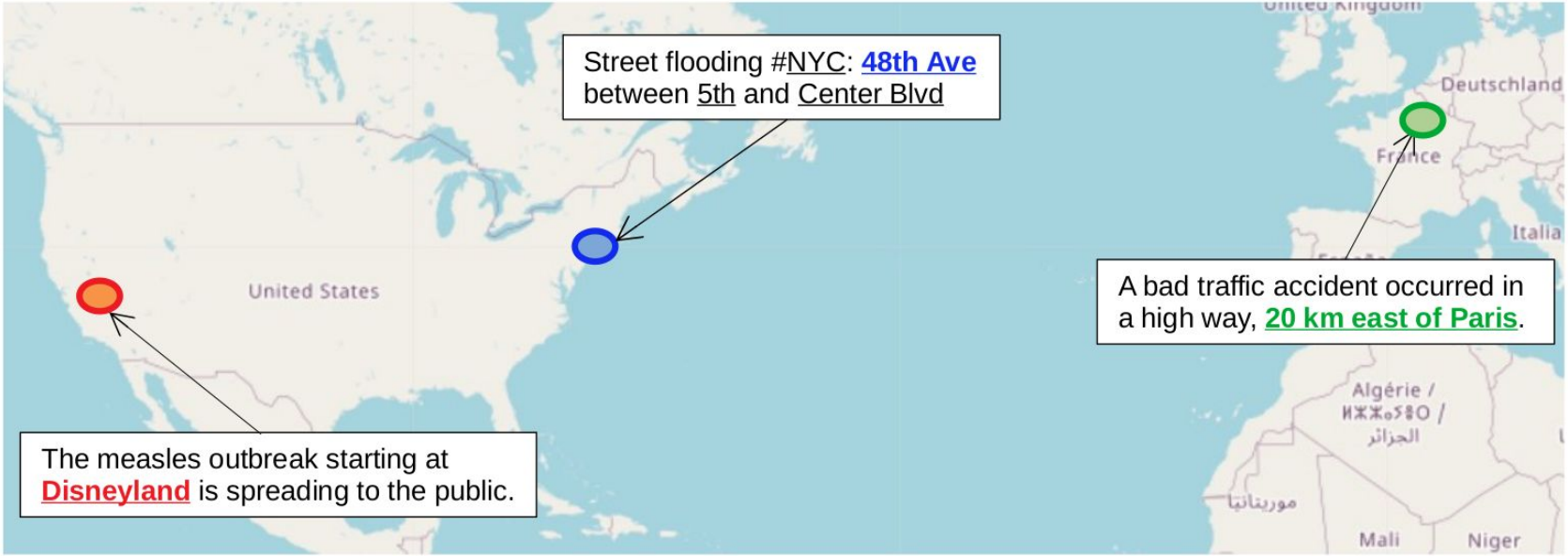
Extract geographic information from natural language texts (GeoExT)

Xuke Hu

Institute of Data Science, German Aerospace Center (DLR)

DLR

Geospatial information in texts



Typical applications

Geographic Information retrieval

Disaster management

Crime management

Geoparsing

Disease surveillance

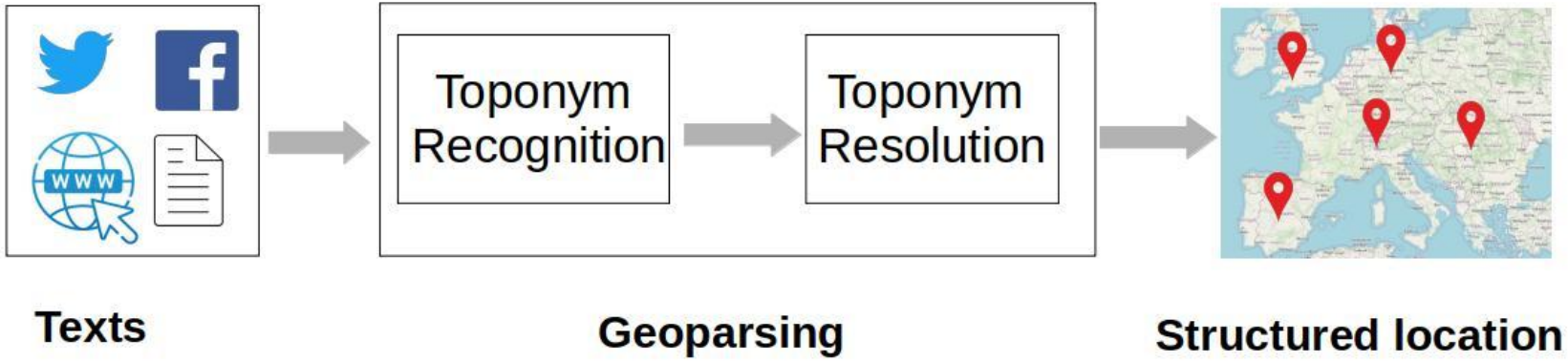
Tourism management

Spatial humanities

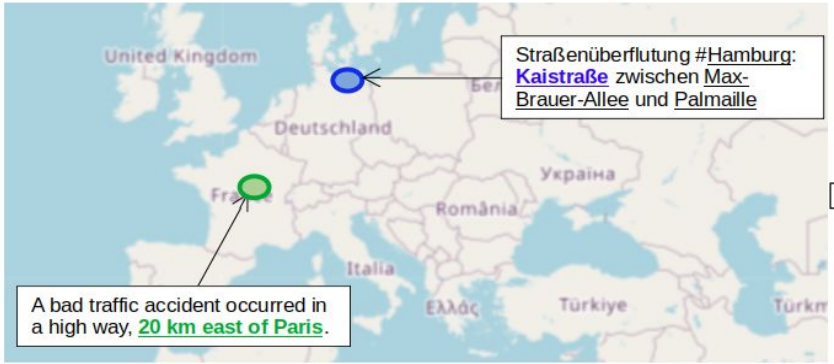
Traffic management

Typical application domains of Geoparsing

Geoparsing

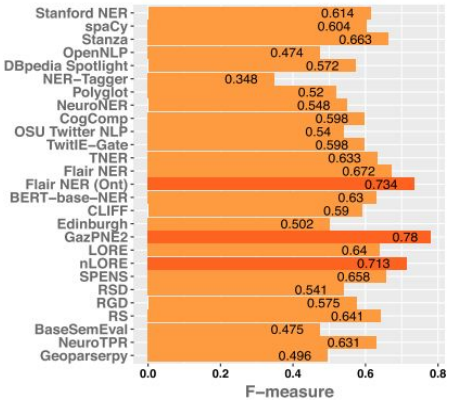

WWW
Toponym Recognition
Toponym Resolution
Texts
Geoparsing
Structured location

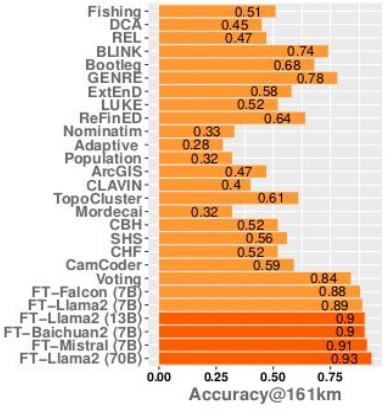
Summary of our work https://github.com/uhuohuy



- Toponym recognition
  - GazPNE: CNN+LSTM+Gazetteers
  - GazPNE2: BERT+BERTTweet+GazPNE
- Toponym resolution
  - Voting
  - LLM + geographic knowledge

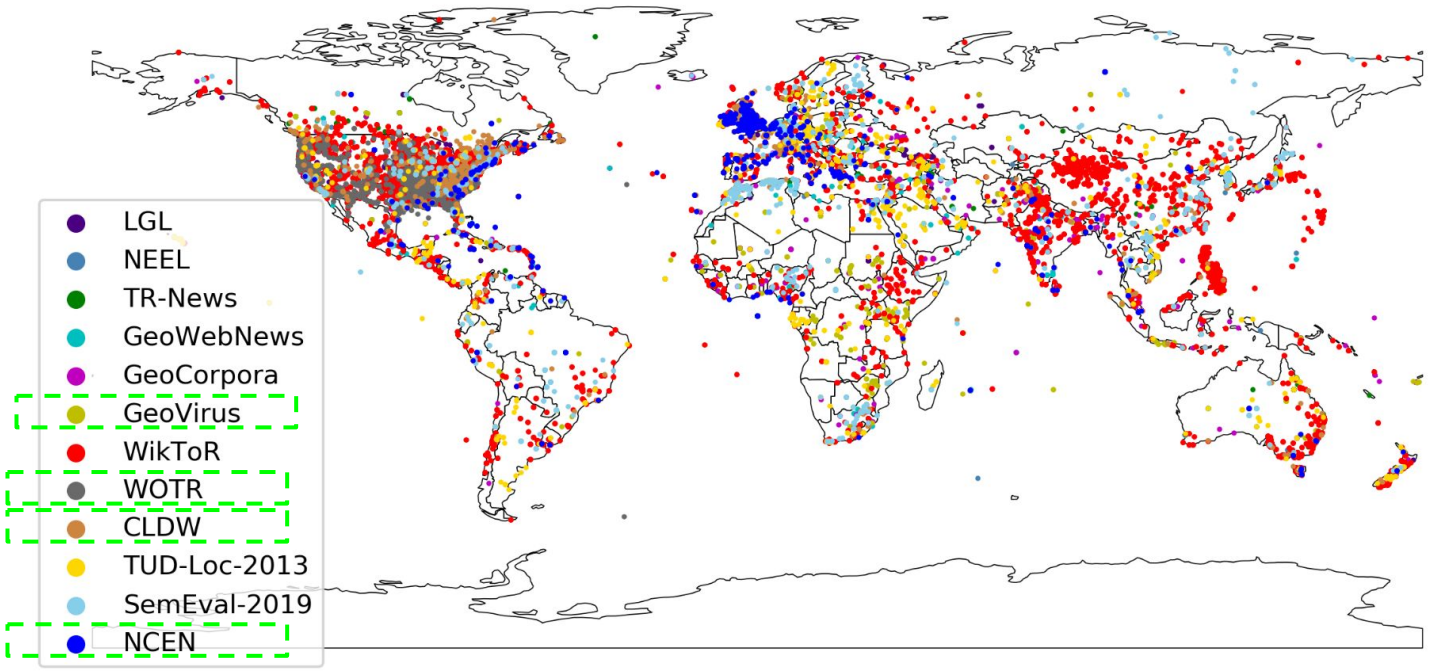
Task	Compared methos	Test dataset	Accuracy
Recognition	27	26	Improve 7%
Resolution	20	12	Improve 17%





Toponym Resolution Datasets

Tweets (2)
History (4)
News (3)
Web Page (1)
Scientific (1)
Wikipedia (1)



Spatial distribution of toponyms in the 12 datasets (98,300)

Publications

1 Hu, X., Kersten, J., Klan, F., et al. (2024). Toponym resolution leveraging lightweight and open-source large language models and geo-knowledge. *International Journal of Geographical Information* Science, 1-28.

2 Hu, X., Elßner, T., Zheng, S., Serere, H. N., Kersten, J., Klan, F., & Qiu, Q. (2024). DLRGeoTweet: A comprehensive social media geocoding corpus featuring fine-grained places. *Information Processing & Management*, 61(4), 103742.

3 Hu, X., Zhou, Z., Li, H., Hu, Y., Gu, F., Kersten, J., & Klan, F. (2023). Location reference recognition from texts: A survey and comparison. *ACM Computing Surveys*, 56(5), 1-37.

4 Hu, X., Sun, Y., Kersten, J., Zhou, Z., Klan, F., & Fan, H. (2023). How can voting mechanisms improve the robustness and generalizability of toponym disambiguation?. *International Journal of Applied Earth Observation and Geoinformation*, 117, 103191.

5 Hu, X., Zhou, Z., Sun, Y., Kersten, J., Klan, F., Fan, H., & Wiegmann, M. (2022). GazPNE2: A general place name extractor for microblogs fusing gazetteers and pretrained transformer models. *IEEE Internet of Things Journal*, 9(17), 16259-16271.

6 Hu, X., Al-Olimat, H. S., Kersten, J., Wiegmann, M., Klan, F., Sun, Y., & Fan, H. (2022). GazPNE: annotation-free deep learning for place name extraction from microblogs leveraging gazetteer and synthetic data by rules. *International Journal of Geographical Information Science*, 36(2), 310-337.

Brainstorm: How can GeoExT benefit UrbanMetaMapping

- Extract overlooked geographic details from archival texts to complement damage maps
- Link textual records to damage maps for deeper analysis
- Cross-check
- …

Third Workshop on Geographic Information Extraction from Texts https://geo-ext.github.io/GeoExT2025/

- **Conference**: 47th European Conference on Information Retrieval
- **Chairs**: Xuke Hu, Ross Purves, Ludovic Moncla, Jens Kersten, Anna Kruspe
- **Location**: Lucca, Italy
- **Submission deadline**: February 2nd, 2025
- **Workshop day**: April 10th, 2025

First GeoExT workshop in Dublin

Second GeoExT workshop in Glasgow