Semantic Network Visualization of Cultural Heritage Data

Shaban Shabani[1,2], Zhan Liu[1], and Maria Sokhn[1]

[1] HES-SO Valais-Wallis, Institute of Information Systems, Sierre, Switzerland
zhan.liu@hevs.ch, maria.sokhn@hevs.ch
[2] University of Basel, Department of Mathematics and Computer Science, Basel, Switzerland
shaban.shabani@unibas.ch

Abstract. Advancement of digital technologies have helped the cultural heritage (CH) organizations such as archives, libraries, and museums to digitize their data collections and improve the accessibility to them. Over the past years various approaches have been developed to enable the visualization of the CH data. In this paper, we present a use case on annotation, linking, visualization and browsing of CH data and suggest some directions to achieve tangible visual analysis. We introduce the semantic linked data search and the method of connected concepts to enhance the users experience when searching for historical data.

Keywords: network visualization · cultural heritage · semantic search · linked data

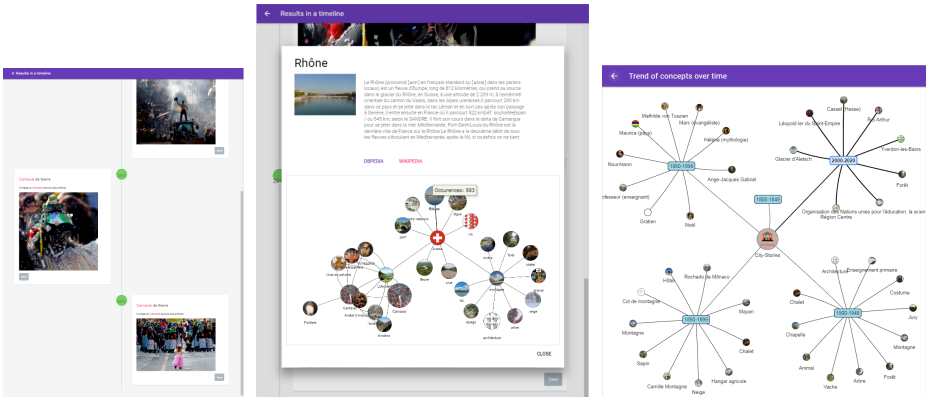
1 Introduction

The role of documenting cultural heritage is a crucial element for preserving the culture for future generations and it has been long recognized. The digitization of cultural collections is changing our perception to cultural organizations. Over the past years, cultural organizations intensively digitize their collections and make them publicly accessible [2]. Digital CH data has importance in fields such as tourism, as it helps to promote regions and cultural diversity [1].

In the cultural heritage, a quick access to the right information has an important role. Visualizing the data in a meaningful representation is a more recent concept used for accessing data, deriving meaning and acquiring knowledge from the data [4]. Data visualization has an incredible power to attract people's attention, consequently it enables users to derive concrete conclusions.

This work aims at finding new approaches of visualizing the CH data to enhance users search experience through semantic search. The aim of semantic technologies is to provide more visually appealing data, by enabling graphical representations of the semantically structured data. Furthermore, it enables meaningful relations of data entities. The meaningful and labeled clustering of data in form of semantic concepts enable new ways to visualize data. As a motivating scenario, we consider the following use case: a user visiting a historical

and tourism attraction site in Valais, uses the City-Stories application to find meaningful relations of data entities about Zermatt. Based on his search, the application recommends to the user the relevant information over Zermatt and it visually displays the most important concepts such as: related locations, people, events and historical sites. In this paper we present an extension of our previous work within the City-Stories project[6]. The rest of this paper is organized as follows: in Section 2 we provide details about the datasets, Section 3 is the main part and it describes the three components: data annotation and linking, semantic linked data search and connected concepts visualization, and Section 4 concludes.



(a) Timeline visualization (b) Network visualization and browsing (c) Top 10 concepts over time

Fig. 1: Screenshots of the front-end application

2 Data Collection and Experimental Setup

Obtaining cultural heritage datasets has a crucial importance for implementing and testing our proposed approach. Hence, in this work we use digital archives datasets obtained from Mediatheque[1] and data collected via the Digital Valais[2] project, summing up to about 45 thousands records. These datasets contain a multi-format content (images, videos, audio and text) including media metadata (title, subject, description, author, date) and span on multi-disciplines such as culture, tradition, archeology, history etc. Part of our previous work focuses on cleaning, integration and alignment [6] of the data. In addition to collecting data from archival and memory institutions, we use data that has been collected via a crowdsourcing application [5], which enables local people to share valuable data and information they have about the cultural heritage of their region.

[1] http://www.mediatheque.ch/
[2] http://www.valais-digital.ch/

3 Semantic Linked Data Search and Network Visualization

Multimedia items in the dataset contain different information in the metadata, including text description, which can be in three languages (mainly French and German, and partially in English). In the first phase, we run the *data annotation and linking*. As there are different languages in the dataset, initially we detect the language used on the text description using the language identification tool[3]. Next, we run the entity linking [3] which automatically annotates mentions of DBpedia[4] resources in the text, and for each entity retrieved, we query the DBpedia SPQARL endpoint[5] to get additional information such as: label name, abstract description, thumbnail and image. The list of extracted entities and external knowledge base data is saved on the dataset repository and this data is the core part for the text search and visualization.

Because the dataset is specific and contains data about a region in Switzerland in local languages, and due to the short available text description in the media metadata, searching and finding relevant information is limited to keywords strict matching. Hence, we add external knowledge base information (if available) in three languages. In this way, we enable the *semantic text search*, i.e. the text search keywords are mapped to the multilanguage linked data available in the dataset.

Since the dataset contains multimedia items mainly about a location, person or event, the data is aggregated based on the title of the items which is more general representation and does not include the specific information. The search results are *timeline visualized*, giving users a better experience on exploring how these places, things and events evolved chronologically. An example of the timeline visualization is depicted in Figure 1a, in which the user is searching for "Carnival of Sierre", and images and videos are visually represented and ordered by time.

Extracted concepts within the text description of the multimedia items are highlighted in distinguishable color. On the same page, users can read more about these concepts: image, description and links to the corresponding Wikipedia and DBpedia webpages are provided. The main part of the visualization is the *network of connected concepts*. The network consists of top k nodes and edges, where the nodes represent the strong concepts related to the chosen concept. An example is illustrated in Figure 1b, where the user is interested to learn more about "Sierre". On the user interface (UI) the network of 10 connected concepts is generated around the chosen concept "Sierre". Size of the circular nodes depends on the importance, i.e. number of occurrences of the concepts with respect to the chosen concept. By clicking on each node, users can read more about other concepts in the network. Additionally, exploring more information about the displayed multimedia is available, allowing users to expand

[3] https://pypi.python.org/pypi/langdetect
[4] http://dbpedia.org
[5] https://dbpedia.org/sparql

and collapse the network. Moreover, the feature of visualizing in a network the trend of connected concepts over period of time is available, enabling users to explore the change of important and relevant things for the chosen place, event or person they are interested in. This is shown in Figure 1c.

4 Conclusion and Future Work

The overall objective of this paper is improving user search experience on CH related data, by understanding user's search context and enabling quick access to the relevant information. This is achieved by retrieving data to generate relevant results as well as visualizing the search results. Our approach based on connected concepts and network visualization, makes the search process easier for the users by showing them the most relevant data on one single UI and it provides better experience when searching for data. Additionally, linking with external knowledge bases, we provide more information about things they are interested in.

In our future work we intend to increase the size of the collections from several archival institutions and touristic locations, and test the performance of the overall system. Moreover, we aim to apply crowdsourcing for improving the accuracy of the semantic connected concepts approach.

Acknowledgement

This work was partly funded by the Hasler Foundation in the context of the project City-Stories. We are grateful to "Mediathque Valais", "Digital Valais Wallis" and "Archives de l'Etat du Valais" for providing the data.

References

1. Liu, Z., Shan, J., Glassey Balet, N., Fang, G.: Semantic social media analysis of chinese tourists in switzerland. Information Technology & Tourism pp. 183–202 (2017)
2. Lynch, C.: Digital collections, digital libraries & the digitization of cultural heritage information. Microform & imaging review pp. 131–145 (2002)
3. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: Dbpedia spotlight: shedding light on the web of documents. In: Proceedings of the 7th international conference on semantic systems. pp. 1–8. ACM (2011)
4. Nazemi, K., Burkhardt, D., Ginters, E., Kohlhammer, J.: Semantics visualization–definition, approaches and challenges. Procedia Computer Science pp. 75–83 (2015)
5. Shabani, S., Sokhn, M.: Gaming as a gateway: Ensuring quality control for crowdsourced data. In: International Conference on Cooperative Design, Visualization and Engineering. pp. 73–76. Springer (2017)
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: Proceedings of the 4th International Workshop on Computational History (HistoInformatics 2017) (2017)