International Cartographic Association
Association Cartographique Internationale

Towards an intuitive User Interface and Geographic Question Answering for an existing spatial Linked Data Endpoint for Dialect Data

Georg Weinberger [a], Johannes Scholz [a,*], Eveline Wandl-Vogt [b]

[a] *RG Geoinformation, Institute of Geodesy, Graz University of Technology, Steyrergasse 30/I, 8010 Graz, Austria, georg.weinberger@student.tugraz.at, johannes.scholz@tugraz.at*
[b] *Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences, Bäckerstraße 13, 1010 Vienna, eveline.wandl-vogt@oeaw.ac.at*

* Corresponding author

Keywords: Linked Data, Digital Humanities, Dialect, GeoAI

Abstract:

In the area of linguistics and dialectology there exists a rich set of data collections on dialect words, their usage, meaning and pronunciation. In the German-speaking area, the Wenker Atlas (Schmidt & Herrgen, 2001) or the Dictionary of Bavarian Dialects in Austria (Österreichische Akademie der Wissenschaften & Bauer, 1985) are well-known first and long-lasting endeavours of such systematic, citizen science alike collections. The data were collected in the early 20th century, using the "indirect method" – which means, using questionnaires linked to specific placenames (which may denote places or areas). Hence, the respective dialect records are also associated to places including a certain time stamp and collectors ID. In recent years, some of these data were digitized and stored in object-relational databases (Scholz et al., 2017). One of the globally first examples of such a database, linked with a dictionary content is the "Datenbank der bairischen Mundarten in Österreich electronically mapped (dbo@ema)" (Wandl-Vogt et al., 2008; Wandl-Vogt 2008). However, it is difficult to combine and query different linguistic, especially dialectal, datasets in an integrated manner, since most datasets lack semantic interoperability (Scholz et al., 2017; Abgaz et al., 2018).

Various projects aim to bridge this semantic gap, additionally offer digital visualizations for a better navigation within the data set and to create better access – such as Benito et al. (2016) and Benito et al. (2017)[1]. Scholz et al. (2017) proposed a spatio-temporal Linked Data model for representing dialect data, originating from the "dbo@ema" dataset. They used existing linguistic data, transformed it into Linked Data, and published it using a virtual RDF graph (Scholz et al., 2017). However, the project did not develop an advanced user interface, capable of querying the developed SPARQL endpoint conveniently.

In this work, we use the approach developed by Scholz et al. (2017) as a starting point and present a user interface (UI) for querying and visualizing spatio-temporal linked dialect data. This approach includes the nearly 32,000 dialect records available from the dbo@ema dataset (Benito et al., 2017), which can also be viewed online on the dbo@ema platform[2]. The advantage of such a user interface is that users do not have to be familiar with the Linked Data query language SPARQL, neither the database schema nor the Linked Data concept to query data from the triplestore. This eases data access, as users can query the database without having to write queries themselves. By interface controls, the users are able to query data, with the help of system generated queries and execution of them. Query results are presented to the user either as text (i.e. the attributes), while the geometries are visualized on a Leaflet map. The user interface is implemented as a web application and is offered as a simple and advanced version - where the advanced version offers a richer set of functionalities.

Software requirements for the user interface are determined in a design workshop with several stakeholders. During the design process of the software, it was considered that users have different interests and levels of experience in the field of linguistics. Therefore, the navigation behavior of users also differs significantly. To meet these circumstances, various access routes to query linguistic and ethnographic data have been implemented (Wandl-Vogt, 2010):

[1] https://exploreat.acdh-dev.oeaw.ac.at/exploreAT-collectionexplorer/
[2] https://dboema.acdh.oeaw.ac.at

Abstracts of the International Cartographic Association, 5, 2022.
European Cartographic Conference – EuroCarto 2022, 19–21 September 2022, TU Wien, Vienna, Austria.
https://doi.org/10.5194/ica-abs-5-68-2022 |

- **Location-Specific Navigation**
  (“Which dialect words from the location XY are stored in the dbo@ema?”)
  User: Inexperienced, does not have to be familiar with the linguistic content;
  Navigation: Query content assigned to a specific location;
  Result: Dialect data assigned to the respective location;

- **Lemma-Specific Navigation**
  (“To which locations is the lemma XY assigned to?”)
  User: Experienced, familiar with the content;
  Navigation: Query content assigned to a specific lemma (standardized headword);
  Result: Dialect data and locations linked to the lemma;

- **Location-Specific Navigation Independent from rigid Boundaries**
  (“Which dialect words are assigned to a specific area?”)
  User: Experienced, familiar with the content;
  Navigation: Query content from a user-defined area instead of locations with rigid boundaries;
  Result: Dialect data assigned to locations intersecting the query area;

- **Spatial Distribution of Similar Dialect Words**
  (“In which locations are dialect words that are similar to the word XY?”)
  User: Experienced, familiar with the content;
  Navigation: Show the spatial distribution of similar dialect words based on a search term;
  Result: Locations, including dialect words similar to the search term;

- **Spatial Distribution of Dialect Data**
  (“How many dialect words are assigned to each location?”)
  User: Experienced, familiar with the content;
  Navigation: Show the spatial distribution of dialect words in relation to locations;
  Result: Locations, including the amount of dialect words assigned to them;

Central to our approach is the implementation of a question answering (QA) system based on Linked Data (Mishra & Chain, 2016). A QA system takes questions in natural language, proposed by the user, analyzes and translates them into Linked Data queries, executes them and presents the results to the user. Our implementation utilizes a template-based approach, which means that the system analyzes the natural language question and executes them using pre-defined template queries based on the question type and the content of the question. As the system also supports geographic questions, it can be referred to as a geographic question answering (GeoQA) system. Geographic questions are questions that contain geographic entities, geographic concepts (e.g. regions, rivers) and/or spatial relationships (e.g. intersects, within) (Mai et al., 2021).

Acknowledgements

This research was partially funded by the Nationalstiftung of the Austrian Academy of Sciences Sciences under the funding scheme: Digitales kulturelles Erbe, grant number DH2014/22, as part of the exploreAT! project, and the PROVIDEDH project, funded within the CHIST-ERA programme under the national grant agreement FWF (Project number I 3441-N33).

References

Abgaz, Yalemisew et al. (2018): A Semantic Model for Traditional Data Collection Questionnaires Enabling CulturalAnalysis. John P. McCrae, Chiarcos, Christian, Declerck, Thierry, Gracia, Jorge, and Klimek, Bettina. Proceedings of the LREC 2018 Workshop "6th Workshop on Linked Data in Linguistics (LDL-2018)". Miyazaki.

Benito, Alejandro et al. (2017): Exploring Lemma Interconnections in Historical Dictionaries. Phoenix.

Benito, Alejandro et al. (2016): A spatio-temporal visual analysis tool for historical dictionaries. Francisco José Garcia-Peñalvo. TEEM '16. Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality. Salamanca, Spain – November 02 - 04, 2016. New York: ACM. doi:10.1145/3012430.3012636.

https://doi.org/10.5194/ica-abs-5-68-2022 |

Mai, G., Janowicz, K., Zhu, R., Cai, L., & Lao, N. (2021). Geographic Question Answering: Challenges, Uniqueness, Classification, and Future Directions. AGILE: GIScience Series, 2, 1-21.

Mishra, A. and Jain, S. K.: A survey on question answering systems with classification, Journal of King Saud University - Computer and Information Sciences, 28, 345–361, 2016.

Österreichische Akademie der Wissenschaften& Bauer, W. (1985). Wörterbuch der bairischen Mundarten in Österreich (WBÖ). Verlag der Österreichischen Akademie der Wissenschaften.

Schmidt, J. E., & Herrgen, J. (2001). Digitaler Wenker-Atlas (DiWA). Bearbeitet von Alfred Lameli, Tanja Giessler, Roland Kehrein, Alexandra Lenz, Karl-Heinz Müller, Jost Nickel, Christoph Purschke und Stefan Rabanus. Erste vollständige Ausgabe von Georg Wenkers "Sprachatlas des Deutschen Reichs.

Scholz, J., Hrastnig, E., & Wandl-Vogt, E. (2017, September). A spatio-temporal Linked Data Representation for modeling spatio-temporal Dialect Data. In International Conference on Spatial Information Theory (pp. 275-282). Springer, Cham.

Wandl-Vogt, E. (2008): ..wie man ein Jahrhundertprojekt zeitgemäß hält: Datenbankgestützte Dialektlexikografie am Institut für Österreichische Dialekt- und Namenlexika (I Dinamlex) (mit 10 Abbildungen). In: Ernst, Peter (Eds.): Bausteine zur Wissenschaftsgeschichte von Dialektologie/germanistischer Sprachwissenschaft im 19. und 20. Jahrhundert. Beiträge zum 2. Kongress der internationalen Gesellschaft für Dialektologie des Deutschen, Wien, 20.-23. September 2006. Wien, PP 93- 112. (= Bausteine zur Wissenschaftsgeschichte von Dialektologie/ germanistischer Sprachwissenschaft im 19. und 20. Jahrhundert).

Wandl-Vogt, E. (2010). Multiple access routes. The dictionary of Bavarian dialects in Austria / Wörterbuch der bairischen Mundarten in Österreich (WBÖ). In Granger, S., Paquot, M. (Eds.), eLexicography in the 21st Century: New Challenges, New Applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22–24 October 2009 (= Cahiers du Cental 7) (pp. 451–455). Louvain-la-Neuve, France: Presses Univ. de Louvain.

Wandl-Vogt, E., Kop, C., Fliedl, G., Nickel, J., & Scholz, J. (2008, July). Database of Bavarian Dialects (DBÖ) electronically mapped (dbo@ema). A System for Archiving, Maintaining and Field Mapping of Heterogeneous Dialect Data for the Compilation of Dialect Lexicons. In Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008). Barcelona: Institut Universitari de Lingüistica Aplicada/Universitat Pompeu Fabra. CD.

https://doi.org/10.5194/ica-abs-5-68-2022