The Journal of Open Source Software

The Irchel Geoparser: A Modular Python Library for Toponym Recognition and Resolution

Diego Gomes [1¶] **and Ross S. Purves** [1]

1 Department of Geography, University of Zurich, Switzerland ¶ Corresponding author

DOI: 10.21105/joss.10365

Software

- Review
- Repository
- Archive

Editor: James Gaboardi

Reviewers:

- @SamHames
- @cforgaci

Submitted: 28 November 2025
Published: 20 November 2026

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

The Irchel Geoparser is an open-source Python library for recognising and resolving place names (toponyms) in unstructured text. Based on a modular architecture, it provides an end-to-end geoparsing pipeline in which the components responsible for toponym recognition, toponym resolution, and the underlying gazetteer data can be adapted or exchanged. The library includes robust, configurable built-in modules and a GeoNames-based gazetteer configuration that together provide a ready-to-use pipeline for English text. Users assemble this pipeline from the modules they choose. They can also fine-tune built-in modules or integrate their own custom recognisers and resolvers. Similarly, it is possible to include domain-specific gazetteers, allowing geoparsing workflows to be tailored to diverse research and practical contexts.

Statement of Need

Geoparsing is an essential step in linking text to geographical locations (Purves et al., 2018), enabling diverse applications including mapping news coverage, exploring historical documents, or analysing discourse across regions (Chesnokova et al., 2019; Grover et al., 2010; Teitler et al., 2008). Typically, geoparsing is a two-stage process involving firstly toponym recognition and subsequently toponym resolution. Toponyms, or place names, are first identified in text, a special case of the more general problem of named entity recognition. Toponym recognition must differentiate between instances of words being used in other contexts - for example to describe generic objects (as for instance in the case of bath (a place to wash oneself and an English city)) or to name people and firms (for example (George) Washington or Zurich (Insurance)). Toponym resolution resolves candidate place names to a unique identifier, dealing with the problem of multiple instances of the same place name (for example London, UK vs. London, Ontario, Canada).

The Irchel Geoparser was designed to provide a modular Python library which could easily be incorporated into text processing chains by assembling a complete pipeline from its built-in modules. As such, it is well suited to teaching applications introducing, for example, graduate students to the geoparsing process. At the same time, because of its modular design, the entire processing chain is configurable to specific use cases - for example by swapping gazetteers to use those more appropriate to historical documents or retraining language models underlying toponym resolution modules on a specific corpus. Thus, the Irchel Geoparser is also well suited for use in research. Finally, we also aimed to provide a well documented Python library with an open licence allowing not only experimentation but also reproducibility of results.

State of the Field

Although several geoparsers already exist, they have limitations that restrict their practical applicability. Systems such as the Edinburgh Geoparser (Grover et al., 2010), CLAVIN (Berico

Technologies, 2012), GeoTxt (Karimzadeh et al., 2019), and the Mordecai family (Halterman, 2017, 2023) have contributed significantly to the field, but they generally require substantial effort for deployment and offer only limited flexibility in the customisation of the geoparsing process. Adapting these tools to a specific task, whether by incorporating new models for languages other than English, using specialised gazetteers or aligning the pipeline to a domain typically involves separate forks of the software, where this is possible, and these are often not well documented.

The Irchel Geoparser addresses these limitations by providing a clearly defined, interchangeable geoparsing pipeline packaged as an installable Python library. Its built-in modules offer good performance for English text and all stages of the pipeline can be adapted or exchanged entirely. This makes the Irchel Geoparser suitable both as a ready-to-use tool and as an experimental platform for researchers developing new geoparsing methods. By developing a modular architecture, we make it possible for other researchers to integrate customised modules addressing individual elements of the geoparsing workflow - for example recent work using LLMs to enhance toponym recognition (Yan & Lee, 2024) or resolution (Anuradha et al., 2025; Hu et al., 2024; Yan & Lee, 2024).

Software Design

The Irchel Geoparser treats geoparsing as a pipeline of recogniser and resolver modules that users assemble and that adhere to simple, well-defined interfaces. Any Python class implementing these interfaces can be used within the pipeline, allowing users to integrate custom methods while relying on the system for data handling, caching, and tracking. Documents and geoparsing results can be persisted in a local SQLite project, keeping raw text, recognised toponyms, resolver outputs and auxiliary metadata together for long-term workflows.

The default toponym recogniser uses spaCy's named-entity recognition functionality to identify toponyms based on entity labels including Geopolitical Entity (GPE), Location (LOC), or Facility (FAC) (Honnibal et al., 2020). Because it builds directly on spaCy models, users can easily switch languages, use fine-tuned models or incorporate custom pipelines by specifying a different model. The default toponym resolver incorporates a transformer-based disambiguation method presented in prior work (Gomes et al., 2024). It retrieves candidate locations from a gazetteer, embeds both the mention context and candidate descriptions using a transformer model, and ranks candidates based on embedding similarity. Users can choose to use the default model that has been fine-tuned using English news articles linked to GeoNames, or they can supply their own transformer checkpoints, which can then be trained for use with any configured gazetteer.

The gazetteer layer itself is also customisable. While a GeoNames-based configuration is provided by default, users can create gazetteer configuration files to build custom gazetteers, combining different data sources as required. This makes integrating gazetteer data created for, for example, historical georeferencing (Grossner et al., 2023) or by national mapping agencies, straightforward. Since modules only depend on the gazetteer interface, switching gazetteers is a configuration change on the resolver and leaves the rest of the pipeline unchanged.

Research Impact Statement

The Irchel Geoparser has been used in both research and teaching contexts. For example, it supported a study on geographic bias in global climate disaster reporting, in which thousands of news articles were analysed and mapped to reveal patterns in the geographic focus of media attention (Kong & Purves, 2025). Hanny et al. (2025) also used the Irchel Geoparser to analyse the locations of disasters described in social media content before carrying out a spatial-temporal analysis of emotions. Subsequent studies extended this work to multimodal disaster detection (Hanny et al., 2026) and to the evaluation of multimodal topic models

of disaster-related social media (Jalilian et al., 2026). Anderson & Giordano (forthcoming) integrated the Irchel Geoparser in a forthcoming publication to explore the locations described in holocaust testimonies. It was also used in a number of Master's theses, for instance to investigate how newspaper discourses reflected the return of the wolf to Switzerland in the last two decades (Besse, 2025), to explore global discourses of disaster resilience (Haus, 2025).

Furthermore, the straightforward installation, comprehensive documentation, and a complete demo available as a Docker image (Gomes, 2026) also make the Irchel Geoparser an accessible tool for teaching and for illustrating key concepts in geographic information retrieval and we have used it in multiple workshops and courses.

AI Usage Disclosure

The Python library described in this paper was developed over a period of about two years. Generative AI was used to assist in software development, for example in brainstorming for potential architectures, to automate and assist in the writing of functions and classes, to create tests and to generate documentation. For brainstorming ChatGPT and Claude were the primary models used, and for coding Claude. Due to the rapid development of these models during the software development period, we do not give specific version numbers here. All code and documentation was reviewed and validated by the authors, who made the core design decisions. Claude was used in the initial drafting of parts of this manuscript. The final version was reviewed and edited by the authors, who take full responsibility for its content.

Acknowledgements

The development of the Irchel Geoparser was supported by the GIS Hub, a project funded by the Digital Society Initiative (DSI) of the University of Zurich. Additional funding was provided by the Public Data Lab of the Digitalization Initiative of the Zurich Higher Education Institutions (DIZH). We thank Michele Volpi for his involvement in developing the first version of the Irchel Geoparser, and Nicolas Spring for his contributions as an open-source collaborator. Finally, we thank the reviewers and editor of the paper and software for their very supportive and constructive comments.

References

Anderson, C. J., & Giordano, A. (forthcoming). Natural Language Processing and the Spatial Humanities: Comparing Geoparsing of Holocaust Testimonies With GPT-4.1, Deep Learning, and Rules-Based NLP. *Forthcoming*. Forthcoming.

Anuradha, I., Sumanathilaka, D. K., Mitkov, R., & Rayson, P. (2025). Toponym Resolution: Will Prompt Engineering Change Expectations? *Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing*, 95–104. https://doi.org/10.26615/978-954-452-098-4-011

Berico Technologies. (2012). *CLAVIN (Cartographic Location And Vicinity INdexer)*.

Besse, M. (2025). *The wolf debate in Switzerland: A spatial and temporal analysis performing geographic information retrieval on newspaper articles* [Master's thesis, University of Zurich]. https://doi.org/10.5167/uzh-436325

Chesnokova, O., Taylor, J. E., Gregory, I. N., & Purves, R. S. (2019). Hearing the silence: finding the middle ground in the spatial humanities? Extracting and comparing perceived silence and tranquillity in the English Lake District. *International Journal of Geographical Information Science*, *33*(12), 2430–2454. https://doi.org/10.1080/13658816.2018.1552789

Gomes, D. (2026). *Irchel Geoparser Demo*. https://docs.geoparser.app/en/latest/demo.html

Gomes, D., Purves, R. S., & Volpi, M. (2024). Fine-Tuning Transformers for Toponym Resolution: A Contextual Embedding Approach to Candidate Ranking. In X. Hu, R. S. Purves, L. Moncla, J. Kersten, & K. Stock (Eds.), *Proceedings of the GeoExT 2024: Geographic information extraction from texts workshop* (Vol. 3683, pp. 43–51). CEUR-WS. https://ceur-ws.org/Vol-3683/paper6.pdf

Grossner, K., Grunewald, S., & Mostern, R. (2023). Bringing places from the distant past to the present: a report on the World Historical Gazetteer. *International Journal on Digital Libraries*, *24*(3), 159–162. https://doi.org/10.1007/s00799-022-00341-2

Grover, C., Tobin, R., Byrne, K., Woollard, M., Reid, J., Dunn, S., & Ball, J. (2010). Use of the Edinburgh geoparser for georeferencing digitized historical collections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1925), 3875–3889. https://doi.org/10.1098/rsta.2010.0149

Halterman, A. (2017). Mordecai: Full Text Geoparsing and Event Geocoding. *Journal of Open Source Software*, *2*(9), 91. https://doi.org/10.21105/joss.00091

Halterman, A. (2023). *Mordecai 3: A Neural Geoparser and Event Geocoder*. arXiv. https://doi.org/10.48550/arXiv.2303.13675

Hanny, D., Ghosh Dastidar, K., Wieland, M., Granitzer, M., & Resch, B. (2026). Towards multimodal geospatial reasoning: a foundation model approach for disaster detection from social media, news, and weather data. *Natural Hazards*, *122*(11), 459. https://doi.org/10.1007/s11069-026-08191-w

Hanny, D., Schmidt, S., & Resch, B. (2025). Bluesky as a social media data source for disaster management: investigating spatio-temporal, semantic and emotional patterns for floods and wildfires. *Journal of Computational Social Science*, *9*(1), 16. https://doi.org/10.1007/s42001-025-00448-x

Haus, L. (2025). *Framing Disaster Resilience: A Media Discourse Analysis of British Newspapers* [Master's thesis, University of Zurich]. https://doi.org/10.5167/uzh-435937

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. Zenodo. https://doi.org/10.5281/zenodo.1212303

Hu, X., Kersten, J., Klan, F., & Farzana, S. M. (2024). Toponym resolution leveraging lightweight and open-source large language models and geo-knowledge. *International Journal of Geographical Information Science*, 1–28. https://doi.org/10.1080/13658816.2024.2405182

Jalilian, E., Hanny, D., & Resch, B. (2026). Beyond unimodal metrics: a multimodal evaluation framework for geo-referenced content on social media. *GeoInformatica*, *30*(2), 25. https://doi.org/10.1007/s10707-026-00584-0

Karimzadeh, M., Pezanowski, S., MacEachren, A. M., & Wallgrün, J. O. (2019). GeoTxt: A scalable geoparsing system for unstructured text geolocation. *Transactions in GIS*, *23*(1), 118–136. https://doi.org/10.1111/tgis.12510

Kong, I., & Purves, R. S. (2025). Analyzing Geographic Bias of Newspaper Articles Reporting Global Climate Disasters. *Annals of the American Association of Geographers*, 1–19. https://doi.org/10.1080/24694452.2025.2564220

Purves, R. S., Clough, P., Jones, C. B., Hall, M. H., & Murdock, V. (2018). Geographic Information Retrieval: Progress and Challenges in Spatial Search of Text. *Foundations and Trends in Information Retrieval*, *12*(2-3), 164–318. https://doi.org/10.1561/1500000034

Teitler, B. E., Lieberman, M. D., Panozzo, D., Sankaranarayanan, J., Samet, H., & Sperling, J. (2008). NewsStand: a new view on news. *Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. https://doi.org/10.1145/1463434.1463458

Yan, Y., & Lee, J. (2024). GeoReasoner: Reasoning On Geospatially Grounded Context For Natural Language Understanding. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 4163–4167. https://doi.org/10.1145/3627673.3679934