We present an infrastructure that comprises a multi-lingual and multi-modal corpus (i.e., a corpus of textual data supplemented with images and videos) that belongs to the humanities domain along with a dedicated database (content management system) with advanced indexing, linking, and search functionalities. This infrastructure also includes a geotagging component and will be integrated into a platform aimed at defining personalized itineraries and providing, thus, a multi-faceted experience to visitors of Eastern Macedonia and Thrace in Northern Greece using mythology as a starting point.

**Abstract**

The Mythotopia corpus is being created as part of an infrastructure aimed at the development of an online platform offering a multifaceted view of the area using mythology as starting point. The mythological content is further enhanced with tangible and intangible elements that pertain to the domains of history, architecture, environment, culture, society, folklore, recreation, gastronomy, travel and tourism, leisure, and more (Vacalopoulou et al., 2021). The platform uses different types of data to facilitate search and retrieve functionalities based on several criteria, also offering the option of defining personalized itineraries in the area based on these criteria. Therefore, the problem of “multimodal location estimation” lies within the heart of the overall project. Consequently, Points of Interest (POIs), i.e., geospatial entities that are (a) characterised by at least a name and a set of coordinates, and (b) correspond to a place of interest to end users, are deemed as core elements handled in the Mythotopia corpus. Places, facilities, artefacts, living entities (i.e., persons, plants, and animals), events, and even intangible cultural heritage items that may be placed on a map are deemed as POIs.

**Rationale & Scope**

The Mythotopia corpus is being created as part of an infrastructure aimed at the development of an online platform offering a multifaceted view of the area using mythology as a starting point. The mythological content is further enhanced with tangible and intangible elements that pertain to the domains of history, architecture, environment, culture, society, folklore, recreation, gastronomy, travel and tourism, leisure, and more (Vacalopoulou et al., 2021). The platform uses different types of data to facilitate search and retrieve functionalities based on several criteria, also offering the option of defining personalized itineraries in the area based on these criteria. Therefore, the problem of “multimodal location estimation” lies within the heart of the overall project. Consequently, Points of Interest (POIs), i.e., geospatial entities that are (a) characterised by at least a name and a set of coordinates, and (b) correspond to a place of interest to end users, are deemed as core elements handled in the Mythotopia corpus. Places, facilities, artefacts, living entities (i.e., persons, plants, and animals), events, and even intangible cultural heritage items that may be placed on a map are deemed as POIs.

**Corpus description**

The corpus comprises three components:

- Sub-corpora of literary texts (LA, GRC, EL, EN) → primary data + accompanying texts (EL, EN)
- The cultural component (images, videos) of artifacts coupled with texts (EL, EN) as accompanying material
- Sub-corpora in the domain of Travel, texts (EL, EN) and images as accompanying material

Selected by experts based on certain criteria defined:

- Relevance to the myths of the area;
- Availability of primary data.

Textual data in the Travel domain are crated ad hoc.

Metadata were added to both primary data and accompanying material (texts, images). Annotations were integrated across the following pillars: (a) efficient documentation aimed at indexing and retrieval of the content; (b) interlinking of the various entries in the database; (c) corresponding to a place of interest to end users, are deemed as core elements handled in the Mythotopia corpus. Places, facilities, artefacts, living entities (i.e., persons, plants, and animals), events, and even intangible cultural heritage items that may be placed on a map are deemed as POIs.

**Storing in a Database**

The corpus is stored in a Database that was ultimately used for searching the data. It serves as a content management system (CMS) and was built in-house.

Database functionalities:

- Documentation of texts, images and video data
- Interlinking of the data, i.e., texts with images, video, and accompanying texts
- Geotagging of the textual data referring to POIs.

The platform follows the database schema (depicted in Figure 1) and provides a menu containing the main data elements of the corpus: myths, reference texts, artistic representations, mythological figures (people), multimedia files and POIs. For each of these entities, the platform provides a common management mechanism.

To better account for quality assurance and security control, different accessibility rights are granted to users. At the lower level, annotators are allowed to access all corpus elements and can create, edit, and delete their own records. Editors, on the other hand, have the same access to system functionalities, but also the right to publish completed records.

**References (selected)**

