Mordecai: Full Text Geoparsing and Event Geocoding

Andrew Halterman

1 MIT

Summary

Mordecai is a new full-text geoparsing system that extracts place names from text, resolves them to their correct entries in a gazetteer, and returns structured geographic information for the resolved place name. Geoparsing can be used in a number of tasks, including media monitoring, improved information extraction, document annotation for search, and geolocating text-derived event data, which is the task for which it was built. Mordecai was created to provide several features missing in existing geoparsers, including better handling of non-US place names, easy and portable setup and use through a Docker REST architecture, and easy customization with Python and swappable named entity recognition models. As a gazetteer, it uses Geonames (Geonames 2016) in a custom-build Elasticsearch database.

References
