Exposing Geospatial Cultural Heritage Content in Map-based Applications

Nikos Bikakis[2,1], Giorgos Giannopoulos[1], Nikolaos Sidiropoulos[1], Christina Flouda[1], Athanasios Doupas[1], Voula Giouli[1], Panagiotis Karioris[1], Paraskevi Botini[1], Anna Vacalopoulou[1], Gregory Stainhaouer[1]

[1]*ATHENA Research Center, Greece* [2] *Hellenic Mediterranean University, Greece*

Abstract

This paper presents our implementation of interactive services related to map-based exploration of cultural heritage geospatial data. We first present our methodology for interlinking geographical entities (e.g., Points of Interest), i.e., the identification of same real-world geographical representations between different data sources. Then, we present efficient methods towards functionalities related to itinerary planning. The methods are part of the Mythotopia infrastructure which includes multilingual and multimodal digital content in the humanities, and advanced exploration-based functionalities. An interface allows users to interact and explore the rich content of the corpus.

Keywords
Map-based exploration, Itinerary planning, Geospatial entity matching, Path queries.

1. Introduction

The Mythotopia infrastructure integrates multilingual and multimodal digital content that is centered around the area of East Macedonia and Thrace (Greece), featuring geographical entities, natural resources, tangible and intangible cultural assets, the wealth of ancient myths and the associated mythological figures. This multifaceted digital content needs advanced exploration and visualization functionalities. The paper presents the infrastructure focusing on the functionalities related to entity matching and enrichment, and itinerary planning.

2. The Mythotopia Digital Content

The digital content comprises: (1) a mythological component (narratives and literary texts); (2) a multimodal corpus of Culture and Archaeology, and (3) a bilingual and multimodal corpus of Tourism and Travel. The three have been collected to better visualizing the essence of the area and providing enhanced experiences to visitors. The narratives for a collection of texts in Modern Greek (EL) and their translations in English (EN) depicting myths of the area, mythological figures and localities. Each myth is supplemented with literary texts in Ancient Greek and Latin and their translations in EL and EN; the literary texts are accompanied with textual material (also EL and EN). The sub-corpus of Archaeology is multimodal including images, while the multimodal corpus of Travel comprises informative texts about the Points of Interest (POIs) and images. The entire corpus is also oral consisting of sound files. POIs-related texts were manually geotagged and inter-linked as appropriate. Metadata were added to the texts and images for indexing, retrieval, and interlinking among entities. Documentation builds on well-established standards (i.e., TEI, CIDOC), while a data-specific thesaurus (taxonomy) with a set of controlled vocabularies were defined [3]. Finally, part of the textual data was also geotagged.

2.1. Basic Entities

Myths (narratives), *Persons*, *Literary texts*, *Images of artifacts*, *Textual data*, *POIs*, as well as *Routes* (linking POIs with the rest) are the corpus entities.

More precisely, *Myths* are the main entities of the platform linked with other entities: literary texts, art photos, POIs (i.e., texts, images, audio, geographical entities) *Literary texts* were inserted manually to the backend, whereas *Multimedia content* was collected by onsite photo shooting. Similarly, *POIs* are geographical entities and tangible/intangible cultural assets, i.e., archaeological/historical sites, museums, galleries, wildlife gastronomy, cultural events, inter *alii*.

Proceedings of the 6th International Workshop on Big Data Visual Exploration and Analytics co-located with EDBT/ICDT 2023 Joint Conference (March 28-31, 2023), Ioannina, GR.

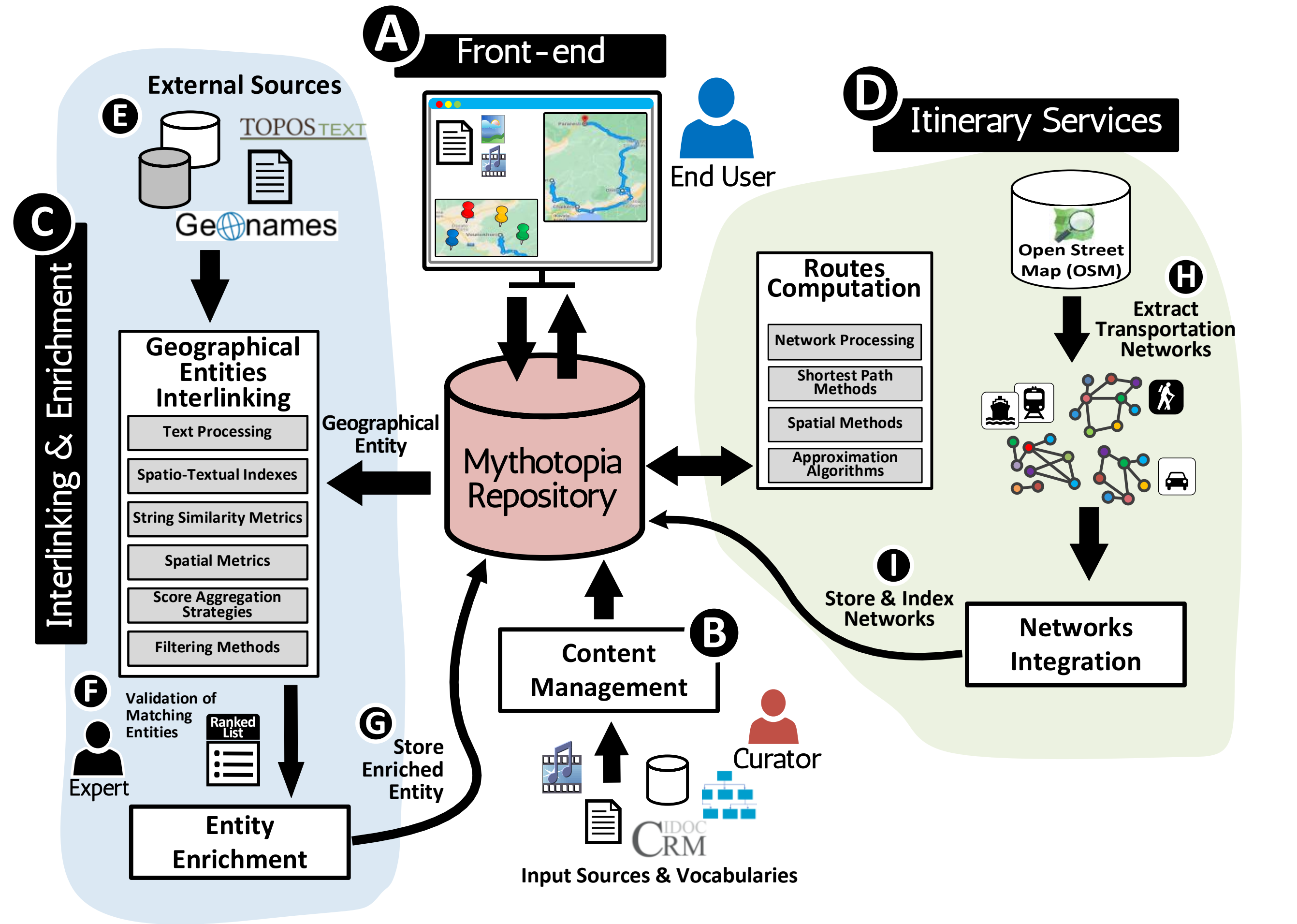

Figure 1: Mythotopia Infrastructure Overview

POIs are linked with myths via a separate interconnection process, providing map-based visualizations. *Routes* are separate entities defining the interconnection of POIs on the map by means of transportation.

3. Infrastructure Overview

Figure 1 presents an overview of the Mythotopia infrastructure, which mainly consists of: (1) *Mythotopia Repository*; (2) *User Interface*; and (3) *Core platform*.

The *User Interface* Ⓐ offers several functionalities to the users, such as, exploration over corpus content, interactive maps, travel itinerary functionalities (details in Sect. 4).

The *Core platform* contains several components related to data curation, semantic annotations, geotagging (*Content Management* Ⓑ component). Here we focus on the components related: to *interlinking and enrichment of geographical entities* with external sources Ⓒ; and *travel itinerary services* Ⓓ.

3.1. Interlink and Enrich Geographical Entities

The *Geographical Entity Interlinking* component offers semi-automatic interlinking of geographical entities, i.e., the identification of same real-world geographical representations between the corpus and external data sources Ⓔ that potentially contain the same entities, e.g., Geonames, ToposText.

In the involved sources, each geographical entity includes these basic attributes: (1) *name*; (2) *alternative name*s; and (3) *geography type* (e.g., Point, Polygon) and (4) *coordinates*. The implemented interlining mechanism applies similarity functions over textual and spatial attributes, and score aggregation techniques, utilizing parts from our previous work [1]. We next briefly outline the interlinking process.

During preprocessing (Figure 1), the *Text processing* subcomponent performs text processing over the textual attributes of the external source by, e.g., removing punctuation marks, stripping accents, and sorting terms alphanumerically. In the next step, the *Spatio-Textual Indexing* subcomponent constructs different textual indexes and an alyzers for textual attributes involved in the interlinking process. Additionally, indexes are built for spatial attributes (i.e., coordinates), enabling spatial-based comparisons, e.g., measuring geographical distances between entities. In our current implementation, the *Spatio-Textual Indexing* subcomponent is built on top of the Apache Lucene system.

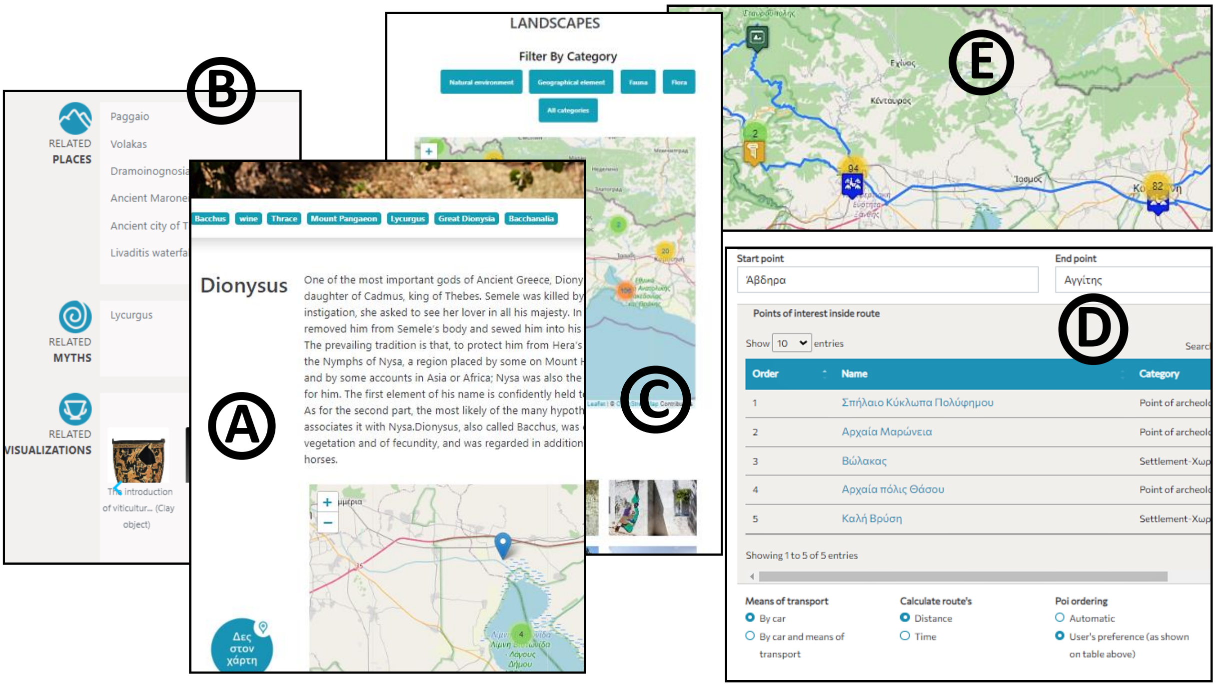

Figure 2: Mythotopia User Interface

In the interlinking phase, a geographical entity from the corpus is used as input. The *Interlinking component* exploits indexes to compare input with external entities. The comparison is performed over: (1) textual attributes by using string similarity metrics, e.g., Damerau-Levenshtein, Cosine N-Grams; and (2) spatial attributes by using spatial metrics, e.g., Euclidean distance between entities' coordinates.

Similarity scores over different textual attributes are aggregated using different scoring strategies (e.g., weighted average, threshold-based average), resulting to an overall matching score. Based on this score, an initial list of candidate matches between the corpus entity and a set of external entities is compiled. Then, the score calculated by the spatial attributes is used to prune matches from the candidate list. The framework returns a ranked list of candidate matchings 🅕.

Next, a domain-expert manually performs a validation step by selecting from the candidate list the entity identified as same with the internal entity.

In the final step 🅖, *Entity Enrichment* component uses the external entity metadata to generate an enriched geographical entity in the Mythotopia repository.

3.2. Travel Itinerary Services

The infrastructure provides numerous routing functionalities 🅓, enabling users to generate different types of travel itineraries based on several criteria, such as, travellers' preferences (e.g., means of transport) and restrictions (e.g., limited time). Efficient algorithms have been developed, offering complex routes computations.

In a nutshell, the user provides a starting and an ending location, as well as a set of POIs which they wish to visit. The system generates a route (roads/paths) that need to be followed by the user, in order to visit all the desired POIs.

Itinerary Planning. Users may request an itinerary using these options: **(1)** *Route optimization objective:* (a) Route Length, (b) Traveling Time. **(2)** *Travel type:* It defines the means of transport to be used in the route, allowing three options: *Driving*, *Walking*, and *Transit* (e.g., train, boat). Travel Type determines the transport network on top of which the route is computed, e.g., streets and the paths that pedestrians, highways. **(3)** *Visiting order mode*: It allows the user to select whether they prefer: (a) visit the desired POIs in a specific order, (b) allow the algorithm to find an optimal order of visitation wrt *Route optimization objective*. **(4)** *Approximate route computation*.

Transportation Networks. The *Networks Integration* component extracts and integrates different transportation networks 🅗 (e.g., road networks, railways, boat routes)[2], enabling route generation based on different means of transport. The resulted networks are indexed and stored in the Mythotopia database 🅘.

Efficient Itinerary Computation. The *Routes Computation* component enables efficient route computation,

[2] The networks are extracted from OpenStreetMap.

which is essential for interactive applications. Routing functionalities involve the combination of shortest path algorithms, spatial processing (e.g., space/network pruning), network processing (e.g., creating overlay networks), multi-criteria optimizations, and approximation methods. Part of implementations is based on methods proposed in our previous work in [1] which allows efficient route planning over multiple POIs.

The main idea of the implemented method is the following. Initially, we identify an adequate subset of the transportation network, that is guaranteed to contain, among others, the optimal route for the user's input. This way, parts of the road network are pruned, reducing the search space of the algorithm, and enabling the efficient route generation.

4. User Interface Functionality

This section briefly introduces the Mythotopia user interface (Fig. 2). The basic functionality includes:

Corpus Exploration & Discovery. User can navigate over the whole corpus (Fig. 2 Ⓐ Ⓑ). Starting from an entity (e.g., myth, person, POI, route) users can discover its related entities, and associated information, such as textual descriptions, multimedia content, references, bibliography.

Map-based Exploration. Users may perform visual exploration over different geographical areas via a map visualization (Fig. 2 Ⓐ Ⓒ). Geotagged entities (e.g., myth, POIs, activities contained in a geographical area are visualized. Faceted exploration is supported for filtering visualized entities.

Itinerary Planning. Route construction UI provides several options, offering the user the capability to specify several preferences and restrictions (see Sect. 0). The user selects a starting and an ending location, POIs, and provides their requirements, such as the means of transportation, the order of visiting the POIs Ⓓ. The computed route is presented on the map Ⓔ. Additionally, for each of the route's POI, the UI presents nearby POIs. This way, the UI suggests to the user other POIs which can be "easily" reached.

Itinerary Recommendations. Beyond user-specified itineraries, users can look for *myth-related routs*. Here, the system considers the POIs that are associated with the myth recommending a set of itineraries that involve related POIs.

Moreover, numerous *predefined itineraries have been assigned to themes* (e.g., highland paths, paths that involve POIs related to urban legends). So, the user can select they preferred theme and examine the routs presented in the UI.

5. Demonstration Outline

Here we outline our demonstration scenario. Attendees will be able to interact with the infrastructure via the UI (Fig. 2). Initially, users will be presented with various supported functionalities, as well as, specific scenarios that will provide better insight UI capabilities. Then, they will be able to interact with the UI.

As a simple demonstration scenario, we assume that user starts their exploration by selecting a myth. Myth-related resources and information are presented to the user Ⓐ Ⓑ. Moreover, a map depicts myth-related POIs Ⓐ. Then, users can select a POI to see more details in its information page.

Next, users examine the route functionalities. First, they select POIs and specify their preferences Ⓓ. Then, the computed route is presented on the map, along with nearby POIs Ⓔ. Users examine the routes related to a specific myth. Finally, they select one of the predefined route themes and inspect associated routes.

Acknowledgement. We acknowledge support of this work by the project "Mythological Routes in Eastern Macedonia and Thrace" (MIS 5047101) which is implemented under the Action "Regional Excellence", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

6. References

[1] Kaffes V., Belesiotis A., Skoutas D., Skiadopoulos S., "Finding shortest keyword covering routes in road networks". SSDBM 2018.

[2] Giannopoulos G., Kaffes V., and Kostoulas G., "Learning Advanced Similarities and Training Features for Toponym Interlinking". ECIR 2020

[3] Voula, G., Vacalopoulou, A., Sidiropoulos, N., Flouda, C., Doupas, A., Stainhaouer, G., "From mythos to logos: A bilingual thesaurus tailored to meet users' needs within the ecosystem of cultural tourism", EURALEX International Congress, 2022

[4] Giouli V., Vacalopoulou A., Sidiropoulos N., Flouda C., Doupas A., Giannopoulos G., Bikakis N., Kaffes V., Stainhaouer G. "Placing multi-modal, and multi-lingual Data in the Humanities Domain on the Map: the Mythotopia Geo-tagged Corpus". In Proc. of 13th Intl. Conf. on Language Resources & Evaluation (LREC 2022)