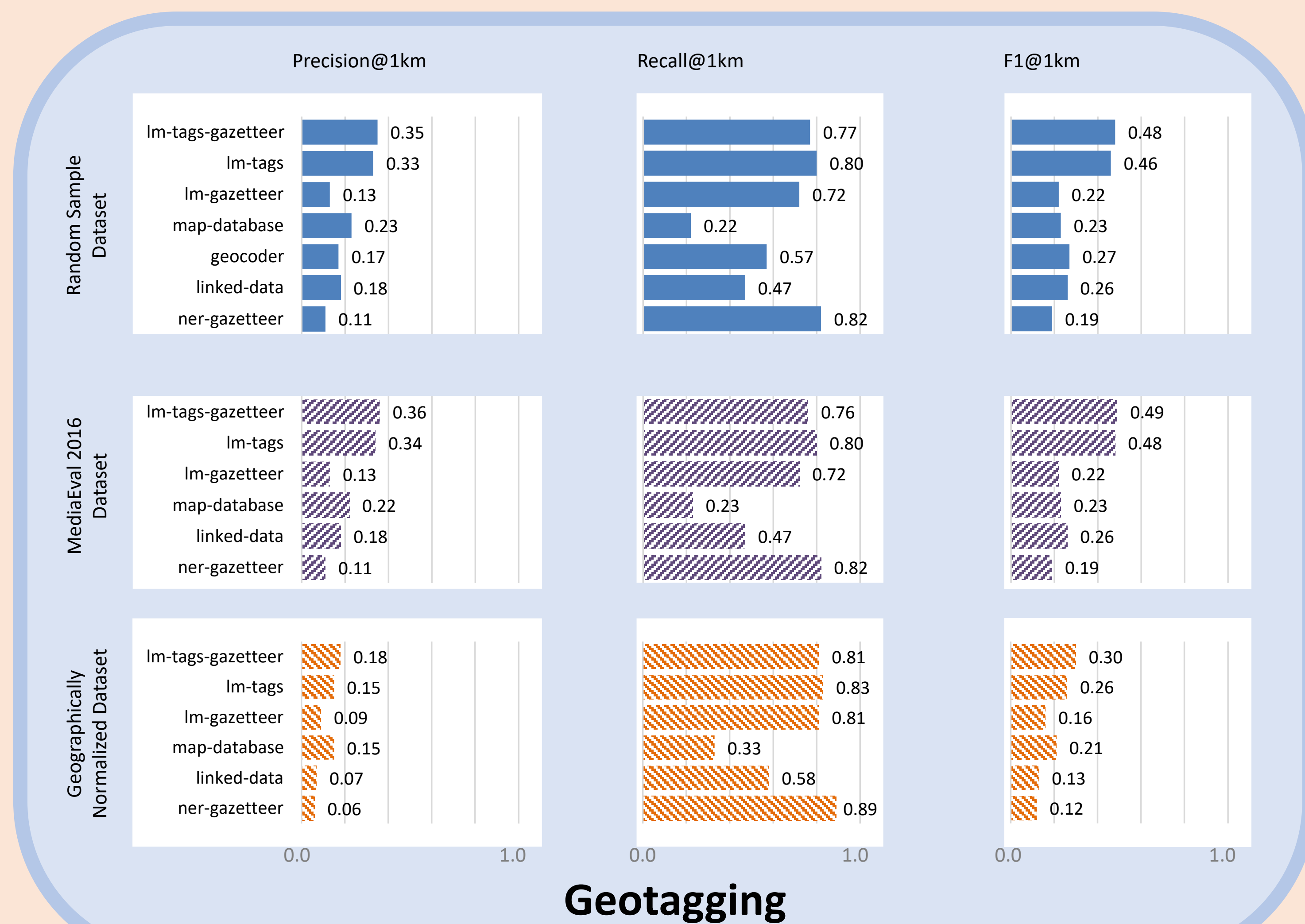
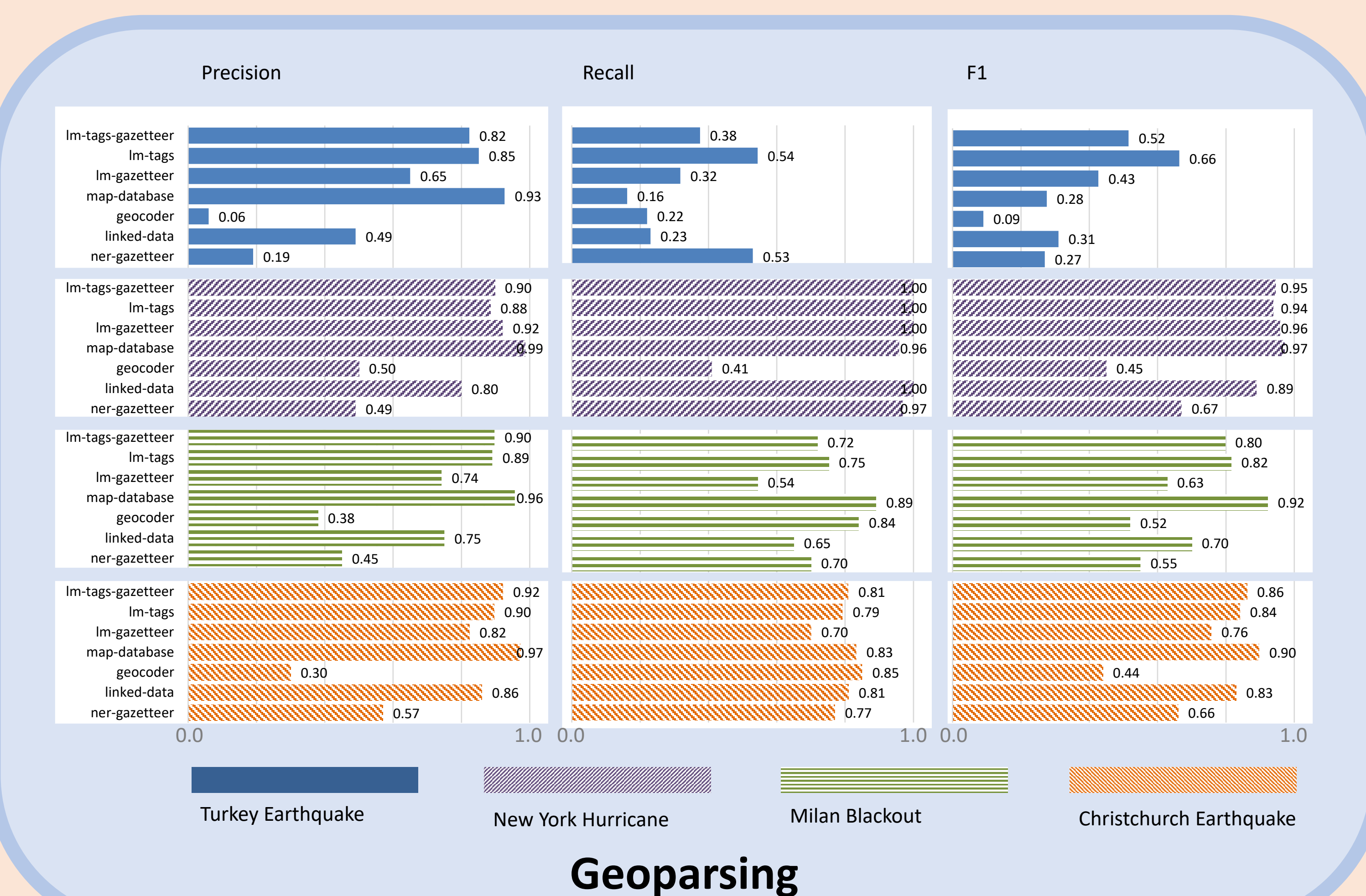
Comparative study among five ‘best of class’ location extraction algorithms for geoparsing and geotagging problems:

- Entity matching using an OpenStreetMap (OSM) database.
- Language model using social media tags and multiple gazetteers.
- DBpedia-based linked-data entity recognition and disambiguation.
- Named entity recognition and GeoNames gazetteer lookup.
- Named entity recognition and the Google Geocoder API.



## Results

Geoparse evaluation using over 6,000 manually labelled tweets. Best overall for English and Italian was **map-database** entity matching with (P 0.96 to 0.99, F1 0.90 to 0.97). For Turkish **Im-tags** approach was best (F1 0.66).



Geotagging evaluation for 1km squares using YFCC100m Flickr dataset with over 39,000,000 geotagged Flickr posts. The **Im-tags-gazetteer** approach was strongest (F1@1km 0.49) and showed the strength of using tag sets for location disambiguation.

## Application

Geoparsing social media to support breaking news verification with Deutsche Welle [REVEAL project]; Geoparsing online forums to map online illegal plant trade with Kew Gardens and UK Border Force [FloraGuard project]; Geoparsing citizen reports to identify trends and opinions about urban development [CUTLER project].

## Acknowledgements

This work is part of the research and development in the 7th Framework and Horizon 2020 Program of the European Commission (REVEAL project 610928, TRIDEC project 258723, InVID project 687786, CUTLER project 770469) and the UK Economic and Social Research Council (FloraGuard project ES/R003254/1).