Empowering Digital Transformation in Tourism through Intelligent Methods for Representation and Exploitation of Cultural Heritage Knowledge

Salvatore Carta, Gianni Fenu, Alessandro Giuliani*, Marco Manolo Manca, Mirko Marras, Leonardo Piano, Alessandro Sebastian Podda, Livio Pompianu and Sandro Gabriele Tiddia

University of Cagliari, Department of Mathematics and Computer Science, Cagliari, Italy

Abstract

Nowadays, the cultural heritage wealth held within museums, archives, libraries, and heritage sites is under continuous threat due to various environmental, geopolitical, and resource factors. The need for proper preservation of such invaluable cultural assets worsens the risk of their loss and hinders global access to their knowledge. This paper describes our first steps towards preserving cultural heritage knowledge through intelligent digital services. Our work is conducted as part of the PNRR e.INS "Ecosystem of Innovation for Next Generation Sardinia" project, Spoke 6, which aims to support digital transformation in various domains, including cultural heritage, in the Sardinian region, Italy. The main objectives of this project include the digitization of cultural heritage knowledge by employing innovative solutions for a proper knowledge representation, the exploitation of such digital knowledge for proposing contextual and personalized services, and suitable user-friendly interfaces and applications that enable both operators and the general public to access, explore, and engage with such knowledge.

Keywords

Cultural Heritage, Knowledge Graph, Knowledge Extraction, Personalization, Ranking, Recommendation.

1. Background

Cultural Heritage (CH) is the legacy of tangible and intangible assets representing human culture, history, identity, and diversity. UNESCO defines it as "*the entire corpus of material signs – either artistic or symbolic – handed on by the past to each culture and, therefore, to the whole of humankind*"[1]. It includes numerous varieties of entities, such as archaeological sites, monuments, museums, artworks, music, or festivals. CH is a priceless educational, touristic, economic, and social cohesion resource [1]. However, it continuously faces challenges threatening its existence and transmission to future generations. Such threats are related to manifold aspects. On the

[1]https://unesdoc.unesco.org/ark:/48223/pf0000084697

one hand, some of them are often unpredictable or hard to handle, such as human-induced risks (e.g., vandalism or looting) or natural phenomena (e.g., natural disasters or climate change). On the other hand, a crucial weakness is the lack of adequate resources, strategies, or policies for the conservation of CH. The preservation of CH is an essential task that requires appropriate resources and policies to protect human culture's diverse and incalculable expressions. In this scenario, numerous CH sites and entities must address various threats to their preservation and sustainability. Some of them include:

- **The lack of funding or expertise.** CH entities need regular maintenance and documentation to avoid deterioration. These actions often require financial and human resources not always available or affordable. For example, some institutions may need more funds, equipment, materials, or experienced specialists to carry out adequate interventions.
- **The scarcity of documentation.** From the point of view of promotion and dissemination, many CH sites or artefacts may not have accurate documents of their history, origin, or general information to assign them the proper value and significance.
- **The shortage of awareness and participation.** Numerous CH sites depend on the attention and participation of the general public and local communities to ensure their appreciation. However, stakeholders often lack knowledge, motivation, or interest to promote the respect and value of the CH the people encounter or to engage them in collateral activities, leading to the scenario where people may be unaware of the significance of particular CH entities or may have adverse or indifferent attitudes toward them.

In such a context, the CH domain must specialize and become more innovative and sustainable to enhance the preservation and conservation of cultural identity. For this purpose, promoting and strengthening the identity and thousand-year history of our territory and culture plays a crucial role in facilitating the growth of awareness and participation of people and local communities and in encouraging stakeholders and institutions to invest effort and resources for CH preservation. Our work aims to propose innovative solutions for addressing the most critical issues in this scenario, as described in the following Sections 2 and 3.

2. Motivation

The lack of appropriate instruments has often led to losing valuable content for preserving and promoting our cultural heritage. To overcome this issue, the typical solution is the integration of specific *digital technologies and services* [2, 3, 4, 5] that (i) ensure stability and integrity over time, (ii) transform content to be more accessible, usable, and interactive, and (iii) enable the use of communication channels, such as the Web or mobile technologies, facilitating dissemination on a global scale. However, the digitization is currently facing several issues [6, 7]. In particular, digital content must be *actionable*, i.e., capable of stimulating concrete actions by users. Therefore, defining or adopting strategies to engage audiences actively and creatively, offering immersive, personalized, and participatory experiences is crucial.

In this context, an adequate knowledge representation is paramount for implementing digital services [8, 9]. Indeed, cultural archives are often digitized by scanning documents and photos, including metadata, pre-processing, and indexing. The digitized data are usually shared as static open data archives to facilitate access, innovation, and collaboration, enabling the development

of virtual reality, mobile, and Web applications. Static archives highly limit the understanding of CH because of the lack of dynamic and interconnected representations of information, and they usually fail to capture the complex relationships between entities, thus limiting the analysis of historical and cultural relationships. For example, let us consider an institutional site that encapsulates information to help tourists discover a historic city in Italy. Static open data archives could be used on a website page to list places of interest with essential details such as names, addresses, and times. Suppose a tourist wants a deeper, more connected perspective. In that case, the limitation of static archives emerges, e.g., they may want to understand the relationships between historical places, such as which monument is related to a particular historical event or which artist created a work of art. These interconnections are complex to represent in a static archive. In contrast, Knowledge Graphs (KGs) [10], i.e., structured representation of information that semantically links concepts, entities, and relationships in a meaningful and easily interpretable way, could visualize connections between monuments, historical events, and figures, providing a more comprehensive view. Unlike static archives, which are isolated data collections, KGs highlight connections between data, allowing for a deeper, more contextualized understanding. This would uncover interesting relationships, linking cultural elements, local traditions, and contemporary contexts.

However, creating KGs is a challenging task [11]. Defining semantic relationships requires domain understanding and specific skills. The lack of intuitive and automated tools hinders widespread KG creation, especially for those without technical skills. To this end, our work focuses on defining and implementing a new no-code methodology for creating and exploiting KGs, including two key elements. First, we aim to design an innovative automated artificial intelligence-based approach that employs Large Language Models (LLMs) [12] to transform unstructured and structured digitized information into coherent semantic triples (entity, relation, entity). These triples will form the nodes and relationships of the KG. In particular, the devised KGs can link and organize data from historical archives, museums, cultural institutions, tourist guides, and other relevant sources, enabling an in-depth understanding of our culture, facilitating the user experience, as well as creating contextual or personalized recommendations. Following the previous example, suppose a tourist explores a Roman historic site or monument through its website. He/she might receive offers on related souvenirs, such as Roman temple miniatures or books on history, guided tours of the site's history, restaurants with Roman cuisine, scheduled events, and personalized suggestions based on user's past visits and preferences. The second innovative contribution of the work focuses on artificial intelligence-based methods that use LLMs for exploiting KGs in contextual and personalized targeting, thus providing accurate and explainable suggestions. The following section describes our contribution in detail.

3. Contribution

This contribution describes our first steps toward promoting the preservation of CH knowledge by devising and implementing innovative digital services. Our work is conducted as part of the PNRR e.INS "Ecosystem of Innovation for Next Generation Sardinia" project, Spoke 6, aimed at supporting digital transformation in various domains, including cultural heritage.

3.1. Overview

In the context of improved digital transformation in tourism, the main goal of our work is the cooperation between technology partners and culture operators to develop a highly innovative service for the automatic digitization and interconnection of cultural heritage through the generation of proper KGs from structured and unstructured information and their exploitation for devising specific novel services also tailored to individual visitors or specific groups. As KGs can link and organize data from historical archives, museums, cultural institutions, tourist guides, and other relevant sources, enabling an in-depth understanding of our cultural heritage and facilitating the user experience, creating proper KGs and related digital services facilitates the enhancement and promotion of our heritage. They also enable knowledge sharing among various communities and make collaboration among cultural institutions possible. In detail, appropriate KGs for cultural heritage through innovative artificial intelligence techniques that exploit the potential of LLMs will be created, and algorithms and platforms that implement specific contextual and personalized targeting systems will be developed. For the sake of completeness, let us point out that nevertheless LLMs may produce wrong facts due to their embeddings [13] and training methodologies [14] or may generate biased content, especially when used for personalization [15], a proper prompting-based strategy may outperform classical or human-based methodologies [16], also in proposing proper recommendations [17].

3.2. Objectives and outcomes

The devised tools and services aim to increase the competitiveness and attractiveness of the Sardinian region. The goal is guided by a series of objectives, including:

- The **digitization** of cultural heritage by creating usable, comprehensive, robust, and coherent knowledge graphs.
- The **exploitation** of the created KGs for innovative, effective, and efficient products and services that take advantage of the potential of the developed knowledge base.
- The **development** of a technological infrastructure that provides the aforementioned contextual and personalized functionalities.

Our work will yield distinct products and outcomes, as follows:

- First, proper **KGs** for the cultural heritage domain, created in accordance with the requirements analysis and use cases defined during the first step.
- The KGs will be fed into **contextual and personalized targeting** systems that provide (groups of) visitors with suggestions based on their context, interests, and preferences.
- A properly accessible, usable, immersive, and customizable **technological infrastructure** that provides the contextual and personalized features devised in the earlier stages.

3.3. Architecture

For the sake of clarity, in Figure 1, the high-level architecture of the general platform is depicted. According to the primary outcomes described in Section 3.2, the platform is composed of three main modules: *knowledge graph generation*, *contextual and personalized targeting*, and *infrastructure*. Each module is briefly described in the following.
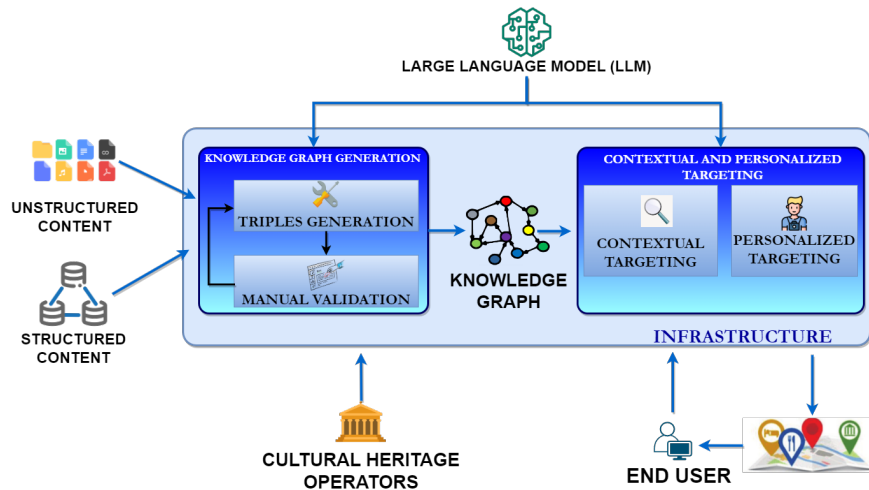
Figure 1: High-level architecture of the devised platform.

3.3.1. Knowledge Graph Generation

This module is aimed at generating the final KGs for CH. In particular, for each use case defined during the project, relevant triples (in the form of `<subject, relation, object>`) are extracted from structured or unstructured textual sources provided by stakeholders by employing innovative algorithms based on a proper LLM-prompting strategy. A specific computer-aided validation methodology will be used to investigate the correctness of the extracted triples. In detail, innovative and efficient active learning algorithms will embrace an incremental approach for determining the correctness of triples. This will be achieved through an iterative query strategy in which, at each step, a human operator is asked to validate a limited set of triples, pre-classified by the algorithm trained in the previous step, with the aim of incrementally correcting the error, and, therefore, improving the effectiveness step by step. The output will be a proper KG composed of the generated triples for a given use case.

3.3.2. Contextual and personalized targeting

This module aims to generate proper *explainable* recommendations by innovative LLM-based algorithms that explore the paths of the generated KGs. For example, if we consider the path [`<user_1>` visited `<Castle of Cagliari>` associated_to `<Ancient Sardinia>` associated_to `<ExhibitionX>`], the resulting text to explain a recommendation for the place `ExhibitionX` to `user_1` could be "*It is related to the Castle of Cagliari, which you visited recently, as both belong to the Ancient Sardinia category*". This approach improves the user experience and confidence. This module first analyzes the KGs by exploiting ad-hoc methodologies to understand entity structures and relationships, identify affinity relationships between entities, and adequately represent entities to overcome the limitations of existing methods. Afterwards, the module generates recommendations with a twofold strategy. On the one hand, specific targets, such as identifying similar entities or forming clusters of related entities, are defined. Recommendations are generated based on clusters and relationships, validating their

quality (*contextual targeting*). On the other hand, personalized representations are created by combining affinity information and specific user characteristics, including interaction history. The personalized representations will allow us to provide recommendations tailored to users' individual needs or specific contexts (*personalized targeting*).

3.3.3. Infrastructure

The final technological infrastructures will integrate (i) the essential services for knowledge *ingestion*, i.e., the information extraction and management from multiple sources of structured and unstructured information; (ii) the KG creation and management services, including the human-assisted triples validation; (iii) the targeting services (both contextual and personalized); (iv) the application layer of end-user services with various management services (e.g., user creation, profiling, ID management and authentication, travel experience historicization, or sharing and social media interfaces). The integrated platform will be accessible to two main categories of users: first, the entities involved in the knowledge construction phase (cultural institutions, touristic operators, or other stakeholders), i.e., users designated for the data ingestion and the iterative manual triples validation, and second, the end-users, such as a website visitor, which will be the beneficiary of the targeting services.

4. Expected Impact

The final integrated tool may enable (a) high quality and customization in the search, retrieval, and fruition of knowledge in the cultural domain referring to a target territory; (b) active support and guidance of the user in the construction of custom services such as route stages and places/resources according to his or her interests; (c) proactive formulation of personalized suggestions and recommendations regarding ideas, visits, experiences, and related products, (d) support in sharing knowledge and experiences with other users and visitors connected to the platform and its services to assist in creating and developing cultural communities.

Our work is expected to positively affect the cultural communities from several perspectives. In particular, we expect an increase in the quality and quantity of available information on the CH and the related services of the involved territories, increased visibility and enhancement of the cultural heritage at a national and international level, greater collaboration and sharing among the different communities and cultural institutions involved in the project and augmented competitiveness and innovation of tourism and culture operators in Sardinia.

Acknowledgments

References

[1] E. Panzera, The Socio-Economic Impact of Cultural Heritage: Setting the Scene, Springer International Publishing, Cham, 2022, pp. 1–41. doi:`10.1007/978-3-030-94468-1_1`.

[2] F. Maietti, Heritage enhancement through digital tools for sustainable fruition—a conceptual framework, Sustainability (Switzerland) 15 (2023). doi:`10.3390/su151511799`.

[3] M. Morandotti, E. Doria, Information system as tool for cultural heritage documentation and preservation. protocol structuring and testing on a case study, volume 48, 2023, p. 1081 − 1088. doi:`10.5194/isprs-Archives-XLVIII-M-2-2023-1081-2023`.

[4] M. Raykovska, K. Jones, I. Vasilev, A photographic documentation workflow for digitization of cultural heritage: The 14 century church of sv. nikola in kalotina, bulgaria, volume 48, 2023, p. 1287 − 1293. doi:`10.5194/isprs-Archives-XLVIII-M-2-2023-1287-2023`.

[5] S. M. Carta, S. Consoli, A. Giuliani, A. S. Podda, D. R. Recupero, Culturai: Semantic enrichment of cultural data leveraging artificial intelligence, IEEE Access 10 (2022) 127328 − 127344. doi:`10.1109/ACCESS.2022.3226070`.

[6] L. Baratin, F. Gasparetto, V. Tronconi, What communication about heritage conservation? a challenge for digital documentation content, volume 48, 2023, p. 183 − 194. doi:`10.5194/isprs-Archives-XLVIII-M-2-2023-183-2023`.

[7] A. Adane, A. Chekole, G. Gedamu, Cultural heritage digitization: Challenges and opportunities, International Journal of Computer Applications 178 (2019) 1–5. URL: http://www.ijcaonline.org/archives/volume178/number33/30748-2019919180. doi:`10.5120/ijca2019919180`.

[8] N. Barbuti, Thinking digital libraries for preservation as digital cultural heritage: By r to r4 facet of fair principles, Int. J. Digit. Libr. 22 (2021) 309–318. URL: https://doi.org/10.1007/s00799-020-00291-7. doi:`10.1007/s00799-020-00291-7`.

[9] Y. Liu, Evaluating visitor experience of digital interpretation and presentation technologies at cultural heritage sites: a case study of the old town, zuoying, Built Heritage 4 (2020). doi:`10.1186/s43238-020-00016-4`.

[10] L. Ehrlinger, W. Wöß, Towards a definition of knowledge graphs, in: International Conference on Semantic Systems, 2016. URL: https://api.semanticscholar.org/CorpusID:8536105.

[11] C. Peng, F. Xia, M. Naseriparsa, F. Osborne, Knowledge graphs: Opportunities and challenges, Artificial Intelligence Review (2023). doi:`10.1007/s10462-023-10465-9`.

[12] S. R. Bowman, Eight things to know about large language models, 2023. `arXiv:2304.00612`.

[13] K. Sun, Y. E. Xu, H. Zha, Y. Liu, X. L. Dong, Head-to-tail: How knowledgeable are large language models (llm)? a.k.a. will llms replace knowledge graphs?, 2023. `arXiv:2308.10168`.

[14] L. Berglund, M. Tong, M. Kaufmann, M. Balesni, A. C. Stickland, T. Korbak, O. Evans, The reversal curse: Llms trained on "a is b" fail to learn "b is a", 2023. `arXiv:2309.12288`.

[15] E. J. Gerritse, F. Hasibi, A. P. de Vries, Bias in conversational search: The double-edged sword of the personalized knowledge graph, in: Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval, ACM, 2020. URL: https:

//doi.org/10.1145%2F3409256.3409834. doi:`10.1145/3409256.3409834`.

[16] J. Baek, A. F. Aji, A. Saffari, Knowledge-augmented language model prompting for zero-shot knowledge graph question answering, 2023. `arXiv:2306.04136`.

[17] R. M. Harrison, A. Dereventsov, A. Bibin, Zero-shot recommendations with pre-trained large language models for multimodal nudging, 2023. `arXiv:2309.01026`.