Geodata Generation and Enrichment via ChatGPT for Location Based Services (LBS)

[1]Andreas Keler*, [2]Jukka M. Krisp

[1]Department of Urban Management, Kyoto University, Uji, Japan, [2]Applied Geoinformatics, University of Augsburg, Germany

Abstract. This short paper focuses on the generation and enrichment of geospatial data using ChatGPT (Chat Generative Pre-Trained Transformer, version 3, April 2023). Can we use ChatGPT to generate & use geospatial data in Locations Based Services? We conduct several example generations using the CSV and GeoJSON data formats for the feature classes points, lines and polygons. First results show the potential for educational and research purposes. The initial experiments, for example for tourism sites in Augsburg, indicate that is possible to generate geographic data, still the quality of the information is questionable. As it is generated data, it has shortages and challenges concerning for example the certainty & vagueness concerning the geographic components. In an outlook, we give first ideas for the application of the more promising approach of geodata enrichment by using ChatGPT and its possible use in location-based services (LBS).

Keywords. Artificial Intelligence Chatbots, Geodata Generation, Location-Based Services, ChatGPT, Geographic Data

1. Introduction

In recent years, the popularity of using generative AI in education and research has changed. As artificial intelligence chatbots are able to facilitate the academic writing process and improve readability in different languages, we can expect qualitative changes in selected outcomes of texts and software. The latter refers to the ability of ChatGPT (Chat Generative Pre-Trained Transformer) to optimize language use - even for programming languages. These expectations are currently being confirmed more and more, as in an initial study by Taecharungroj (2023).

In this work in progress research, we investigate the possibility of using ChatGPT to generate geospatial data in different feature classes such as points, trajectories, lines and polygons. The focus of this work is on generating geospatial data for the purpose of providing test applications for location-based services. This can be seen as an alternative method for the case that no or only sparse geodata are available. It also applies to the presence of geodata with only little attribute information. Thus, besides the generation of geodata (and the associated attribute information), another idea would be the enrichment of already existing geodata with additional attributes.

One possible application for ChatGPT is the enrichment of geodata for tourism sites. It could provide an option for capturing local knowledge (along with event information). This is partly related to ongoing research activities on the formalization of Points of Interests (POIs), such as by Psyllidis et al. (2022), and would in our opinion greatly benefit current developments.

Geodata Generation via ChatGPT – Selected Examples

The steps for creation and enrichment are provided by using ChatGPT prompts. We document our previous steps mainly using the responses and data sets generated by ChatGPT.

1.1. Generation of Point Data – POIs as an Example

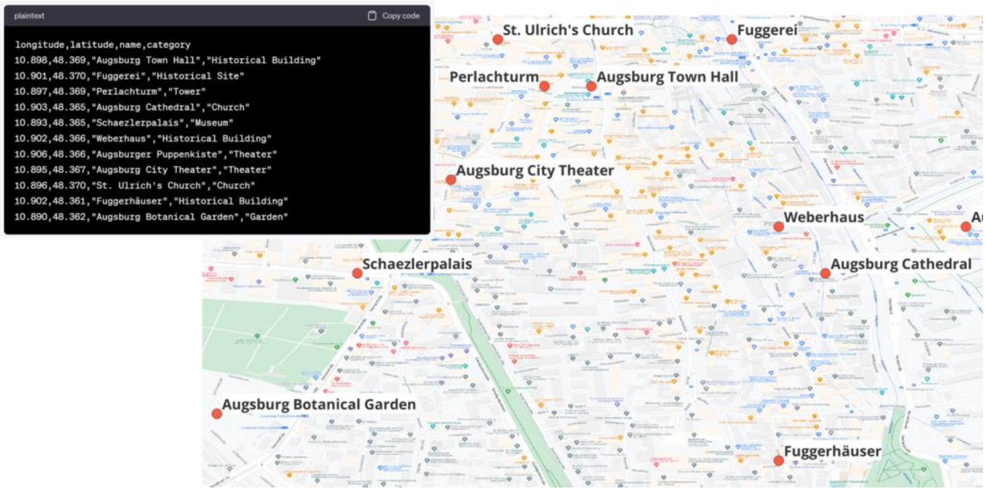

Figure 1. Visualized CSV file of landmarks in Augsburg, Germany.

In *Figure 1* we see a visualization of 11 landmarks (tourist sites) in Augsburg, only a few of which actually correspond to the exact location (prompt: "Create a csv file of all landmarks in Augsburg"). This could indicate that the quality of the location data is rather low. Nevertheless, the selection of landmarks in Augsburg, Germany, is acceptable (based on local knowledge).

1.2. Generation of Trajectory Data Sets

Figure 2. The generated trajectory of Odysseus with labeled events and directional arrows (based on a CSV file).

In *Figure 2*, (prompt: “create a csv file with the trajectory of Odysseus with record number, longitude, latitude, time stamp, year, name of location, name of event, time duration of Odysseus’ stay at the specific location”). In this case, selected events from the novel are well represented, but other attribute data - especially the timestamps - are inappropriate.

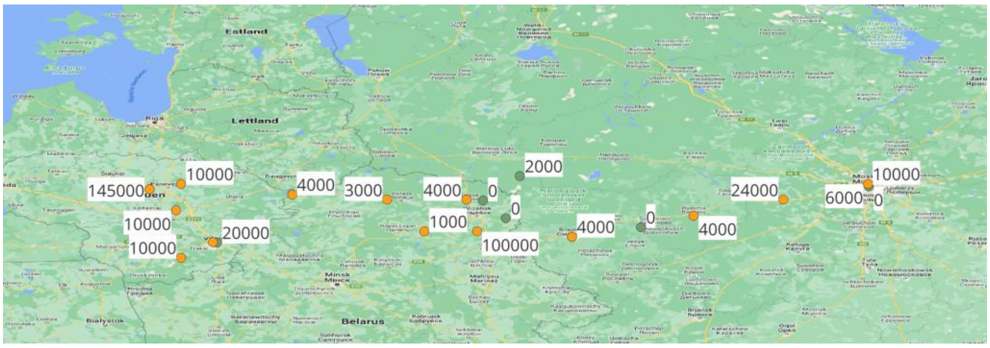

Figure 3. Recreating the data of Minard's map for Napoleon's Russian campaign (in orange) with labels showing the number of troop losses after every battle (compare to the data in green for Napoleon's Russian campaign in 1812).

In *Figure 3*, we compared the data in Minard's map with the losses of Napoleon's army (prompt: "create a csv file Minards map"). The locations of the battles partially match and, surprisingly, so do selected records from the historical map. On the other hand, the data on Napoleon's campaign of 1812 (in green in Figure 3) show far less detailed information.

1.3. Generation of Lines and Polygons

Figure 4. Generation of a line segment (in Magenta) representing the length of the reflection lake in Washington D.C..

Figure 4 shows surprising information: the reflection lake is shorter and not in the right place, but relatively close to the location of the reflection lake in Washington D.C. (prompt: “create a GeoJSON file with a line segment between the start and end of the reflection lake in Washington DC”).

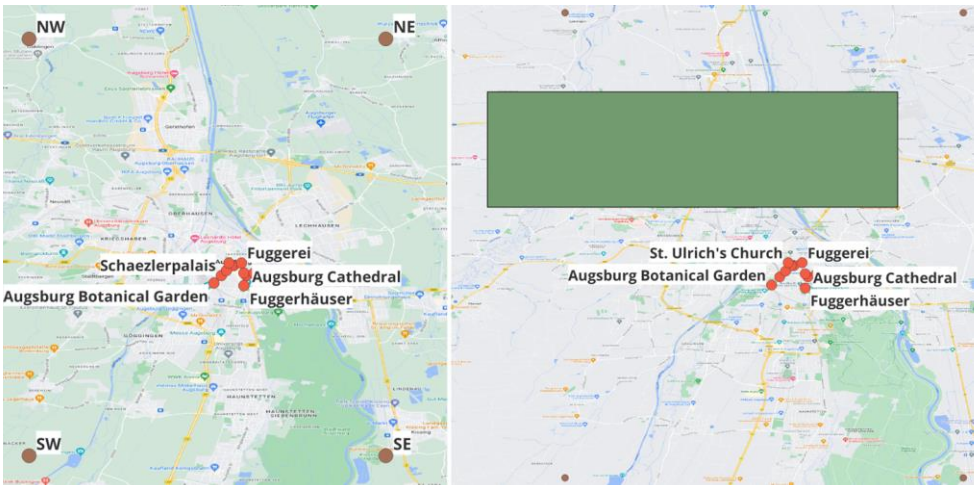

Figure 5. Examples for a bounding box around the German city of Augsburg with 4 cornerpoints (based on a CSV file) (left) and another attempt with a polygon (based on a GeoJSON file) (right).

Figure 5 shows the two tested approaches to polygon generation, resulting in a CSV file (left) and a GeoJSON file (right) with different dimensions and shapes (prompt: "create a csv file with four corner points of a bounding box around Augsburg"; "Create a GeoJSON file with a bounding box around Augsburg"). Nevertheless, there is a close proximity to the real city boundaries of Augsburg, Germany.

Geodata Enrichment through ChatGPT?

The development in this field is highly dynamic. It seems impossible to predict how the development will continue. Still, we can conclude that it is possible to generate geospatial data with the ChatGPT AI algorithms. As expected, these yield different qualities. Rather "fuzzy" concepts like "top 10 sights' seem to be generated in a reasonable list, geocoding seems reasonable on a large scale (e.g. the market place in Augsburg is not exactly on the right coordinates, but "approximately" in the right area. Nevertheless, the exact coordinates are not correct. Features that are measured in space (e.g. reflections of lakes or country borders) are created with estimated coordinates. These features are not geocoded correctly.

The idea of enriching geospatial data with generative AI differs from traditional approaches by selectively adding or extending attribute information for specific features. Jang et al. (2023) propose enriching geospatial data with place identities that can incorporate local knowledge, including dynamic components such as event information. This approach can benefit voluntary mapper initiatives such as the one presented by Polous et al. (2015). Our focus is on enriching point data on landmarks in Augsburg and establishing relationships between landmarks and POIs in OpenStreetMap (OSM). The formalization of POIs (Psyllidis et al. 2022) could benefit from AI-generated information enrichment. We are also exploring web-scraping approaches (Brenning & Henn 2022) to enrich scraped data with additional information during the scraping process.

Another avenue worth considering is the use of linked open data and semantic web approaches to generate and enrich geospatial data. Freely accessible databases such as Geonames (Ahlers 2013) offer analysis possibilities for location and street information.

References

Ahlers D (2013) Assessment of the accuracy of GeoNames gazetteer data. In Proceedings of the 7th workshop on geographic information retrieval (pp. 74-81)

Brenning A, Henn S (2023) Web scraping: a promising tool for geographic data acquisition. arXiv preprint arXiv:2305.19893

Jang KM, Chen J, Kang Y, Kim J, Lee J, Duarte F (2023) Understanding Place Identity with Generative AI. arXiv preprint arXiv:2306.04662

Polous K, Krisp JM, Meng L, Shrestha B, Xiao J (2015) OpenEventMap: A volunteered location-based service. Cartographica: the International Journal for Geographic Information and Geovisualization 50(4):248-258

Psyllidis A, Gao S, Hu Y, Kim EK, McKenzie G, Purves R, Yuan M, Andris C (2022) Points of Interest (POI): a commentary on the state of the art, challenges, and prospects for the future. Computational Urban Science 2(1):20

Taecharungroj V (2023) "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. Big Data and Cognitive Computing 7(1):35