City-Stories: Combining Entity Linking, Multimedia Retrieval, and Crowdsourcing to Make Historical Data Accessible[*]

Laura Rettig[1,**][✉][0000−0002−9765−0549], Shaban Shabani[2,3,**][0000−0003−4710−6091], Loris Sauter[2,**][0000−0001−8046−0362], Philippe Cudré-Mauroux[1][0000−0003−2588−4212], Maria Sokhn[3][0000−0001−7586−0564], and Heiko Schuldt[2][0000−0001−9865−6371]

[1] University of Fribourg, Fribourg, Switzerland
`firstname.lastname@unifr.ch`
[2] University of Basel, Basel, Switzerland
`firstname.lastname@unibas.ch`
[3] University of Applied Sciences Western Switzerland (HES-SO), Switzerland
`firstname.lastname@hes-so.ch`

Abstract. Digitized historical image collections as provided by individuals or memory institutions often suffer from limited or a complete lack of metadata. In this paper, we present the *City-Stories* system that combines entity linking, multimedia retrieval, and crowdsourcing to make historical images searchable even across collections.

Keywords: Multimedia Retrieval · Entity Linking · Semantic Data · Crowdsourcing.

1 Introduction

Collecting, managing, and accessing historical data is essential for digital preservation of cultural heritage. This is particularly important for advanced applications that use digitized historical content shared across cultural heritage institutions and archives [6]. Sharing such data opens the door to several exciting possibilities. First, it makes it possible to integrate heterogeneous multimedia collections from different sources, formats, and metadata schemata, to ensure access via a homogeneous interface. Additionally, descriptive metadata opens the possibility to extract the context of the documents for meaningful concepts and to link the documents across media types and external collections. Second, integrated historic multimedia content allows for interactive approaches to retrieval which support different content and context-based query types such as keyword queries, query-by-example, query-by-sketch, semantic queries, spatio-temporal queries, and any combination thereof. Third, these features allow the

---

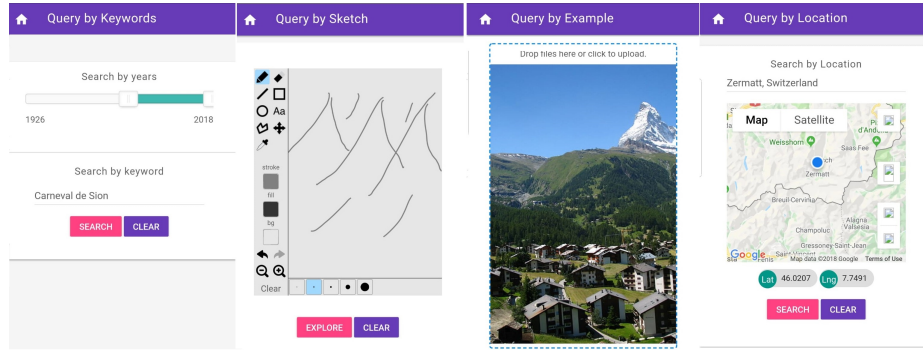[**] These authors contributed equally to this work.

Fig. 1. Screen shots of City-Stories front-end supporting different query types.

users of these applications to not only become content consumers but also content providers. Citizens own valuable private collections such as photo albums, audio, and video archives. Enabling crowdsourcing as a service allows them to share important content that can be of great public interest and contribute to the digital preservation of cultural heritage of their region. Moreover, by sharing their knowledge, citizens can play a crucial role in curating existing data.

In this paper, we present *City-Stories*, a hybrid system consisting of modules for multimedia retrieval, entity recognition and linking, and crowdsourcing for cultural heritage data. *City-Stories* enables the management, collection, and presentation of heterogeneous multimedia data in applications for cultural heritage, leveraging both content and metadata for the multimedia documents.

2   System Overview

The *City-Stories* system consists of three major components: (i) a module for *semantic data expansion*, (ii) a *spatio-temporal content browser* based on the *vitrivr* system, and, (iii) a *crowdsourcing and knowledge visualisation* module. During the offline phase, multimedia collections are extracted by a general data ingestion module and subsequently processed by the browser's underlying multimedia retrieval engine while simultaneously a semantic expansion is performed. In the online phase, additional information is gained via crowdsourcing, data that is further enhanced on-the-fly by the semantic data expansion module.

2.1   Semantic Data Expansion

The multimedia data present in this project is often accompanied by *textual metadata* (usually in the form of title and description of an item). In our previous work [6], we detail how we use *entity linking* to enrich the data presented to users with information from a knowledge base, specifically, WikiData[4]. In this

---

[4] https://www.wikidata.org

component, we use textual metadata provided alongside media items in the DigitalValais[5] and Mediathèque[6] datasets to extract and link related entities. This allows us to enhance our data with relevant information on the entities present in both the metadata and in the knowledge base and enables linking data from further sources to the same entities for integration.

Due to the specificity of our data, we encounter many *rare* entities: Such entities are not recognized by our previous entity linking pipeline. However, using *word embeddings*, we can still infer relationships between entities in absence of entries to the knowledge base. In particular, since we are dealing with short textual information—too small to learn embeddings from the entirety of the *City-Stories* corpus—we employ methods that combine different textual sources for domain-specific textual information [4] and temporal term evolution [5]. For this, we localize the closest embedding (obtained via the aforementioned methods) in the vector space of an identified entity that is present in the knowledge base.

2.2 Spatio-Temporal Content Browser

City-Stories leverages a modified version of the *vitrivr* [2] content-based multimedia retrieval system, tailored to the search in historic multimedia collections. In particular, Cineast [3], *vitrivr*'s retrieval engine, provides a plethora of query modes, of which *query-by-example* (QbE), *query-by-sketch* (QbS), *query-by-location* (QbL) and *query-by-time* (QbT) are enabled in *City-Stories*. QbE enables users to provide a sample image to be looked for. Using QbS, users might sketch the query freely or modify an existing image with a superimposed sketch. Spatial (QbL) and temporal (QbT) queries allow users to search for the time and/or place where historic objects have been captured.

QbS and QbE in *City-Stories* is based on a content-based similarity search along various features. Metadata for QbL and QbT are either extracted from the multimedia objects or provided externally. Often, historical documents lack appropriate metadata like EXIF for images and thus we heavily rely on additional data provided by the two other modules, either provided by human annotation or via semantic expansion. This data is stored in corresponding MongoDB and PostgreSQL databases, while the extracted multimedia retrieval features are stored in *vitrivr*'s database CottontailDB [1]. The *City-Stories* system communicates with the *vitrivr* system via RESTful API, leveraging the OpenAPI standard[7].

2.3 Crowdsourcing and Knowledge Visualization

The crowdsourcing component of *City-Stories* allows platform users to share multimedia content. Citizens can share their private digitized historical collections and contribute to the digital preservation of cultural heritage of their region. The cross-platform capability enables users to share their collections from

---

[5] `http://www.valais-wallis-digital.ch`

[6] `https://www.mediatheque.ch/`

[7] `https://www.openapis.org/`

desktop or mobile devices and provide metadata that cover descriptive aspects of the shared items (title, description, tags, and categories) and the spatio-temporal properties (date and location).

The type of multimedia collections considered in *City-Stories* come without or only with little geographical and temporal information. For instance, finding the location and time where and when an image was taken is a rather challenging machine learning classification task. In contrast, humans can perform better especially if the annotation tasks are properly matched with annotators' capabilities. We deploy four crowdsourcing tasks and leverage the wisdom of crowds to improve the metadata of historical collections: i.) *Location-Finder*, used for finding the place depicted in images; ii.) *Year-Finder* for identifying the year images have been captured; iii.) *Annotation-Competition* for competitive image tagging, and iv.) *Validator* to validate automatically generated tags and image categories [7]. Gamification approaches are considered to incentivize the users participating in the tasks as well as for data quality control.

3   Conclusion

In this paper, we have presented the *City-Stories* system and how it seamlessly combines content-based retrieval with entity-based navigation and leverages the wisdom of the crowd for enhancing the metadata and extending the historical collections. It allows users to browse, perform different interactive query types and explore historical data from archives and museum collections.

References

1. Gasser, R., Rossetto, L., Heller, S., Schuldt, H.: Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis. In: Proceedings of the 28th ACM International Conference on Multimedia. ACM (2020)
2. Gasser, R., Rossetto, L., Schuldt, H.: Multimodal Multimedia Retrieval with vitrivr. In: Proceedings of the International Conference on Multimedia Retrieval (2019)
3. Heller, S., Sauter, L., Schuldt, H., Rossetto, L.: Multi-Stage Queries and Temporal Scoring in vitrivr. In: Proceedings of the IEEE International Conference on Multimedia Expo Workshops (ICMEW) (2020)
4. Rettig, L., Audiffren, J., Cudré-Mauroux, P.: Fusing Vector Space Models for Domain-Specific Applications. In: Proceedings of the 31st IEEE International Conference on Tools with Artificial Intelligence, ICTAI (2019)
5. Rettig, L., Hänggli, R., Cudré-Mauroux, P.: The Best of Both Worlds: Context-Powered Word Embedding Combinations for Longitudinal Text Analysis. In: Proceedings of the IEEE International Conference on Big Data (2020)
6. Shabani, S., Sokhn, M., Rettig, L., Cudré-Mauroux, P., Beck, L., Tanase, C., Schuldt, H.: City-Stories: A Multimedia Hybrid Content and Entity Retrieval System for Historical Data. In: HistoInformatics@CIKM. pp. 22–29 (2017)
7. Shabani, S., Sokhn, M., Schuldt, H.: Hybrid Human-Machine Classification System for Cultural Heritage Data. In: Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia HeritAge Contents (2020)