Assessing Performance in Extracting Topological, Direction and Distance Spatial Relations from Reddit using LLMs

Paddy Smith[1,*], Ed Manley[1] and Myles Gould[1]

[1]Institute for Spatial Data Science, School of Geography, University of Leeds, UK

Abstract

This paper provides an initial exploration of the capabilities of large language models (LLMs) to extract spatial relations from unstructured social media text. The approach examines the performance of GPT-4o and Gemini 1.5-Pro using a diverse set of spatial relation terms, and seeks to determine whether certain spatial relation types (topological, distance and direction) are more challenging to extract. To evaluate, GPT-4o and Gemini 1.5-Pro output is compared to manually labeled spatial relation triplets from Reddit place descriptions. The findings demonstrate challenges in extracting spatial relations for LLMs, with the highest model only achieving 0.48 precision. However, performance varied across spatial relation types, as direction relations were extracted with higher precision (0.75) compared to distance relations (0.62) and topological relations (0.35).

Keywords

Spatial relation extraction, Large Language Models (LLMs), Social media, Reddit

1. Introduction

In recent years, there has been renewed interest in improving extraction of geographic information from unstructured text, such as social media or user-generated sources. This has been driven partly by recent restrictions on social media APIs (e.g., X/Twitter) [1], but also the potential of large language models (LLMs) in extracting geographic information from unstructured text [2].

Reddit has attracted attention in the geospatial domain, due to its place-based communities (i.e. r/London, r/Melbourne) and user-generated discussions which can contain rich descriptions of place. As a result, Reddit provides information on how people perceive and describe places in everyday language, which are often absent from geographic datasets. The availability of Reddit data has enabled research into toponym recognition methods for extracting place names [3, 4]. However, research on the extraction of spatial relations from Reddit, and other social media text, has been limited.

Spatial relations describe how objects or entities are positioned in space, relative to one another [5]. Generally, they are categorized as either topological, direction or distance, and represented in natural language with terms like 'in', 'above' or 'near' [6]. Methods for extracting spatial relations from natural language have developed in computer science and GIScience [7], and common methods include; dictionary-based, supervised machine learning and deep neural networks [8].

Extracting spatial relations from unstructured text, like Reddit, remains a challenge for established methods, for several reasons. Spatial relation triplets may not follow a simple syntactic structure of <subject, relation, object> (e.g. *"check out the nice parts of Bradford, such as Ilkley and Saltaire."*), terms used can be complex and contextual (e.g. *'about 10 minute drive'*), and applied presuming spatial knowledge (e.g. *"Camden is just past King's Cross."*) [9]. LLMs may be able to tackle these challenges. For instance, identifying informal spatial relation terms (abbreviations, slang and creative spelling) via their vast training corpus, recognize ambiguously used spatial relations due to greater contextual semantic understanding or infer implicit references using spatial reasoning.

GeoExT 2025: Third International Workshop on Geographic Information Extraction from Texts at ECIR 2025, April 10, 2025, Lucca, Italy

*Corresponding author.
gy17pls@leeds.ac.uk (P. Smith); e.j.manley@leeds.ac.uk (E. Manley); m.i.gould@leeds.ac.uk (M. Gould)
0009-0006-6955-1144 (P. Smith); 0000-0002-8904-0513 (E. Manley); 0000-0002-7104-0312 (M. Gould)

This study is an initial exploration into the capabilities of using LLMs to extract explicit spatial relation triplets from a manually annotated Reddit text corpus. Preliminary experiments test LLM's ability to identify spatial relation terms across general types and sub-types. By focusing on explicit spatial relations, LLMs semantic interpretation of text is explored, not its ability to infer implicit relations through reasoning.

2. Related Work

LLMs have already shown impressive capabilities in information extraction tasks, including NER [10] and relation extraction [11], and also toponym recognition and resolution [12, 13]. Initial attempts have been made towards assessing LLMs capabilities in spatial relation tasks. A subset of this work has explored LLMs spatial reasoning ability to infer spatial relations [14, 15]. A few papers have also evaluated LLMs ability to extract explicit spatial relations from text. Ramrakhiyani et al. [16] combines natural language inference (NLI) with LLMs to extract border orientations between countries within Wikipedia text data, achieving high precision results. Hu et al. [17] uses GPT-4 to extract a range of spatial relations from Wikipedia text, however performance is not evaluated. Haris et al. [18] tests three separate LLMs (GPT, Llama and Gemini) to extract a selection of spatial relation terms from the Corpus of the Lake District Writings (CLDW). The results showed high performance for GPT-4, however the models produced inaccuracies, including extracting spatial relations that did not exist in the text.

There are several directions this preliminary work can be extended. So far, there have been limited attempts to extract spatial relations from unstructured text [18], and from social media data. Furthermore, no study has compared LLM's performance in extracting spatial relation terms across spatial relation types (topological, distance and direction). Finally, the use of prompts has been relatively unexplored, as previous work has only used zero-shot prompting.

3. Approach

3.1. Test set

Reddit data collected by Berragan et al. [19] was used in this study. The dataset consists of Reddit posts and comments from 186 UK place-based subreddits, posted between 2011 to 2022. To evaluate LLMs performance, text from the subreddit r/Leeds were manually annotated to create the test set. Triplets were labelled if they contain a subject (a reference to a place to be located), a spatial relation term and an object (a reference to a place already located), in the form <subject, relation, object>. A place was defined as any named location (e.g., 'Leeds', 'Headingley'), or generic place descriptor (e.g., 'home', 'city center').

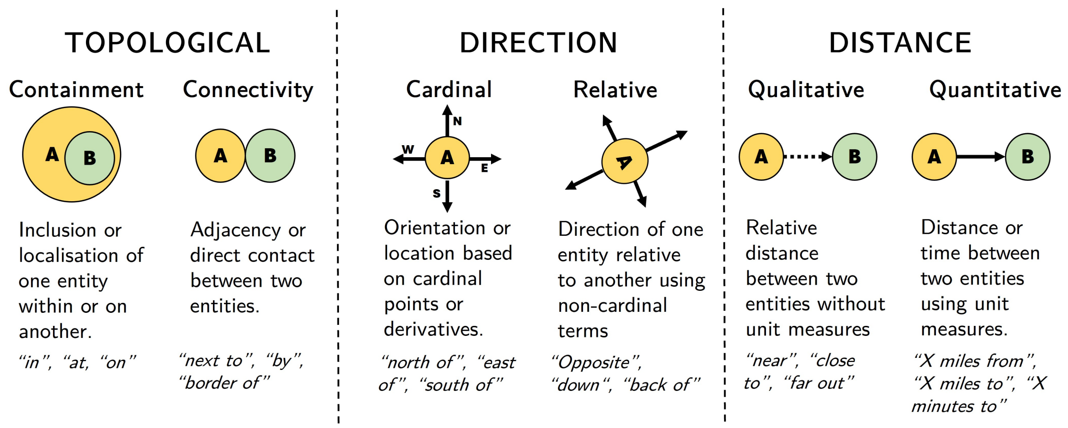

Figure 1: General type and sub-type spatial relations with definition and common natural language terms

Each triplet within the test set was assigned to spatial relation general-type (topological, distance or direction) and six sub-types (see Figure 1), which have been defined in [20]. Only containment and connectivity relations were chosen for topological spatial relations, as terms relating to other topological relations were not frequently present in the Reddit text corpus. Natural language terms were assigned to a general type and sub-type, aided by previous work [21, 22].

In total, 300 spatial relation triplets were labeled from the Reddit text corpus. A comment or post may contain multiple spatial relation triplets. It was ensured that equal amounts of triplets for each spatial relation general type and sub-type were present within the test set. An additional 600 comments or posts that did not contain spatial relation triplets (non-spatial) were included in the test set to better reflect the sparsity of such relations in the Reddit text corpus. This also enabled the evaluation of false positives, allowing assessment of whether LLMs incorrectly extracted spatial relationships that were not present in the text.

3.2. Prompt-engineering

Prompts were designed to include definitions for spatial relation sub-types, and to constrain the LLM to only extract explicit spatial relations (see Appendix A). Additionally, place names were not provided as an input, therefore the model is required to extract them. The base prompt is used for the zero-shot approach, which does not provide any training data. It is then adapted for two different prompt approaches (Appendix B). Few-shot prompting uses examples, and chain-of-thought provides reasoning steps to improve model performance [23, 24].

To extract the spatial relation triplets, GPT-4o and Gemini 1.5-Pro was used via OpenAI [25] and Google AI [26] API services. The outputs were compared to the test set, and performance is evaluated using precision, recall and F1-score. A correct spatial relation triplet includes exact place names and spatial relation term found in the text, in the correct order and with the correct general type and sub-type labels. Performance scores for each spatial relation general-type and sub-type were calculated as an average across the three prompt approaches.

Table 1
Evaluation results for GPT-4o and Gemini 1.5-Pro. Precision (P), recall (R) and F1 scores. Bold numbers indicate best score.

	Sample			GPT-4o			Gemini 1.5-Pro		
	Triplets	Non-spatial	Total	P	R	F1	P	R	F1
Zero-shot	300	600	900	0.48	0.52	0.50	0.38	0.54	0.45
Few-shot	300	600	900	0.45	0.68	0.54	0.42	0.83	0.56
Chain of Thought	300	600	900	**0.48**	0.63	0.55	0.42	**0.83**	**0.56**
Topological	100	600	700	0.35	0.75	0.48	0.35	**0.92**	0.50
Containment	50	600	650	0.27	0.79	0.40	0.27	**0.97**	0.42
Connectivity	50	600	650	0.79	0.68	**0.73**	0.62	0.86	0.71
Direction	100	600	700	**0.75**	0.53	**0.62**	0.49	0.70	0.56
Relative Direction	50	600	650	0.66	0.54	0.59	0.39	0.75	0.50
Cardinal Direction	50	600	650	**0.89**	0.53	0.66	0.79	0.64	0.68
Distance	100	600	700	0.62	0.52	0.56	0.54	0.53	0.53
Quantitative Distance	50	600	650	0.71	0.45	0.54	0.51	0.34	0.41
Qualitative Distance	50	600	650	0.57	0.59	0.58	0.55	0.71	0.61

4. Experiments and Evaluation

The results of the experiment are presented in Table 1. Overall, GPT-4o and Gemini 1.5-Pro exhibited moderate performance scores, with lower than expected precision. Few-shot and chain-of-thought prompting appeared to improve performance across the models with higher recall scores. Precision and recall scores varied by spatial relation type considerably, as certain spatial relations were more challenging to extract (Figure 2).

Direction relations were less challenging compared to topological and distance relations. In particular, cardinal relations were extracted with high precision (example *a*). Likewise, topological connectivity relations were easily identified and captured by the models (example *b*).

Most spatial relation sub-types had lower recall compared to precision, therefore the models was confident in predictions but missed relevant cases. In particular, quantitative distance relations proved challenging, as both models frequently missed numerical spatial relation terms (example *c*). In contrast, relative direction and qualitative distance had lower precision scores, as the terms often rely on spatial references and interpretation, making them more prone to inaccuracies (example *d*).

Finally, containment relation proved the most challenging for GPT-4o and Gemini 1.5-Pro. Contrasting to other spatial relations, it produced very low precision scores. Examining the output, there was a high frequency of implied 'in' relations. Some of these were geographically correct however did not explicitly appear in the text, whilst others were presumed due to the sentence structure (example *e*). Similar findings have been reported in previous studies [18], and have been attributed to LLM's tendency to hallucinate information that appears plausible but is not grounded in the input text.

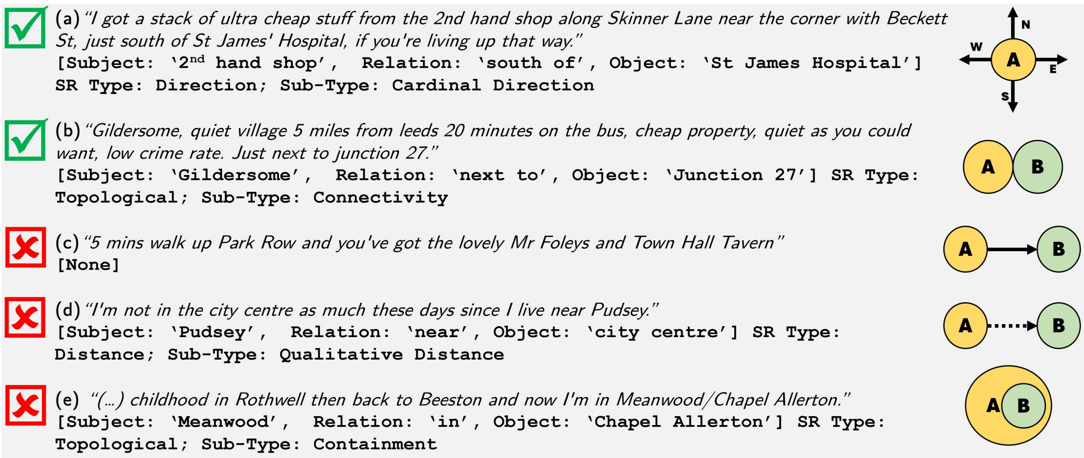

Figure 2: Example GPT-4o output with spatial relation sub-type. (✓) denotes successes and (✗) challenges.

5. Conclusion

This paper presents initial experiments using LLMs to extract explicit spatial relations from Reddit. Both GPT-4o and Gemini 1.5-Pro encountered significant challenges in this task, exhibiting numerous inaccuracies and missed cases. However, ability to extract spatial relations varied across different types, warranting further investigation. One possible explanation is the presence of biases in training datasets. For instance, topological and directional relations, such as *"next to"* or *"north of"*, may be overrepresented, leading to better generalization for these types of relations. Additionally, understanding the extent to which LLMs rely on memorization versus reasoning remains an open question [27]. This is particularly relevant for Reddit-based data, given its reported inclusion in LLM training corpora [28].

Future research should explore techniques to enhance LLM performance, including the integration of geographic knowledge [17]. For topological relations, resources such as gazetteers and spatial ontologies could validate containment relationships. Retrieval-augmented generation (RAG) approaches may help mitigate hallucinations by incorporating external knowledge sources [29].

Although the findings are specific to the manually annotated dataset, further investigation using benchmark datasets is needed. Nevertheless, this experiment establishes an initial foundation for future assessments of LLM's capabilities in spatial relation extraction.

Acknowledgments

This research has been funded by Economic and Social Research Council (ESRC) via the White Rose Doctoral Training Partnership (WRDTP).

Declaration on Generative AI

The authors used ChatGPT to rephrase sentences to improve clarity and conciseness. Any suggestions were reviewed and edited as needed, and the authors take full responsibility for the publication's content.

References

[1] B. I. Davidson, D. Wischerath, D. Racek, D. A. Parry, E. Godwin, J. Hinds, D. Van Der Linden, J. F. Roscoe, L. Ayravainen, A. G. Cork, Platform-controlled social media APIs threaten open science, Nature Human Behaviour 7 (2023) 2054–2057. doi:10.1038/s41562-023-01750-2.

[2] H. Kim, S. Lee, POI GPT: Extracting POI Information from Social Media Text Data, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-4/W10-2024 (2024) 113–118. doi:10.5194/isprs-archives-XLVIII-4-W10-2024-113-2024.

[3] C. Berragan, A. Singleton, A. Calafiore, J. Morley, Transformer based named entity recognition for place name extraction from unstructured text, International Journal of Geographical Information Science 37 (2023) 747–766. doi:10.1080/13658816.2022.2133125.

[4] M. Stillman, A. Kruspe, Geolocation Extraction From Reddit Text Data, in: GeoExT 2024: Second International Workshop on Geographic Information Extraction from Texts at ECIR 2024, Glasgow Scotland, UK, 2024.

[5] K. Stock, C. B. Jones, S. Russell, M. Radke, P. Das, N. Aflaki, Detecting geospatial location descriptions in natural language text, International Journal of Geographical Information Science 36 (2022) 547–584. doi:10.1080/13658816.2021.1987441.

[6] C. Loglisci, D. Ienco, M. Roche, M. Teisseire, D. Malerba, Toward Geographic Information Harvesting: Extraction of Spatial Relational Facts from Web Documents, in: 2012 IEEE 12th International Conference on Data Mining Workshops, IEEE, Brussels, Belgium, 2012, pp. 789–796. doi:10.1109/ICDMW.2012.20.

[7] Q. Qiu, Z. Xie, K. Ma, Z. Chen, L. Tao, Spatially oriented convolutional neural network for spatial relation extraction from natural language texts, Transactions in GIS 26 (2022) 839–866. doi:10.1111/tgis.12887.

[8] H. M. Rawsthorne, N. Abadie, E. Kergosien, C. Duchêne, É. Saux, Automatic Nested Spatial Entity and Spatial Relation Extraction From Text for Knowledge Graph Creation: A Baseline Approach and a Benchmark Dataset, in: Proceedings of the 7th ACM SIGSPATIAL International Workshop on Geospatial Humanities, ACM, Hamburg Germany, 2023, pp. 21–30. doi:10.1145/3615887.3627754.

[9] Q. Qiu, Z. Xie, K. Ma, L. Tao, S. Zheng, NeuroSPE : A neuro-net spatial relation extractor for natural language text fusing gazetteers and pretrained models, Transactions in GIS 27 (2023) 1526–1549. doi:10.1111/tgis.13086.

[10] S. Wang, X. Sun, X. Li, R. Ouyang, F. Wu, T. Zhang, J. Li, G. Wang, GPT-NER: Named Entity Recognition via Large Language Models, 2023. doi:10.48550/ARXIV.2304.10428.

[11] X. Xu, Y. Zhu, X. Wang, N. Zhang, How to Unleash the Power of Large Language Models for Few-shot Relation Extraction?, 2023. doi:10.48550/ARXIV.2305.01555.

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

[13] K. Kopanov, Comparative Performance of Advanced NLP Models and LLMs in Multilingual Geo-Entity Detection, in: Proceedings of the Cognitive Models and Artificial Intelligence Conference, ACM, İstanbul Turkiye, 2024, pp. 106–110. doi:10.1145/3660853.3660878.
[14] A. G. Cohn, R. E. Blackwell, Evaluating the Ability of Large Language Models to Reason about Cardinal Directions, 2024. doi:10.48550/ARXIV.2406.16528.
[15] N. Fulman, A. Memduhoğlu, A. Zipf, Distortions in Judged Spatial Relations in Large Language Models, The Professional Geographer (2024) 1–9. doi:10.1080/00330124.2024.2372792.
[16] N. Ramrakhiyani, V. Vasudeva, K. Girish, Extracting Orientation Relations between Geo-Political Entities from their Wikipedia Text, GeoExT@ ECI (2023) 44–50.
[17] L. Hu, W. Li, J. Xu, Y. Zhu, GeoEntity-type constrained knowledge graph embedding for predicting natural-language spatial relations, International Journal of Geographical Information Science (2024) 1–24. doi:10.1080/13658816.2024.2412731.
[18] E. Haris, A. G. Cohn, J. G. Stell, Exploring Spatial Representations in the Historical Lake District Texts with LLM-based Relation Extraction, 2024. doi:10.48550/ARXIV.2406.14336.
[19] C. Berragan, A. Singleton, A. Calafiore, J. Morley, Mapping cognitive place associations within the United Kingdom through online discussion on Reddit, Transactions of the Institute of British Geographers (2024) tran.12669. doi:10.1111/tran.12669.
[20] P. Kordjamshidi, M. Van Otterlo, M.-F. Moens, Spatial Role Labeling Annotation Scheme, in: N. Ide, J. Pustejovsky (Eds.), Handbook of Linguistic Annotation, Springer Netherlands, Dordrecht, 2017, pp. 1025–1052. doi:10.1007/978-94-024-0881-2_38.
[21] J. Chen, A. G. Cohn, D. Liu, S. Wang, J. Ouyang, Q. Yu, A survey of qualitative spatial representations, The Knowledge Engineering Review 30 (2015) 106–136. doi:10.1017/S0269888913000350.
[22] M. Egenhofer, D. M. Mark, R. Shariff, Natural-Language Spatial Relations Between Linear and Areal Objects: The Topology and Metric of English-Language Terms, International Journal of Geographical Information Science 12 (1998) 215–246.
[23] S. Min, X. Lyu, A. Holtzman, M. Artetxe, M. Lewis, H. Hajishirzi, L. Zettlemoyer, Rethinking the Role of Demonstrations: What Makes In-Context Learning Work?, 2022. doi:10.48550/ARXIV.2202.12837.
[24] A. Gao, Prompt Engineering for Large Language Models, SSRN Electronic Journal (2023). doi:10.2139/ssrn.4504303.
[25] OpenAI, Models: Gpt-4o (128k context window), 2023. URL: https://platform.openai.com/docs/models#gpt-4o, large language model.
[26] Google, Gemini 1.5 pro: A multimodal mixture-of-experts model with extended context window, 2024. URL: https://ai.google.dev/gemini-api/docs/models#gemini-1.5-pro, large language model.
[27] J. Roberts, T. Lüddecke, S. Das, K. Han, S. Albanie, GPT4GEO: How a Language Model Sees the World's Geography, 2023. URL: https://arxiv.org/abs/2306.00020. doi:10.48550/ARXIV.2306.00020.
[28] H. Field, Reddit soars after announcing OpenAI deal that allows use of its data for training AI models (2023). URL: https://www.cnbc.com/2024/05/16/reddit-soars-after-announcing-openai-deal-on-ai-training-models.html#:~:text=Reddit%20soars%20after%20announcing%20OpenAI%20deal%20that,previously%20announced%20a%20similar%20deal%20with%20Google.
[29] W. Fan, Y. Ding, L. Ning, S. Wang, H. Li, D. Yin, T.-S. Chua, Q. Li, A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models, in: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, ACM, 2024, pp. 6491–6501. URL: https://dl.acm.org/doi/10.1145/3637528.3671470. doi:10.1145/3637528.3671470.

A. Base Prompt Used

```
prompt_template = """
Given the Reddit comments, extract spatial relation triplets, when a place is found to be in a spatial relation with another place. The output format is:
[{{
        "subject": "<place name>",
        "relation": "<spatial relation>",
        "object": "<place name>",
        "general_type": "<Topological, Distance, or Direction>",
        "sub_type": "<Containment, Connectivity, Relative Direction, Cardinal Direction, Quantitative Distance, Qualitative Distance>"
    }}]
The subject is the reference to a place to be located, the relation is the term describing the spatial relation, and the object is the reference to a place
already located. Both subject and object must be a name of a place. Do not fabricate any place names or spatial relations. The place name and spatial
relation term must be explicitly found in the text. A comment may have zero, one, or multiple triplets to be extracted.

**Strict Rules for Extraction:**
- **- Both **subject** and **object** must be the name of a place.
- **Do NOT extract triplets where either subject or object is a pronoun or non-geographic entity** (e.g., "one", "I", "it", "people").
- **The spatial relation term must be explicitly found in the text. Do not infer any relations.**
-   **Do NOT assume that places in a list or separated by commas share a spatial relation.**
-   If a sentence contains **two place names without a connecting spatial relation term, do not extract a triplet**.

If a spatial relation term is identified, classify it as follows:

1) **Topological**
   - Containment: Inclusion or localization of one entity within or on another (e.g., "in", "at", "on").
   - Connectivity: Adjacency or direct contact between two entities without overlap (e.g., "next to", "beside", "by").

2) **Direction**
   - Relative Direction: Orientation or direction using non-cardinal terms (e.g., "above", "below", "up", "down").
   - Cardinal Direction: Location based on cardinal points or their derivatives (e.g., "north", "south", "southeast").

3) **Distance**
   - Quantitative Distance: Measurable distance or time between two entities using numerical or metric terms (e.g., "5 meters away", "10 minutes to").
   - Qualitative Distance: Perceived or relative distance without precise measurements (e.g., "close to", "far away").

Use these classifications to identify the correct **general type** and **sub-type**. These categories are for classification purposes only and should not
be used as spatial relation terms themselves.

Text: "{comment}"
"""
```

Figure 3: Base prompt used for zero-shot setting.

B. Prompt Approaches

Given the Reddit comments, extract spatial relation triplet, when a place is found to be in a spatial relation with another place. The output format is:

{{ "subject": "<place name>",
"relation": "<spatial relation>",
"object": "<place name>",
"general_type": "<Topological, Distance, or Direction>",
"sub_type": "<Containment, Connectivity, Relative Direction, Cardinal Direction, Quantitative Distance, Qualitative Distance>"}}

[...]

Here are a few examples:
Comment 1: "I love living near Gateshead, it's great and in Newcastle!."
Output (topological-containment relation): [{{"subject": "Gateshead", "relation": "in", "object": "Newcastle"}}]

(more examples follow..)

(a)

Given the Reddit comments, extract spatial relation triplet, when a place is found to be in a spatial relation with another place. The output format is:

{{ "subject": "<place name>",
"relation": "<spatial relation>",
"object": "<place name>",
"general_type": "<Topological, Distance, or Direction>",
"sub_type": "<Containment, Connectivity, Relative Direction, Cardinal Direction, Quantitative Distance, Qualitative Distance>"}}

[...]

Here are a few examples:
Comment 1: "I love living near Gateshead, it's great and in Newcastle!.“
Reasoning: Gateshead and Newcastle are both places. 'in' is a spatial relation term and refers to Gateshead as the subject and Newcastle as the object.
Output (topological-containment relation): [{{"subject": "Gateshead", "relation": "in", "object": "Newcastle"}}]

(more examples follow..)

(b)

Figure 4: Examples of prompt approaches: (a) few-shot, (b) chain-of-thought.