

GeoTxt

(Web Service to Geolocate Places Mentioned in Microblogs and News Text)

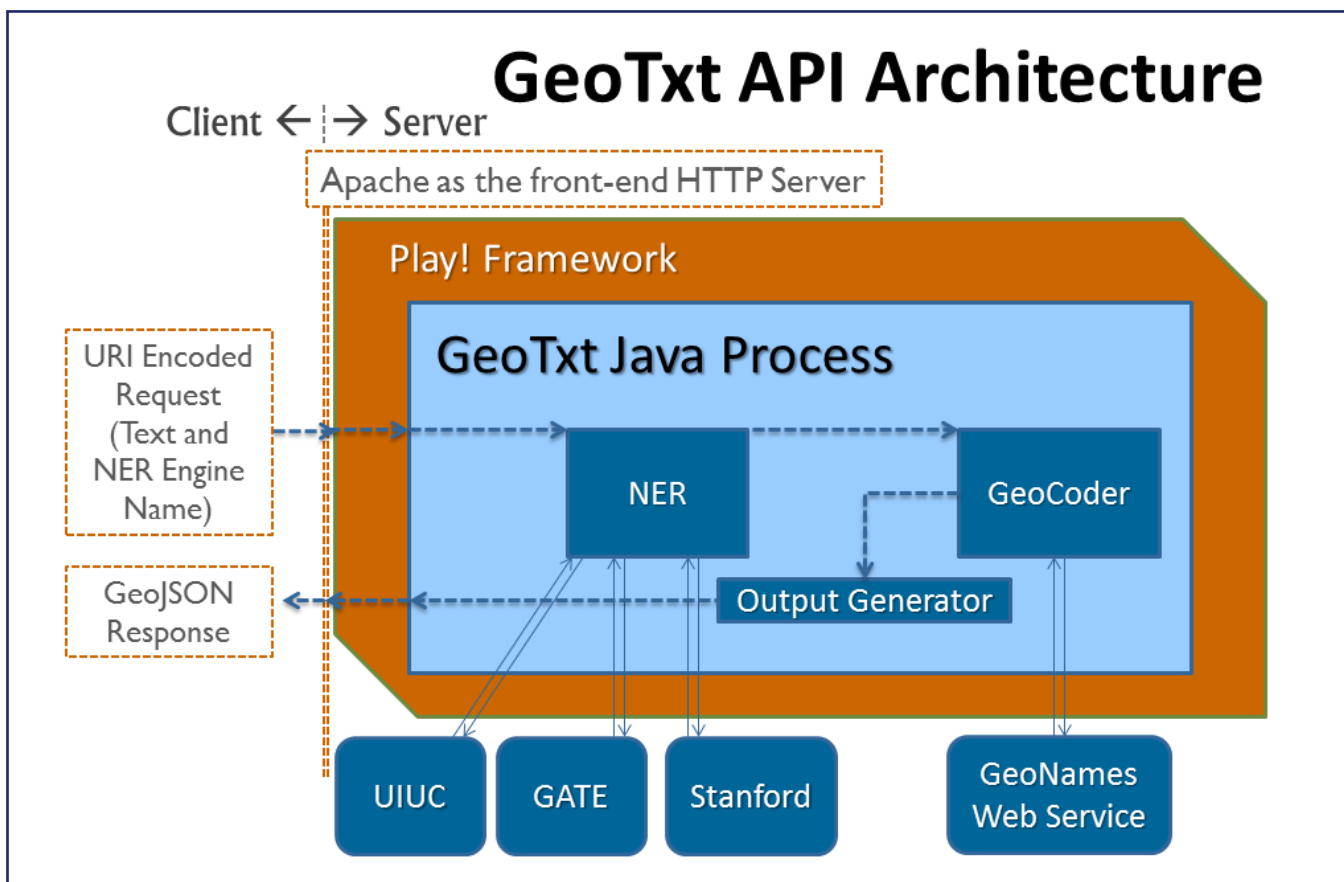
PRODUCT

The Research

Place is critical to homeland security; crisis events happen in place, resources need to be moved from place to place, vulnerable people and infrastructure are at particular places. Place references found in text documents (ranging from news stories and intelligence reports to social media feeds) provide an important complement to formal place-based data already encoded in geospatial databases — if those place references can be recognized, extracted and turned into geospatial data. That capability is what the GeoTxt web service provides, with a specific emphasis on informal text in social media.

What GeoTxt Does

GeoTxt uses techniques from a wide array of research areas — applied linguistics, natural language processing, search engine optimization and geographic information science — to parse out places, people and events explicitly or implicitly mentioned in text (e.g. tweets), and then analyze and contextualize in order to locate them in geographic space.



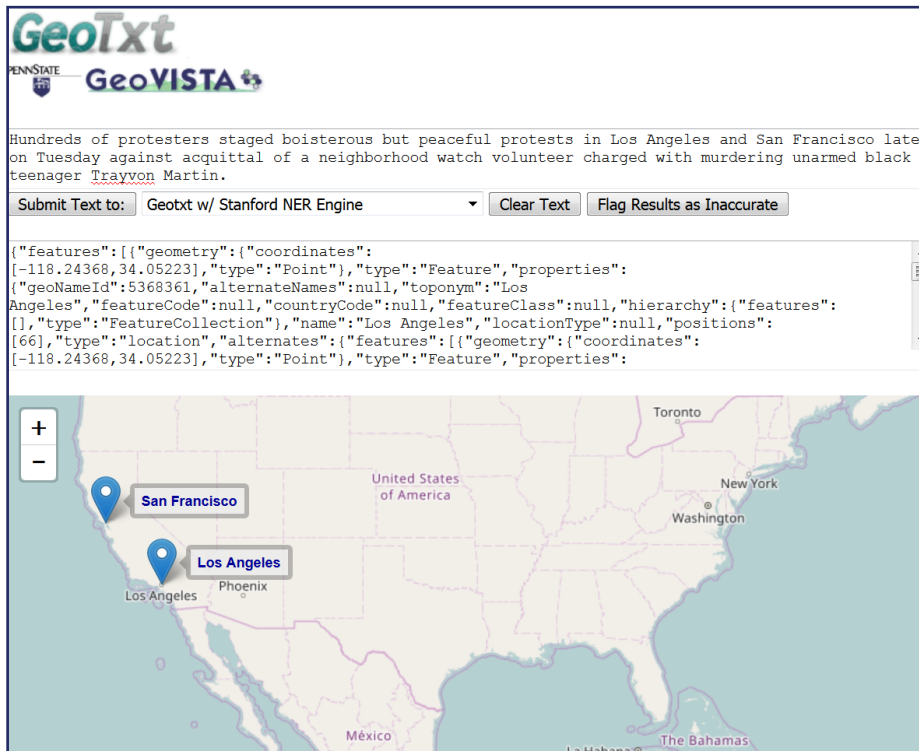
The architecture of GeoTxt's application programming interface.

Why it is Useful

GeoTxt detects locations, people and organizations mentioned in text and assigns geographic coordinates to those locations. This enables references to geographic locations that appear in public posts on social networks, newspapers and other text sources to be harnessed for a variety of analytical tasks.

Accurate geoparsing must address a range of challenges, particularly with Twitter, because users often abbreviate, use nonstandard syntax (e.g. don't capitalize proper nouns) or simply make grammatical mistakes. Additionally, most place names can refer to more than one location (the mean for place names globally is over 100 different locations/names); thus determining the intended location is a challenge.

GeoTxt is specialized to cope with challenges like these. For instance, a tweet that reads "Finally landing in London. I love Canada!" would be geolocated to London, Ontario, instead of London, UK, based on the two locations mentioned in the text — not based upon location population or ranking in GeoNames.



GeoTxt web user interface.

Contact Us

Want to find out how VACCINE's research can help your organization? Email vaccine@purdue.edu or visit www.visualanalytics-CCI.org.

EA/EOU Produced by Purdue Marketing and Media EVPRP-17-8983

