Economic and Social Research Council

The National Science Foundation

Lancaster University

UNIVERSITY OF LEEDS

University of BRISTOL

MANCHESTER 1824 The University of Manchester

Stanford University

IUPUI

Spatial Narratives Project

Understanding imprecise space and time in narratives through qualitative representations, reasoning, and visualisation

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

Project Team

Ian Gregory
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

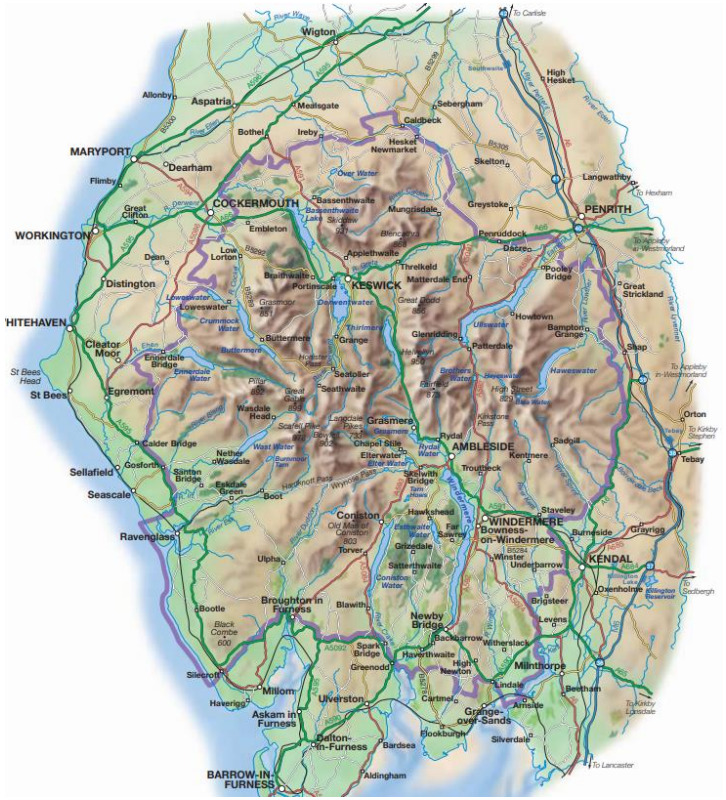
Zephyr Frank
Stanford University, US

Erum Haris
Leeds University, UK

Joanna Taylor
Manchester University, UK

https://spacetimenarratives.github.io/

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 usually use a toponym (place name) 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:
  - the subjective and emotional things that make the place a unique community or landscape. For our purposes,
  - the events that occur at a place will contribute to this.

Note: ‘Sense of place’ cannot exist without a place (either a location or a locale) to attach it to.

Key Objectives

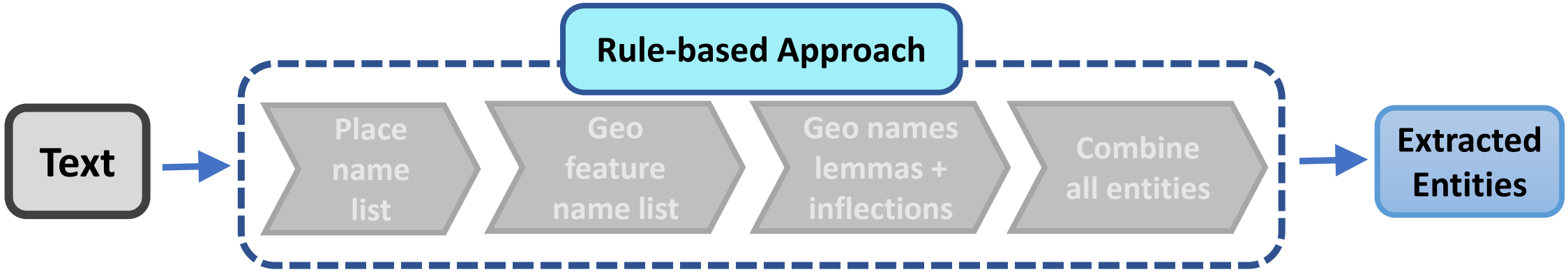
1. Create new and/or modified NLP techniques to identify and extract all quantitative and qualitative references to place and time?
   - toponyms (place names):
     - Penrith, Pooley Bridge, Eamont, Eamont Bridge
   - geographical feature nouns,
   - times and dates and
   - spatial and temporal relationships
2. Evaluate the performance of the techniques on place name identification in written texts?

Given this example text

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.

- We want to extract from the text
  - all place names
  - all geographical feature names (or nouns)
  - all times and dates
  - all relationships between places and times

NLP Pipeline: Rule-Based Approach

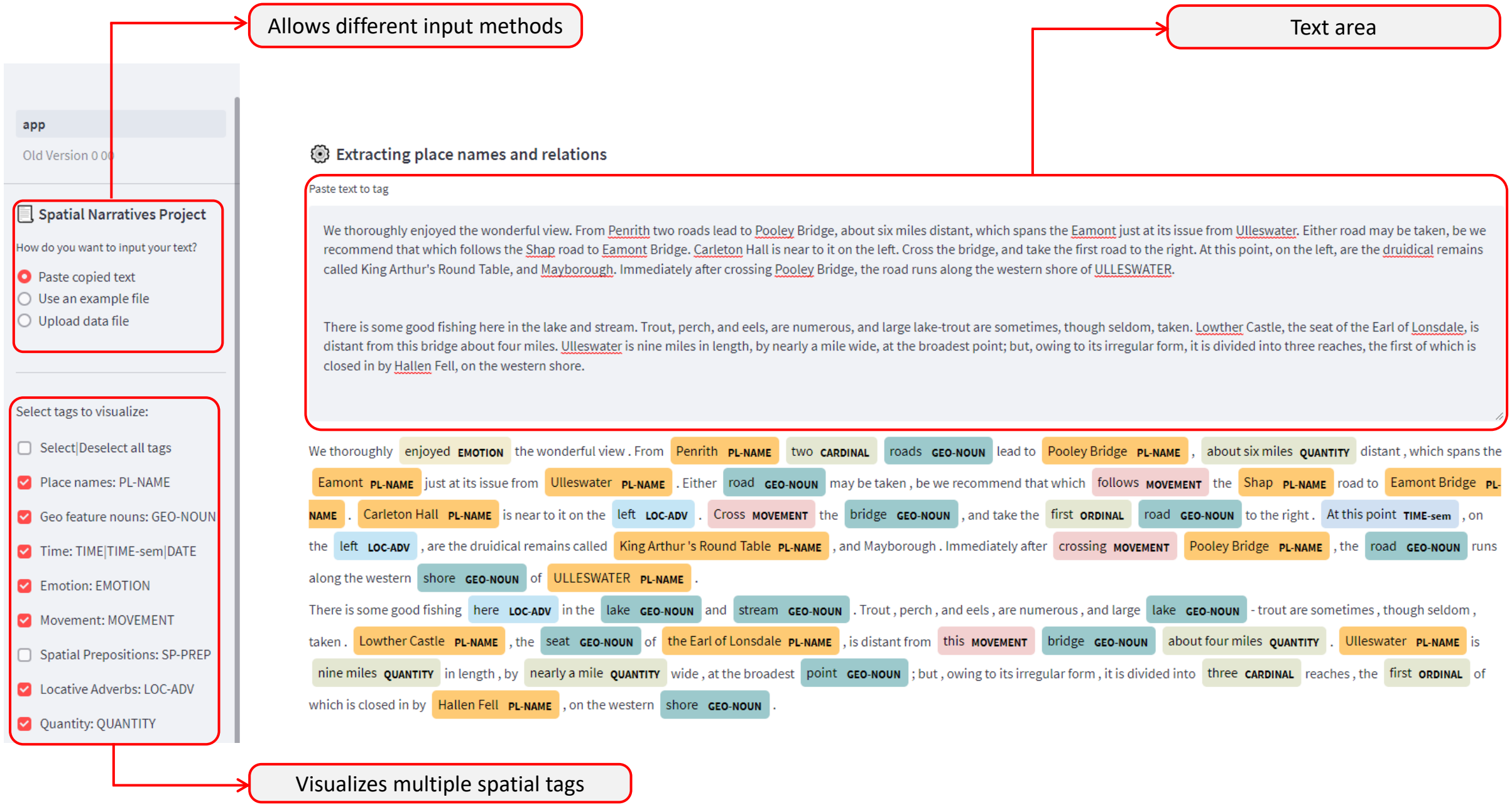


- Uses Regular expressions to extract ‘known’ place names and geographical feature names (nouns)
- Relies on existing list of entities (placenames, geo feature names etc)

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.

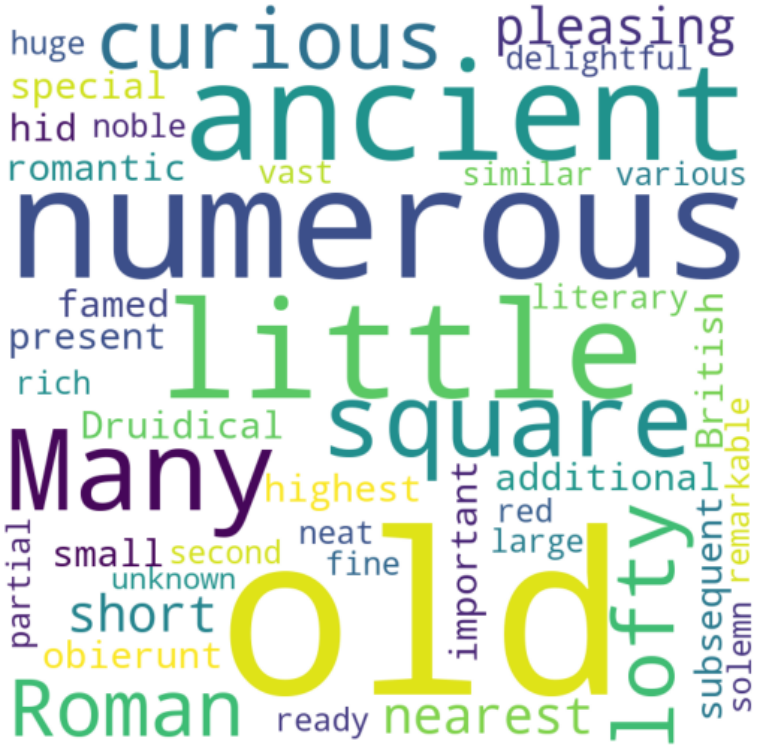
Completing the Pipeline

- We applied 2 Basic rules
- No repetition or over-lap
  - i.e. if entity is found in by the Rule-Based method, drop any **spaCy** or **PyMUSAS** entity within the span.
- Convert all **GPE**, **LOC**, **FAC** and **ORG** tags from **spaCy** to **PL-NAME**
  - And **PERSON** tag too? 🤔 Maybe not!


Allows different input methods
Text area
app
Old Version 0 00
Spatial Narratives Project
How do you want to input your text?
Paste copied text
Use an example file
Upload data file
Select tags to visualize:
Select|Deselect all tags
Place names: PL-NAME
Geo feature nouns: GEO-NOUN
Time: TIME|TIME-sem|DATE
Emotion: EMOTION
Movement: MOVEMENT
Spatial Prepositions: SP-PREP
Locative Adverbs: LOC-ADV
Quantity: QUANTITY
Extracting place names and relations
Paste text to tag
We thoroughly enjoyed the wonderful view. 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.
There is some good fishing here in the lake and stream. Trout, perch, and eels, are numerous, and large lake-trout are sometimes, though seldom, taken. Lowther Castle, the seat of the Earl of Lonsdale, is distant from this bridge about four miles. Ulleswater is nine miles in length, by nearly a mile wide, at the broadest point; but, owing to its irregular form, it is divided into three reaches, the first of which is closed in by Hallen Fell, on the western shore.
We thoroughly enjoyed EMOTION the wonderful view . From Penrith PL-NAME two CARDINAL roads GEO-NOUN 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-NOUN may be taken , be we recommend that which follows MOVEMENT the Shap PL-NAME road to Eamont Bridge PL-NAME . Carleton Hall PL-NAME is near to it on the left LOC-ADV . Cross MOVEMENT the bridge GEO-NOUN , and take the first ORDINAL road GEO-NOUN to the right . At this point TIME-sem , on the left LOC-ADV , are the druidical remains called King Arthur 's Round Table PL-NAME , and Mayborough . Immediately after crossing MOVEMENT Pooley Bridge PL-NAME , the road GEO-NOUN runs along the western shore GEO-NOUN of ULLESWATER PL-NAME .
There is some good fishing here LOC-ADV in the lake GEO-NOUN and stream GEO-NOUN . Trout , perch , and eels , are numerous , and large lake GEO-NOUN - trout are sometimes , though seldom , taken . Lowther Castle PL-NAME , the seat GEO-NOUN of the Earl of Lonsdale PL-NAME , is distant from this MOVEMENT bridge GEO-NOUN about four miles QUANTITY . Ulleswater PL-NAME is nine miles QUANTITY in length , by nearly a mile QUANTITY wide , at the broadest point GEO-NOUN ; but , owing to its irregular form , it is divided into three CARDINAL reaches , the first ORDINAL of which is closed in by Hallen Fell PL-NAME , on the western shore GEO-NOUN .
Visualizes multiple spatial tags

Sense-of-place: Co-occurring Adjectives

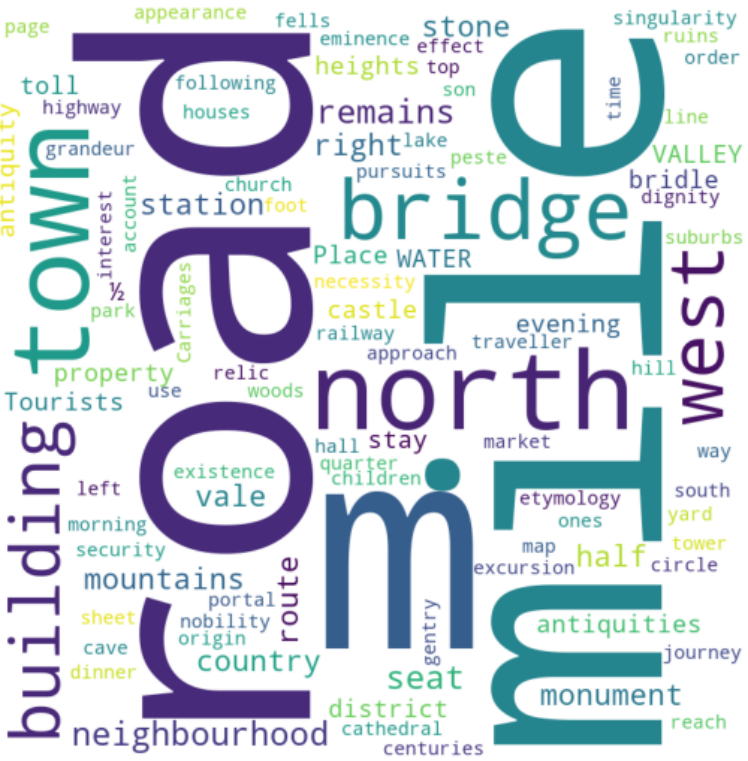
Penrith (Adjs)



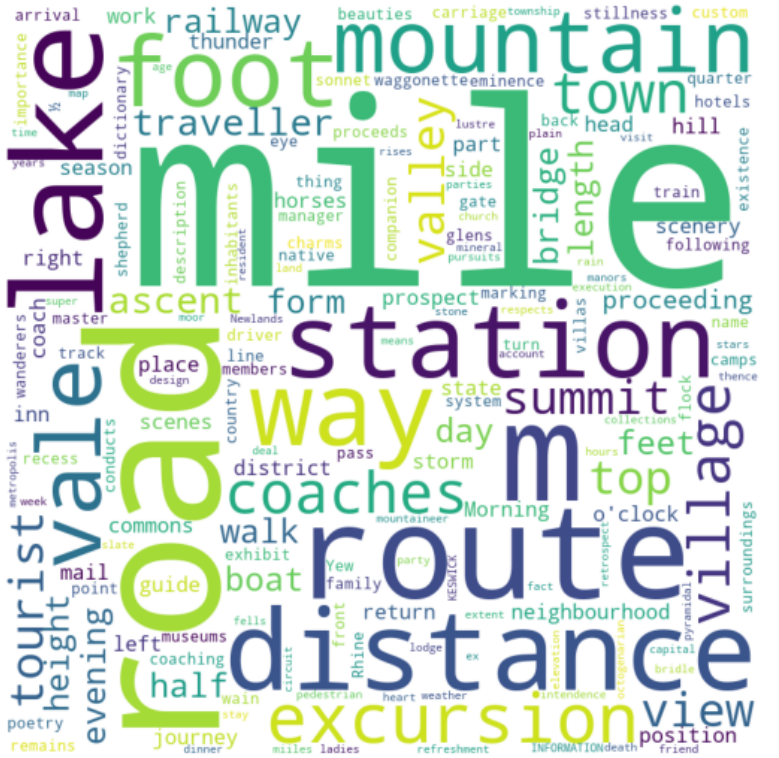
Keswick(Adjs)

Sense-of-place: Co-occurring Nouns

Penrith (Nouns)



Keswick(Nouns)



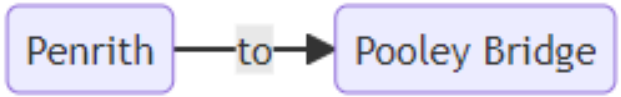
Rule-Based Method

- **Challenges**:
  - Requires a comprehensive list of the entities
  - Needs hand-crafted rules for all possible scenarios
    - e.g. spelling errors, inconsistent capitalisation, lemmas, and their inflections, etc.
  - Difficult to capture time and date in all possible formats
  - It will not generalise well with other corpora
- **Possible solution:**
  - Approach as *named entity recognition*[1] task
  - Adapt an off-the-shelf statistical model for named entity recognition
  - Spacy NLP tool has a built-in NER (named-entity recognizer) in its pipeline

[1]**Named entity**: "lexical unit referring to a real-world entity (typically a proper noun or an acronym) in certain specific domains, notably the human, social, political, economic and geographic domains" (Novel, Erhman, Rosset 2016)

Future work…

- So far, we have built a baseline tool that quickly extracts
  - *place names, geo feature nouns, time and date mentions etc.*
- Next is
  - To extract and analyze **relationships** and other qualitative and quantitative details in the text.
    - "*From Penrith **to** Pooley Bridge*"



  - Extracting and analyzing **sense-of-place** with
    - sentiments classification or similar techniques
    - Example, looking at the frequency list of surrounding words around an entity
  - Complete the annotation of our dataset
  - Training deep neural net model for the extraction task

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

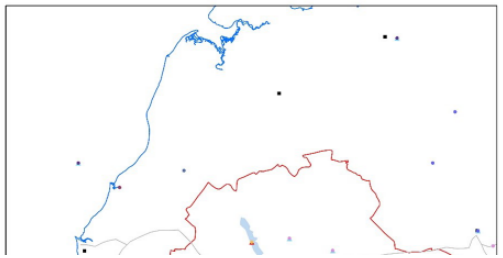
Lancaster University, Lancaster, UK

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



Thanks for your attention!