Economic and Social Research Council

The National Science Foundation

Understanding

IMPRECISE SPACE AND TIME IN NARRATIVES

through qualitative representations, reasoning, and visualisation

Ignatius Ezeani (Lancaster) & John Stell (Leeds)

Spatial Narratives Project

Lancaster University

UNIVERSITY OF LEEDS

University of BRISTOL

MANCHESTER 1824 The University of Manchester

Stanford University

IUPUI

Funded by: ESRC and NSF; **Duration:** 2022 to 2025; **Reference:** ES/W003473/1 **Website:** https://spacetimenarratives.github.io/

The Project

- ESRC and NSF funding, 42 months from April 2022 sources contain multiple representations of place and geography:
  - explicit: e.g. placename – ‘Keswick’
  - implicit: ‘the town’, ‘a mountain’
  - relational: ‘Shortly after leaving Keswick we crossed a stream where we turned off the road and began to climb the hill.’
- To identify, visualise and analyse all these types using:
  - natural language processing (NLP)
  - geographical information science (GISc)
  - qualitative spatio-temporal representation and reasoning
  - visual analytics

Project Team

Ian Gregory (PI)
Lancaster University, UK

Paul Rayson
Lancaster University, UK

Ignatius Ezeani
Lancaster University, UK

Tony Cohn
Leeds University, UK

John Stell
Leeds University, UK

Tim Cole
University of Bristol, UK

Erik Steiner
Stanford University, UK

Jackie Olson
Stanford University, US

David Bodenhamer
Indiana University, US

Neil Devadasan
Indiana University, UK

Zephyr Frank
Stanford University, US

Erum Haris
Leeds University, UK

Joanna Taylor
Manchester University, UK

https://spacetimenarratives.github.io/

Corpora

- Lake District writings including travel literature, fiction, histories, letters, and diaries
- Large collection of Holocaust materials, including
  - survivor testimonies in the form of oral histories and
  - bureaucratic records kept of the events recalled by victims.
- Survivors' testimonies lack temporal and spatial exactness due to the time between the recalled events and the recorded testimonies.

Questions

- How do writers as a group shape the sense of place?
- Is the sense of place temporally and geographically continuous or discontinuous?

- How and when do tropes relate to a sense of place arise — and how and when do they change?

- What spatial and linear trajectories are embedded in these corpora?
- Do these trajectories group by experience, genre, or subject?

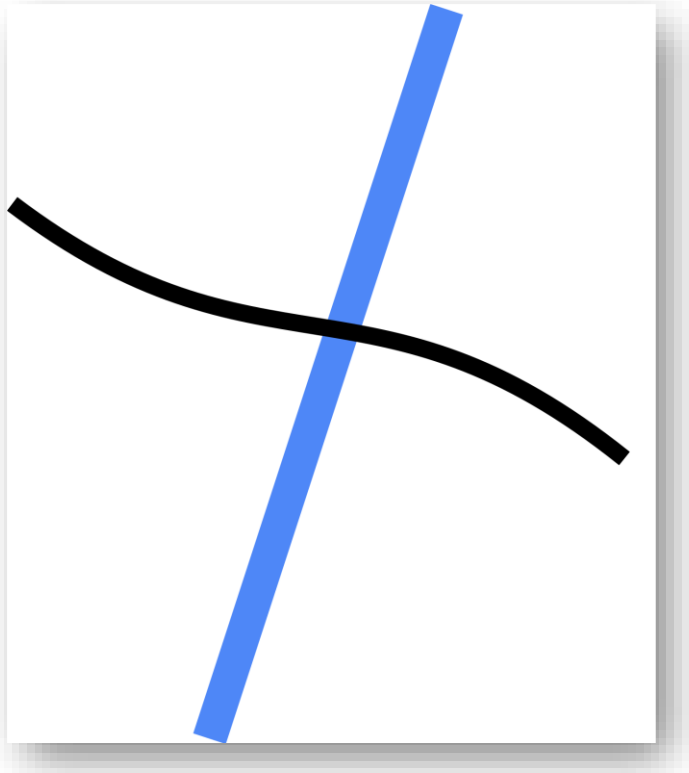
Spatial and Temporal Relationships in Text

- . . . this point, on the left, are the druidical remains called King Arthur's Round Table, and Mayborough. Immediately after crossing Pooley Bridge the road runs along the western shore of Ulleswater to Patterdale, a distance of ten miles; but, before. . .

Qualitative Spatial (and Temporal) Representation

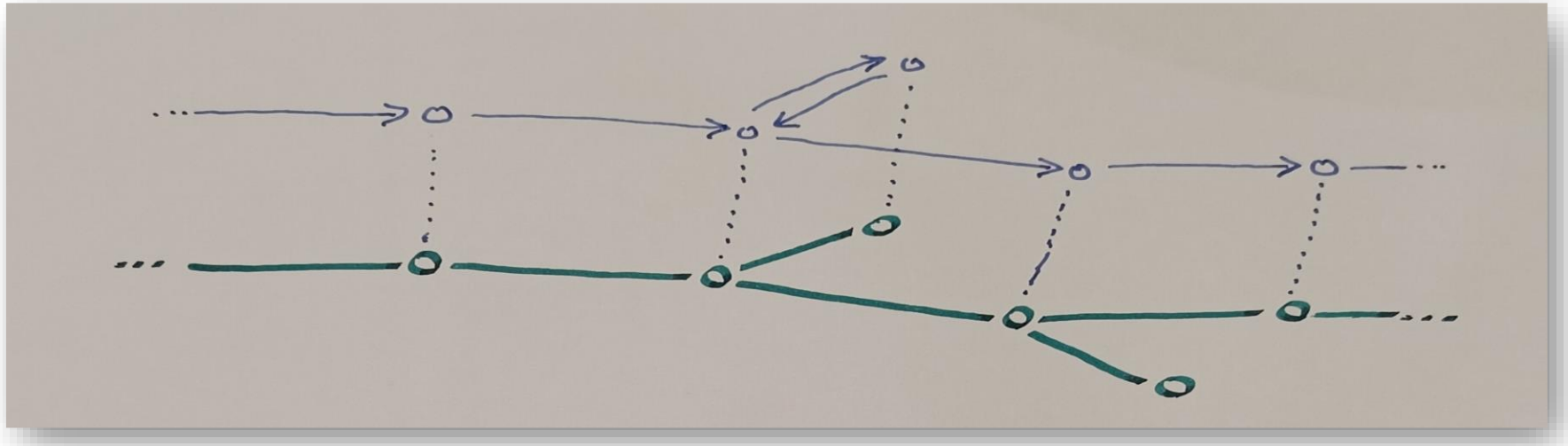
- QSR deals with relationships computationally.
- Space of human experience vs ideal mathematical abstraction.

- *and Reasoning*
  - **if** A and B are on opposite sides of the river
  - **then**: you have to cross the river to move from A to B

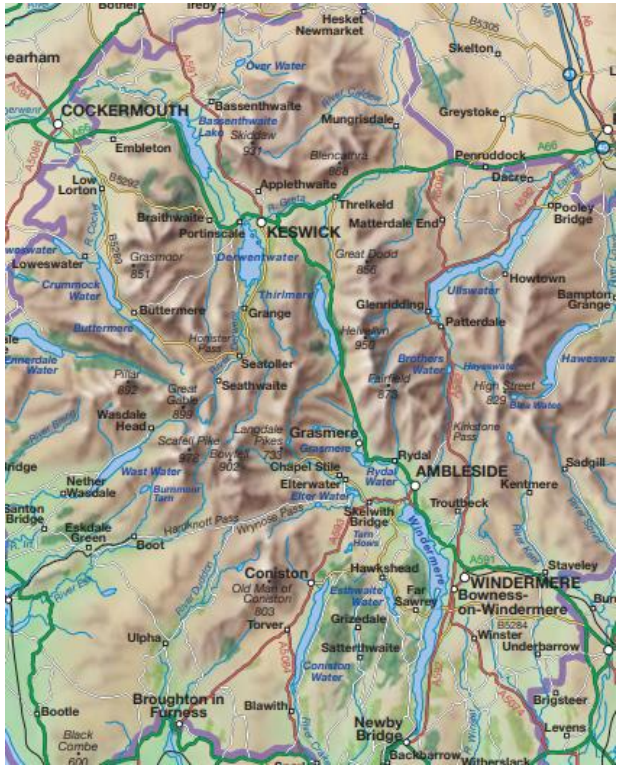


Qualitative Spatial (and Temporal) Representation

- Spatial relationships in text generate a graph
- Journeys are routes in this network
- Nodes can be tied to geographical locations
- Sense of place as pattern in network?



Spatial Entity Extraction



- Corpus of Lake District Writing (CLDW)
  - CLDW consists of 80 manually annotated texts
  - Comprises of over 1.5-million-word tokens
  - Texts originally composed between 1622 – 1900
  - Represent a range of different genres and authors
  - The gold standard subset is a representative sample of the corpus
    - 28 texts, 242k word tokens, one-sixth of the corpus
  - The gold standard marked up all place-name entities with the `<cdplace>` tag.

Project Aim

Given a piece of writing, can we build a narrative around its:

- Location:
  - Texts often use toponyms (place names) for locations often reduced to a coordinate for mapping.
- Locale:
  - Usually nouns for geographical features such as ‘house’, ‘lake’, ‘mountain’ etc. often vague or ambiguous and cannot be directly mapped in Euclidean space.
- Sense of place:
  - subjective and emotional expressions by a person that make the place a unique community or landscape.
  - Events that occur at a place will contribute to this.
  - Various forms of date, time, and other temporal references

Project Aim

Given a piece of writing, can we build a narrative around its:

- Location:
  - Texts often use **toponyms** (place names) for locations often reduced to a coordinate for mapping.
- Locale:
  - Usually nouns for **geographical features** such as ‘house’, ‘lake’, ‘mountain’ etc. often vague or ambiguous and cannot be directly mapped in Euclidean space.
- Sense of place:
  - subjective and **emotional** expressions by a **person** that make the place a unique community or landscape
  - **events** that occur at a place will contribute to this.
  - various forms of **date**, **time,** and other temporal references.

Objectives

Corpus annotation	Annotate the corpus to reflect the occurrence of spatial elements of interest: *toponyms, geo nouns, emotions, events, time*
Spatial entity extraction	Develop NLP techniques to appropriately identify, extract, and categorize these elements.
Spatial visualization	Add a user-friendly visualization interface to the extraction tool to look at the special entities
Analysis	Perform analysis on the extracted entities to build a narrative

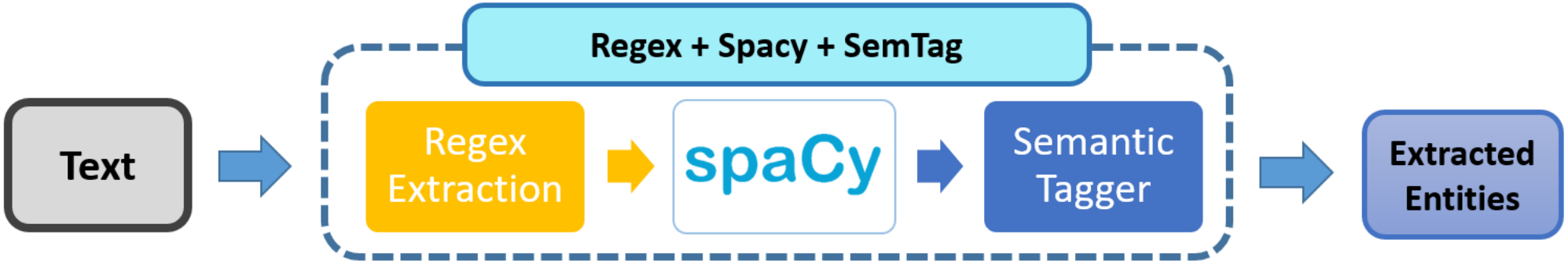
Corpus Annotation Schema and Tagset

- **PLNAME:** Place names
  - From gazetteer, annotated corpus, spacy NER
- **GEONOUN:** Geographical feature nouns
  - From 128 nouns manually created (river, road, waterfall, etc) + their inflections (*road* + *roads*)
- **EVENT:** Any description of an event or activity
  - e.g. delightful *excursions*. From manual annotations, PyMUSAS (movement)
- **EMOTION:** Sentiments and emotional expressions
  - *delightful* excursions, manual annotations, PyMUSAS (emotion)
- **DATE + TIME:** All date and time references
  - Spacy NER, manual annotation, PyMUSAS (Time)
- **PERSON:** Names and references to people
  - Spacy NER and manual annotation

*We also considering:

- **spatial prepositions** (*above*, *behind*, *below*, *through*, *within*)
- **locative adverbs** (*alongside*, *downtown*, *homewards*, *southward*, *up-country*)
- **sentiment classification** of sentences beyond identifying the emotion-bearing words

NLP pipeline



- Rule-based approach – Basic regular expression extracts placenames,
- No repetition or overlap
  - i.e. if an entity is found in by **Regex**, drop any **spaCy** entity within the span.
- Convert all **GPE**, **LOC**, **FAC** and **ORG** tags from **spaCy** to **PL-NAME**
- And **PERSON** tag too? Maybe not.

Corpus Annotation Tool

There is also here a singular floating island, which rises to the surface at intervals, and then sinks again. It varies in size, is covered with vegetation, and appeared last in 1842. It is supposed that its rise to the surface is caused by accumulation of gas formed by the decomposition of vegetable matter. The circuit of Keswick lake is a delightful excursion, affording many fine prospects of surpassing grandeur, wildness, and magnificence. The best views of it are to be had from Crow Park, Friar's Crag, and the Vicarage. The whole vale extends almost in a straight line from the head of Derwentwater to the foot of Bassenthwaite Lake, and is connected with Borrowdale on the south, and Thirlmere on the east, while on the west is Newlands Vale. Castlehill, one mile from Keswick on the Ambleside road, and Ashness, afford good bird's-eye views, while more distant views may be had from Latrigg, Ormathwaite, Applethwaite, and Braithwaite. Skiddaw's Cub is worth ascending.

LIGHT TAG

Has a free plan, with features: *annotate*, *review*, and *analyze* (IAA, Precision & Recall) and example classification. Also with pre-annotation suggestions. The schema includes specified tags in the project aims. Outputs JSON file

Gold Standard

Loughrigg Fell Should be ascended. It is not much above 1100 feet in height, and commands a fine view of all the lakes and mountains in its immediate vicinity. It is little more than a mile from the village.


```
{'value': 'Loughrigg Fell',
 'tag': 'PLNAME', 'start': 1, 'end': 15, 'reviewed': True, 'correct': True,
 'annotated_by':
    [{'annotator_id': 8, 'timestamp': '2023-02-09T15:48:00.266+00:00'},
     {'annotator_id': 1, 'timestamp': '2023-02-01T15:23:52.416+00:00'},
     {'annotator_id': 4, 'timestamp': '2023-02-06T12:56:58.315+00:00'},
     {'annotator_id': 5, 'timestamp': '2023-01-18T15:24:10.713+00:00'}],
}
---
{'value': 'view',
 'tag': 'GEONOUN', 'start': 98, 'end': 102, 'reviewed': False, 'correct': None,
 'annotated_by': [{'annotator': None, 'timestamp': None, 'annotator_id': None}]
}
---
{'value': 'Fell',
 'tag': 'GEONOUN', 'start': 11, 'end': 15, 'reviewed': True, 'correct': False,
 'annotated_by': [{'annotator': None, 'timestamp': None, 'annotator_id': None}],
}
```

Entity extraction

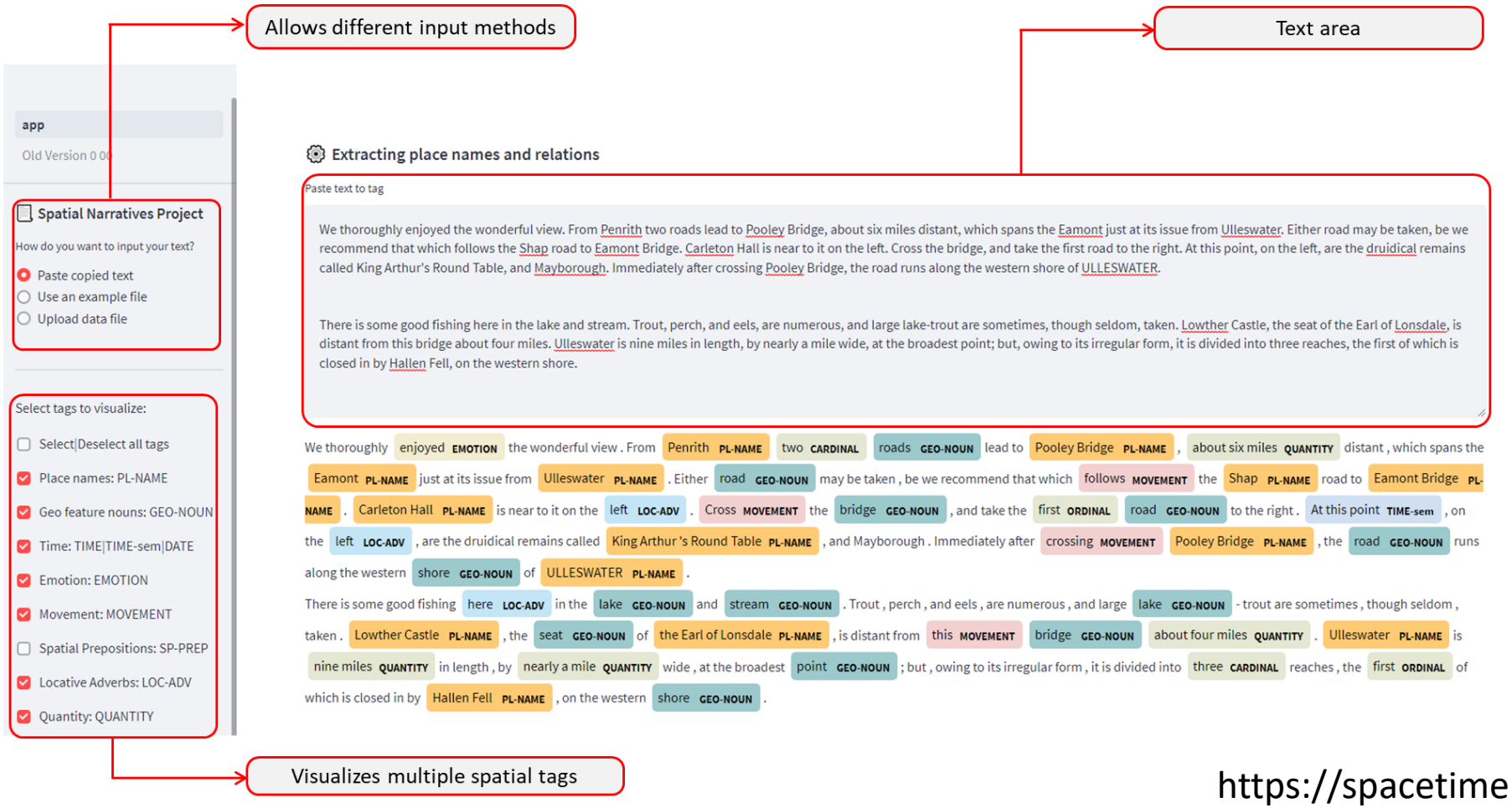
- **Task:** Extract from this text, *place names*, *geographical feature nouns*, *events*, *emotions*, *dates and times*, *people*

From Penrith two roads lead to Pooley Bridge, about six miles distant, which spans the Eamont just at its issue from Ulleswater. Either road may be taken, be we recommend that which follows the Shap road to Eamont Bridge. Carleton Hall is near to it on the left. Cross the bridge, and take the first road to the right. At this point, on the left, are the druidical remains called King Arthur's Round Table, and Mayborough. Immediately after crossing Pooley Bridge, the road runs along the western shore of ULLESWATER to Patterdale, a distance of ten miles; but, before proceeding along it, the tourist would do well to take a walk of a few miles along the eastern shore, in the direction of Martindale, from several points on which he will obtain a good view of the lake.

Entity extraction pipeline

From Penrith PL-NAME two CARDINAL roads GEO-FEATURE lead to Pooley Bridge, PL-NAME about six miles QUANTITY distant, which spans the Eamont PL-NAME just at its issue from Ulleswater. PL-NAME Either road GEO-FEATURE may be taken, be we recommend that which follows the Shap PL-NAME road GEO-FEATURE to Eamont Bridge. PL-NAME Carleton Hall PL-NAME is near to it on the left. Cross the bridge, GEO-FEATURE and take the first ORDINAL road GEO-FEATURE to the right. At this point, GEO-FEATURE on the left, are the druidical remains called King Arthur's Round Table PL-NAME , and Mayborough PL-NAME . Immediately after crossing Pooley Bridge, PL-NAME the road GEO-FEATURE runs along the western shore GEO-FEATURE of ULLESWATER PL-NAME to Patterdale, PL-NAME a distance of ten miles QUANTITY ; but, before proceeding along it, the tourist would do well to take a walk of a few miles QUANTITY along the eastern shore, GEO-FEATURE in the direction of Martindale, PL-NAME from several points GEO-FEATURE on which he will obtain a good view of the lake. GEO-FEATURE

Visualization



https://spacetimenarratives.streamlit.app/

Text2Story @ ECIR'23

Extracting Imprecise Geographical and Temporal References from Journey Narratives (demo)

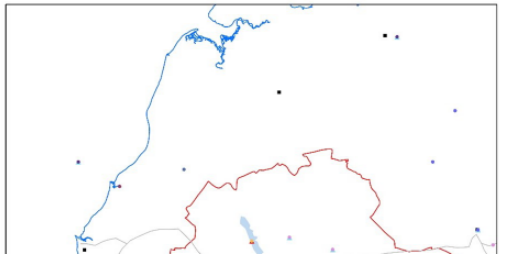
Ignatius Ezeani, Paul Rayson, Ian Gregory
{i.ezeani, p.rayson, i.gregory}@lancaster.ac.uk
Lancaster University, Lancaster, UK

Text2Story 2023:
6th Int'l Workshop on Narrative Extraction from Texts @ 45th Euro Conf. Information Retrieval, Apr 2–6, 2023 - Dublin, Ireland.

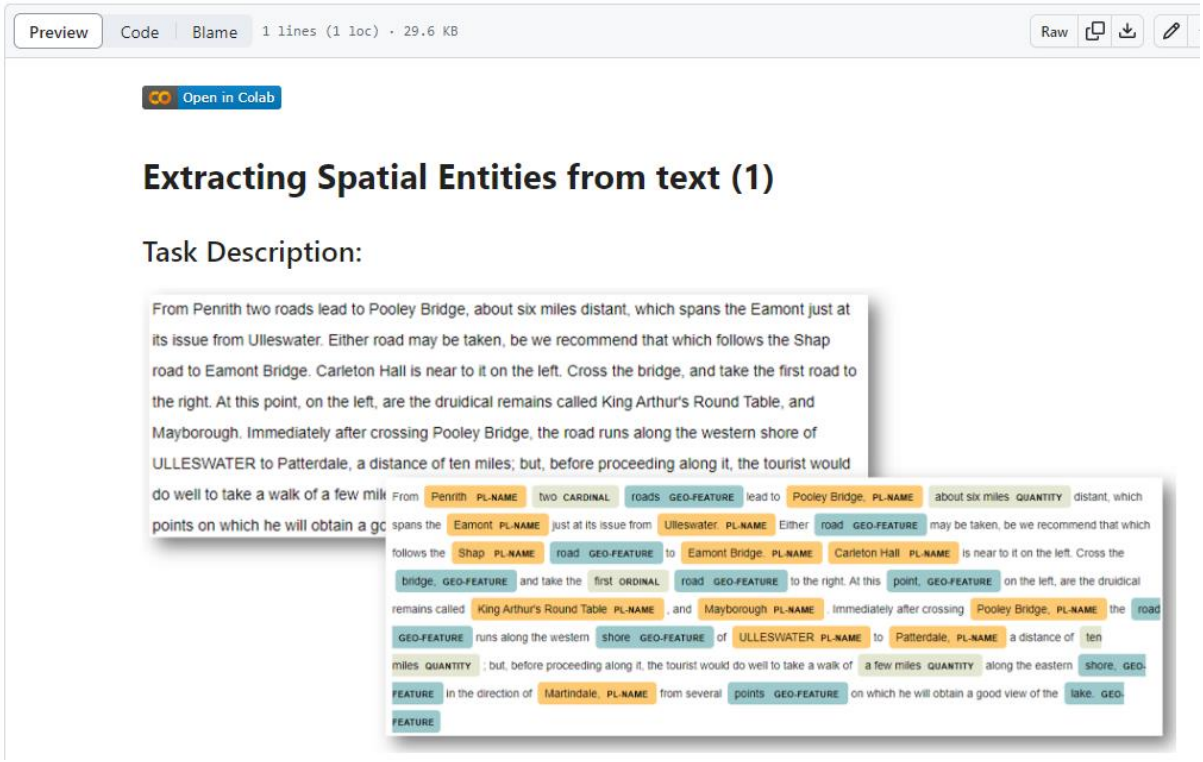
Project Overview

Approaches to understanding geographies in text are highly quantitative and are limited to named places for which coordinates can be found, with little concept of time. But human experiences of geography are often subjective and more suited to qualitative representation. "Geography" is not limited to named places; rather, it incorporates the vague, imprecise, and ambiguous references, for example, "*the camp*", or "*the hills in the distance*", and includes the relative locations using terms such as "*near to*", "*on the left*", "*north of*" or "*a few hours' journey from*". Here, we describe our research prototype to extract and analyse qualitative and quantitative references to place and time in two corpora of English Lake District travel writing and Holocaust survivor testimonies.

The Lake District Map



CL Pre-conference Workshop



GitHub: https://github.com/SpaceTimeNarratives/demo
Workshop website: https://wp.lancs.ac.uk/cl2023/pre-conference-workshops/

Economic and Social Research Council

The National Science Foundation

Thanks for your attention

Spatial Narratives Project

Lancaster University

University of Leeds

University of Bristol

Manchester 1824 The University of Manchester

Stanford University

IUPUI

Funded by: ESRC and NSF; **Duration:** 2022 to 2025; **Reference:** ES/W003473/1 **Website:** https://spacetimenarratives.github.io/