Geospatial Interlinking with JedAI-spatial

Manolis Koubarakis
National and Kapodistrian University of Athens, Greece
koubarak@di.uoa.gr

ABSTRACT

In this talk I will present the framework JedAI-spatial for the interlinking of geospatial data encoded in RDF.

ACM Reference Format:
Manolis Koubarakis. 2022. Geospatial Interlinking with JedAI-spatial. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3487553.3526093

SHORT CV

Manolis Koubarakis is a Professor and Director of Graduate Studies in the Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens. He leads the Artificial Intelligence team (http://ai.di.uoa.gr). He holds a Ph.D. in Computer Science, from the National Technical University of Athens, an M.Sc. in Computer Science, from the University of Toronto, and a diploma (B.Sc.) in Mathematics, from the University of Crete. He is a Fellow of EurAI (European Association for Artificial Intelligence) since 2015 and President of the Hellenic Association for Artificial Intelligence. He is a member of the Advisory Board that implements the Hellenic National Strategy for Artificial Intelligence. He has published more than 200 papers that have been widely cited (7324 citations and h-index 45 in Google Scholar) in the areas of Artificial Intelligence (especially Knowledge Representation), Databases, Semantic Web and Linked Geospatial Data (especially Earth observation data). His research has been financially supported with a total amount exceeding 8 million Euros by the European Commission, the Hellenic Foundation for Research and Innovation, the Greek General Secretariat for Research and Technology, the European Space Agency and industry. Manolis currently participates in H2020 project AI4Copernicus (2021-2023, as Technical Manager of this project which brings Copernicus data the the AI4EU platform) and DeepCube (2021-2023, where he leads the work on Semantic Data Cubes).

MORE DETAILS OF MY TALK

Geospatial data has escalated tremendously over the years. The outbreak of Internet of Things (IoT) devices, smartphones, position tracking applications and location-based services has skyrocketed the volume of geospatial data. For example, 100TB of weather-related data is produced everyday[1]; Uber hit the milestone of 5 billion rides among 76 countries already on May 20, 2017[2]. Web platforms like OpenStreetMap[3] provide an open and editable map of the whole world. Earth observation programmes like Copernicus[4] publish tens of terabytes of geospatial data per day on the Web[5]. For these reasons, geospatial data constitutes a considerable part of Semantic Web data, but the links between its data sources and their geometries are scarce in the Linked Open Data cloud [6, 10].

Geospatial Interlinking aims to cover this gap by associating pairs of geometries with topological relations like those of the Dimensionally Extended 9-Intersection Model (DE-9IM) [2–4]. In Figure 1 for instance, LineString g_3 intersects LineString g_4 and touches Polygon g_1, which contains Polygon g_2. Two are the main challenges of this task: (i) its inherently quadratic time complexity, because it has to examine every pair of geometries, and (ii) the high time complexity of examining a single pair of geometries, which amounts to $O(N \log N)$, where N is the size of the union set of their boundary points [1]. As a result, Geospatial Interlinking involves a high computational cost that does not scale to large Web datasets.

Numerous algorithms aim to address these challenges by enhancing the time efficiency and scalability of Geospatial Interlinking.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France.
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3526093

[1]https://www.ibm.com/topics/geospatial-data
[2]https://www.uber.com/en-SG/blog/uber-hits-5-billion-rides-milestone
[3]https://www.openstreetmap.org
[4]https://www.copernicus.eu
[5]https://www.copernicus.eu/sites/default/files/Copernicus_DIAS_Factsheet_June2018.pdf

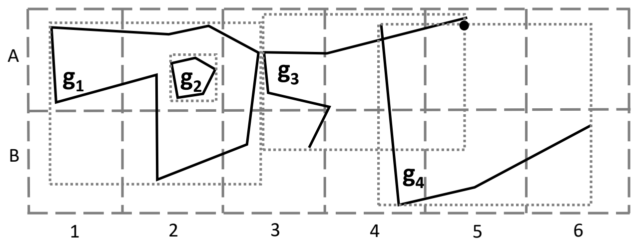

Figure 1: Example of four topologically related geometries.

The most recent ones operate in main memory, reducing the search space to pairs of geometries that are likely to be topologically related according to a geospatial index [8, 14, 16]. However, no open-source system organizes these algorithms into a common framework so as to facilitate researchers and practitioners in their effort to populate the LOD cloud with more topological relations. Systems like Silk [5] and LIMES [7] convey only the methods developed by their creators, Silk-spatial [15] and RADON [13] respectively, while systems that could act as a library of established methods, such as stLD [11, 12], are not publicly available. Moreover, no system supports progressive methods, neither for serial nor for parallel processing, even though they are indispensable for applications with limited computational or temporal resources [9, 10].

To address these issues, I present *JedAI-spatial*, an open-source system that supports a broad range of Geospatial Interlinking applications on the Web by implementing the state-of-the-art methods in the literature. JedAI-spatial makes the following contributions:

- It organizes the main algorithms into a novel taxonomy that facilitates their use and adoption by practitioners and researchers, enabling them to select the most appropriate algorithm based on the requirements of their application.
- Its intuitive user interface supports both novice and expert users.
- Its modular and extensible architecture allows for easily incorporating new algorithms and improvements to the existing ones.
- It optimizes the implementation of existing algorithms, some of which have not been applied to Geospatial Interlinking before.
- We have publicly released the code of JedAI-spatial at: https://github.com/giantInterlinking/JedAI-spatial.

This is joint work with George Papadakis, Nikos Mamoulis (University of Ioannina), Marios Papamichalopoulos, George Mandilaras and Maria Despoina Siampou.

ACKNOWLEDGEMENTS

This work was supported by the H2020 program of the European Commission, projects ExtremeEarth (http://earthanalytics.eu/, contract number 82525) and AI4Copernicus (https://ai4copernicus-project.eu/, contract number 101016798).

REFERENCES

[1] E. P. F. Chan and J. N. H. Ng. A general and efficient implementation of geometric operators and predicates. In *SSD*, pages 69–93, 1997.
[2] E. Clementini, P. D. Felice, and P. van Oosterom. A small set of formal topological relationships suitable for end-user interaction. In *SSD*, pages 277–295, 1993.
[3] E. Clementini, J. Sharma, and M. J. Egenhofer. Modelling topological spatial relations: Strategies for query processing. *Comput. Graph.*, 18(6):815–822, 1994.
[4] M. J. Egenhofer and R. D. Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information System*, 5(2):161–174, 1991.
[5] A. Jentzsch, R. Isele, and C. Bizer. Silk - generating RDF links while publishing or consuming linked data. In *Proceedings of the ISWC 2010 Posters & Demonstrations Track: Collected Abstracts*, 2010.
[6] A. N. Ngomo. ORCHID - reduction-ratio-optimal computation of geo-spatial distances for link discovery. In *ISWC*, pages 395–410, 2013.
[7] A. N. Ngomo and S. Auer. LIMES - A time-efficient approach for large-scale link discovery on the web of data. In *IJCAI*, pages 2312–2317, 2011.
[8] V. Pandey, A. Kipf, T. Neumann, and A. Kemper. How good are modern spatial analytics systems? *Proc. VLDB Endow.*, 11(11):1661–1673, 2018.
[9] G. Papadakis, G. Mandilaras, N. Mamoulis, and M. Koubarakis. Static and dynamic progressive geospatial interlinking. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 2022. Forthcoming.
[10] G. Papadakis, G. M. Mandilaras, N. Mamoulis, and M. Koubarakis. Progressive, holistic geospatial interlinking. In *WWW*, pages 833–844, 2021.
[11] G. M. Santipantakis, C. Doulkeridis, A. Vlachou, and G. A. Vouros. Integrating data by discovering topological and proximity relations among spatiotemporal entities. In *Big Data Analytics for Time-Critical Mobility Forecasting, From Raw Data to Trajectory-Oriented Mobility Analytics in the Aviation and Maritime Domains*, pages 155–179. 2020.
[12] G. M. Santipantakis, A. Glenis, C. Doulkeridis, A. Vlachou, and G. A. Vouros. stld: towards a spatio-temporal link discovery framework. In *Proceedings of the International Workshop on Semantic Big Data, SBD@SIGMOD*, pages 4:1–4:6, 2019.
[13] M. A. Sherif, K. Dreßler, P. Smeros, and A. N. Ngomo. Radon - rapid discovery of topological relations. In *AAAI*, pages 175–181, 2017.
[14] D. Sidlauskas and C. S. Jensen. Spatial joins in main memory: Implementation matters! *Proc. VLDB Endow.*, 8(1):97–100, 2014.
[15] P. Smeros and M. Koubarakis. Discovering spatial and temporal links among RDF data. In *Workshop on Linked Data on the Web, LDOW*, 2016.
[16] B. Sowell, M. A. V. Salles, T. Cao, A. J. Demers, and J. Gehrke. An experimental analysis of iterated spatial joins in main memory. *Proc. VLDB Endow.*, 6(14):1882–1893, 2013.