NLP&Geospatial-Powered Scholarly Analysis and Visualization of Texts

SeyedAli Ghasempouri

University of Bologna, Department of Classical Philology and Italian Studies

1. Why do we need a NLP & GeoSpatial based scholarly digital edition of texts and how are we supposed to benefit from it?

Enhanced Understanding: Helps to understand the content of the texts in a more nuanced way. It can reveal patterns, themes, and sentiments. This can provide new insights into thoughts, feelings, and experiences, as well as the broader historical and cultural context [1,2].

Efficient Analysis: Analyzing historical texts manually can be a time-consuming process, especially when dealing with a large number of documents. NLP can automate much of this process, making it possible to analyze the letters more efficiently [3].

Interactive Exploration: The interactive features of the SDE, such as word clouds, frequency plots, and part-of-speech histograms, can provide a more engaging and immersive experience for users. They can explore the text in a dynamic way, generating visualizations based on their interests [4].

Facilitating Research: The NLP-based SDE can be a valuable resource for scholars and students It can facilitate research by providing a rich, interactive, and easily accessible dataset for analysis [5].

2. What do we need to create a NLP-based SDE and what kind of edition is best suited to our case?

Selection of NLP Tools: The choice of NLP tools depends on the specific tasks we want to perform (e.g., tokenization, part-of-speech tagging, named entity recognition, sentiment analysis). Python has several libraries for NLP, such as NLTK, SpaCy, and TextBlob, which can be used for these tasks [7].

Development of the Web Application: for developing interactive web applications should allows us to create interactive charts and could work with different data formats such as CSV and XML [8].

Inclusion of TEI: The NLP analysis should be complemented with Text Encoding Initiative as a standardized markup for digital texts, aiding in detailed encoding for preservation and analysis. TEI's benefits are exemplified by the ability to differentiate character names and filter occurrences.

3. What is the foundational framework for the project's methodology?

Oyvind Eid's Critical Stepwise Formalization(CSF): it aims to use the computer to get beyond human meaning-seeking interpretations of media expressions.

CSF consists of studying a media expression through the process of adapting it into a new expression in another qualified medium. It is done through capturing the spatial understanding expressed in a text in a computer-based conceptual model and then transferring that model stepwise into a map.

The stages are: (1) The source text, which is the starting point, (2) The primary model, (3) The formalised model, (4) Geographical vector data, (5) Visualisations in the form of maps

Facilitating Research: The NLP-based SDE can be a valuable resource for scholars and students It can facilitate research by providing a rich, interactive, and easily accessible dataset for analysis [5].

4. What's the significance of applying geospatial analysis in literary studies and digital humanities?

Spatial Understanding: it can offer illuminating the significance of place and space in a work. Many literary works, for example, reference specific geographic locations, or their narratives are deeply influenced by the characteristics of the environments in which they are set. Seeing these settings visually can greatly enhance our understanding of these aspects.

Comparative Analysis: Geospatial prototyping can facilitate comparative studies of different works by the same author, or different authors writing about the same place. Seeing these works in a geographical context can shed light on thematic similarities and differences.

Pattern Recognition: A visual representation can help identify patterns that might be less apparent when reading the text. This might include recurring themes related to specific locations, trends in the movement of characters, or the evolution of a place over time as depicted in a series of works.

5. Which Python libraries and Packages can help achieve the expected results?

- **Streamlit:** A framework for building web applications.
- **spaCy:** A library for advanced NLP tasks.
- **neattext:** A text cleaning library.
- **TextBlob:** A library for sentiment analysis.
- **nltk:** A library for working with human language data (text).
- **wordcloud:** A library for generating word clouds.
- **matplotlib, seaborn:** Libraries for creating static, animated, and interactive visualizations.
- **pandas:** A library for data manipulation and analysis.
- **folium:** A library for creating interactive maps.
- **xml.etree.ElementTree:** It implements a simple and efficient API for parsing and creating XML data.
- **datefinder:** A python module for locating dates inside text.
- **Gensim:** An open-source library for unsupervised topic modeling and natural language processing.
- **Seaborn:** A Python data visualization library based on Matplotlib, used for drawing attractive statistical graphics.

Mary Mitford's "Our Village" & Walter Christaller's Central Places

Introduction

Christaller's theory posits that settlements (in this case, the village) act as central places providing services to surrounding areas, arranged in a hierarchical manner and following geometric patterns like hexagons or concentric circles.

Mitford's narrative mapping challenges linear spatial concepts and emphasizes a circular structure, focusing first on interpersonal relationships closer to the village center, and then on natural spectacles and collective events in the periphery. This pattern aligns with Christaller's notion of how central places serve different purposes and hierarchies within their spheres of influence.

We aim to provide a multi-dimensional, data-driven analytical framework that synergizes the spatial theories of Christaller with the narrative innovations of Mitford, enriching our understanding of both.

Objectives

- **Semantic Depth:** "TEI XML Generation" feature goes beyond mere representation of a text's structure; it can represent the semantics of textual features. For instance, named entities like persons, places, or dates can be semantically marked, facilitating richer analysis.
- **Enhanced Visualization:** TEI-encoded documents can be transformed and visualized in various ways using stylesheets, facilitating different modes of textual analysis and presentation.
- **Sentiment Analysis:** For stories focusing on personal relationships ('Ellen', 'Hannah', 'Cousin Mary'), a sentiment analysis could provide insights into the narrative's emotional weight as it correlates with geographic proximity to the village center.

Thanks for your attention