City-Stories: A Multimedia Hybrid Content and Entity Retrieval System for Historical Data

Shaban Shabani, Maria Sokhn
Institute of Information Systems
HES-SO Valais-Wallis, Switzerland
{firstname.lastname}@hevs.ch

Laura Rettig, Philippe Cudré-Mauroux
eXascale Infolab
University of Fribourg, Switzerland
{firstname.lastname}@unifr.ch

Lukas Beck, Claudiu Tănase, Heiko Schuldt
Databases and Information Systems
University of Basel, Switzerland
{firstname.lastname}@unibas.ch

ABSTRACT

Information systems used in tourism rely mostly on up-to-date content on attractive places. In addition, these systems increasingly make use of archived photographs, documents, films, or even ancient paintings and other artwork by integrating such curated content from museums and memory institutions, possibly enriched with user-provided content. Hence the distinction between cultural heritage applications and tourism more and more blurs. Users are not only interested in the current appearance of landscapes, monuments, or buildings, but also in the evolution of these places over time. This requires large multimedia collections which integrate content from several cultural heritage institutions. As a consequence, interactive retrieval systems for historical multimedia are needed that support homogeneous content-based and semantic querying despite the heterogeneity of these collections. In this paper we present *City-Stories*, a multimedia hybrid content and entity retrieval system. *City-Stories* is based on a state-of-the-art open source multimedia retrieval system. Multimedia features in *City-Stories* represent multiple semantic levels: low-level (e.g., color, edge, motion), mid-level (e.g., date, location, objects), and high-level features (e.g., semantic entities, scene category). For the latter, *City-Stories* applies entity recognition and entity linking for identifying semantic concepts and linking objects across media types. Consequently, *City-Stories* supports various types of cross-modal queries. Moreover, *City-Stories* uses a map-based visualization layer that facilitates spatial queries and browsing. Finally, *City-Stories* follows a crowdsourcing approach for content annotation and for enriching curated content with multimedia objects and documents provided by users. The paper shows how the *City-Stories* system seamlessly combines content-based search with entity-based navigation and leverages the wisdom of the crowd for manual annotations.

CCS CONCEPTS

Information Systems → Multimedia and Multimodal Retrieval; *Social tagging systems; Multimedia Databases;* Spatial-temporal Systems; •**Computing Methodologies** → *Semantic Networks;* •**Human-centered Computing** → *Collaborative and Social Computing Systems and Tools;*

KEYWORDS

Content-based Retrieval, Spatio-temporal Querying, Multimedia Databases, Multimodal Interaction, Historical Multimedia, Crowdsourcing, Entity Linking

ACM Reference format:
Shaban Shabani, Maria Sokhn, Laura Rettig, Philippe Cudré-Mauroux, and Lukas Beck, Claudiu Tănase, Heiko Schuldt. 2017. *City-Stories*: A Multimedia Hybrid Content and Entity Retrieval System for Historical Data. In *Proceedings of the 4th International Workshop on Computational History, Singapore, November 2017 (HistoInformatics 2017),* 8 pages.

1 INTRODUCTION

Multimedia data on places of interest like documents, photos, videos, or user ratings are the most important sources used in information systems for tourists. While systems have so far focused on up-to-date content, historical material taken from archives is increasingly gaining importance in order to give tourists more information on how particular touristic sites have developed over time, i.e., how they looked 20, 50, 100, or even more years ago. With the content from museums and memory institutions, the distinction between cultural heritage applications and information systems for tourism increasingly blurs, despite the fact that content may differ significantly (in terms of media types and formats, age, availability and degree of detail of metadata/annotations, etc.). Moreover, content provided by users out of their private archives is also gaining importance due to the proliferation of social networks and crowdsourcing platforms.

In order to provide integrated access to such heterogeneous content, several important technical challenges need to be addressed:

Multimedia Retrieval. The integrated content should be accessible by a very broad range of different query types, such as keyword queries to search in (manual) textual annotations, query-by-example (multimedia search with sample objects), query-by-sketch (multimedia similarity search on the basis of hand-drawn sketches), semantic queries that exploit semantic concepts and links between objects, spatio-temporal queries (i.e., queries on the location and/or time where/when a particular object has been created), and any combination of these modes.

Entity Recognition and Linking. Content coming from different sources, in different formats, and possibly also with different metadata structures has to be integrated to make sure that it can be accessed via a homogeneous interface. This includes standard approaches to schema and data integration, but also more advanced and innovative challenges like entity recognition and entity linking to make sure that links between objects (of the same or even of different media types) can be identified, stored as part of the metadata, enhanced with further external sources, and subsequently exploited for query purposes.

Crowdsourcing. In addition to cultural heritage content curated by archives, user-generated content from private collections is gaining importance in touristic information systems. In order to attract the attention to potential content providers, the awareness of such touristic platforms has to be raised, the technical barrier for contribution has to be lowered, and users have to be encouraged to actively participate. This is not only true for the provision of new content but also for annotations to existing content (e.g., ratings or experience reports). The rapid adoption of smartphones has made it possible to also exploit mobile crowdsourcing [1] as an efficient and easy way of reaching and using human intelligence and machine computation for solving Human Intelligence Tasks (HITs).

In this paper, we introduce *City-Stories*, a novel and innovative system for collecting, managing, and accessing heterogeneous cultural heritage content for touristic applications. The *City-Stories* browser allows to retrieve spatio-temporal knowledge and supports real-time interactive search in large databases of historical multimedia collections. From a systems perspective, *City-Stories* is based on vitrivr[1] [2] which in turn uses the retrieval engine Cineast [3] and the distributed database backend ADAM$_{pro}$ [4]. Information is extracted both from *content* and *metadata* and can be simultaneously queried in both modes. *City-Stories*' content-based search extends the functionality of the vitrivr engine, which allows interactive, efficient multi-feature retrieval in large multimedia collections. Metadata of the multimedia objects (automatically generated or manually added) is used to enrich the description of content in several ways:

- *Low-level features* like color, edge, motion, *mid-level features* like date, location, and *high-level features* like semantic entities or scene categories.
- *Spatio-temporal metadata* in the collection is directly imported as vector location and timestamp features; the *City-Stories* frontend allows for spatial, temporal, and spatio-temporal queries and the results are displayed in a map and on a timeline, respectively.
- *Manual annotations* and user ratings are provided via crowdsourcing.
- *Textual metadata* (usually in the form of title and description of an item) is subjected to entity extraction, which yields Uniform Resource Identifiers (URIs) in a knowledge base. These entity URIs are then used to pre-compute semantic entity distances between collection items.

Hence, *City-Stories* combines content-based search with entity-based navigation and leverages the wisdom of the crowd for manual annotations.

The contribution of the paper is threefold. First, we show how multi-feature content-based similarity search providing a plethora of query types can be enriched by entity recognition and entity linking, thereby allowing for cross-media retrieval based on semantic concepts. Second, we show how curated collections and automatically generated metadata can be extended by user-generated content and user-provided metadata, which is particularly relevant in applications for tourists and which complements manually curated cultural heritage collections. Third, we show how all these elements can be seamlessly combined in the *City-Stories* system.

[1]https://vitrivr.org

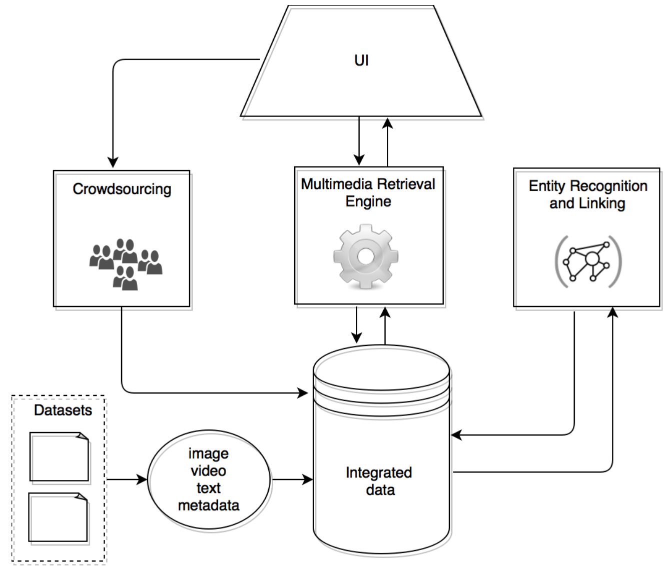


Figure 1: *City-Stories* conceptual architecture.

The remainder of the paper is structured as follows: Section 2 motivates the *City-Stories* approach with a tourism use case. Section 3 discusses the components needed for the integrated multimedia content and entity retrieval system and Section 4 presents details of the *City-Stories* system. Section 5 summarizes related work and Section 6 concludes.

2 MOTIVATION

Consider, as an example for *City-Stories*, the following use case: Sophia, a tourist from Dublin, is visiting the city center of Berlin. When standing in front of the Brandenburg Gate, one of Berlin's neoclassical city gates, she has a variety of questions regarding the building and its neighborhood, like *'What building is this, what was its purpose, when and by whom has it been built?'* or *'How did the neighborhood of the gate look like around 1900, in the so-called 'golden' 1920's, shortly after the end of WW2, in the 1970s, before and after the fall of the iron curtain in 1989 — or at any other point in time in the past?'*.

Sophia holds a smartphone on which she accesses the *City-Stories* query interface. The *City-Stories* system integrates several cultural heritage multimedia collections, e.g., from the German federal archive or from focused museum collections. Moreover, *City-Stories* encompasses a large number of photos and associated metadata provided by local citizens. Using *City-Stories*, Sophia is able to directly browse the content and submit queries of different types:

- *Simple location queries*: Using the GPS coordinates of her current location, Sophia is able to identify the building

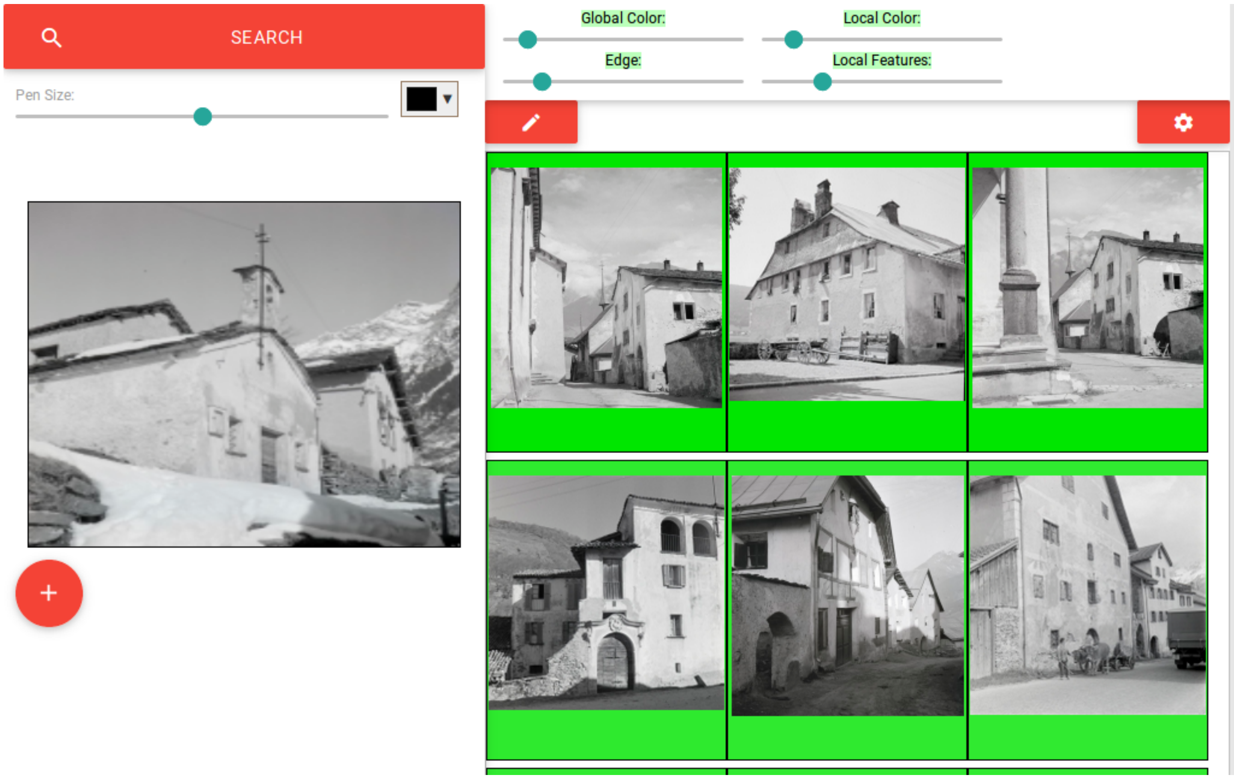


Figure 2: Content-based similarity search in Cultural Heritage content.

she is currently looking at and get access to basic information regarding this building, combined from several data sources on the web.

- *Temporal queries*: On the basis of information from various sources that have been integrated beforehand into the *City-Stories* system, Sophia is able to query details of the building's history (photos or historical paintings from different stages of the building). Moreover, she will also get information on the building's neighborhood at different points in time, on historical events that took place there, and statistical information (e.g., population of the city at different points in time). The latter is based on linked metadata, i.e., metadata enriched with links between objects after entity recognition has been applied to the content.
- *Combined spatial and multimedia similarity queries*: These queries allow to search for similar buildings (or buildings that take a similar role) in the vicinity of Sophia's current location.
- *Multimedia similarity queries*: Sophia takes a photo of the Brandenburg Gate with the camera of her smartphone. She will use this photo for a similarity query in order to find other buildings (in Berlin or in any other European city) that look similar. Here, similarity can be defined either by intrinsic image features, on the base of the object's metadata or links, or any combination of these.
- *Combined sketch-image similarity queries*: Sophia provides one of the photos of the Brandenburg Gate taken with her smartphone as query input and adds a superimposed sketch (e.g., she draws a typical Berlin double-deck coach in the foreground).

Most importantly, Sophia does not want to use an earmarked smartphone app provided by a local tourist organization with manually curated content specialized for a particular touristic site, as she would have to newly install such an app every time she visits another place. Rather, Sophia is interested in *City-Stories*, a generic approach that can be used to integrate and access content independent of a concrete location, so she could use this system for her next trips to Singapore, to visually explore the recent growth of the Marina area, or to New York, for instance to allow her to visualize the development of Lower Manhattan over the past 120 years.

3 CONCEPTS

Figure 1 illustrates the different contributions of the *City-Stories* system and their interplay.

3.1 Spatio-Temporal Retrieval Engine

The multimedia retrieval engine of *City-Stories*, which is based on the vitrivr system, offers multiple different query modes: query-by-example (QbE), query-by-sketch (QbS), temporal, and spatial queries. In a given query we can freely combine these different modes, e.g., by providing an exemplary image, drawing a sketch on top of an existing or a provided image, and/or specifying a location. To provide this functionality, we extend the software stack

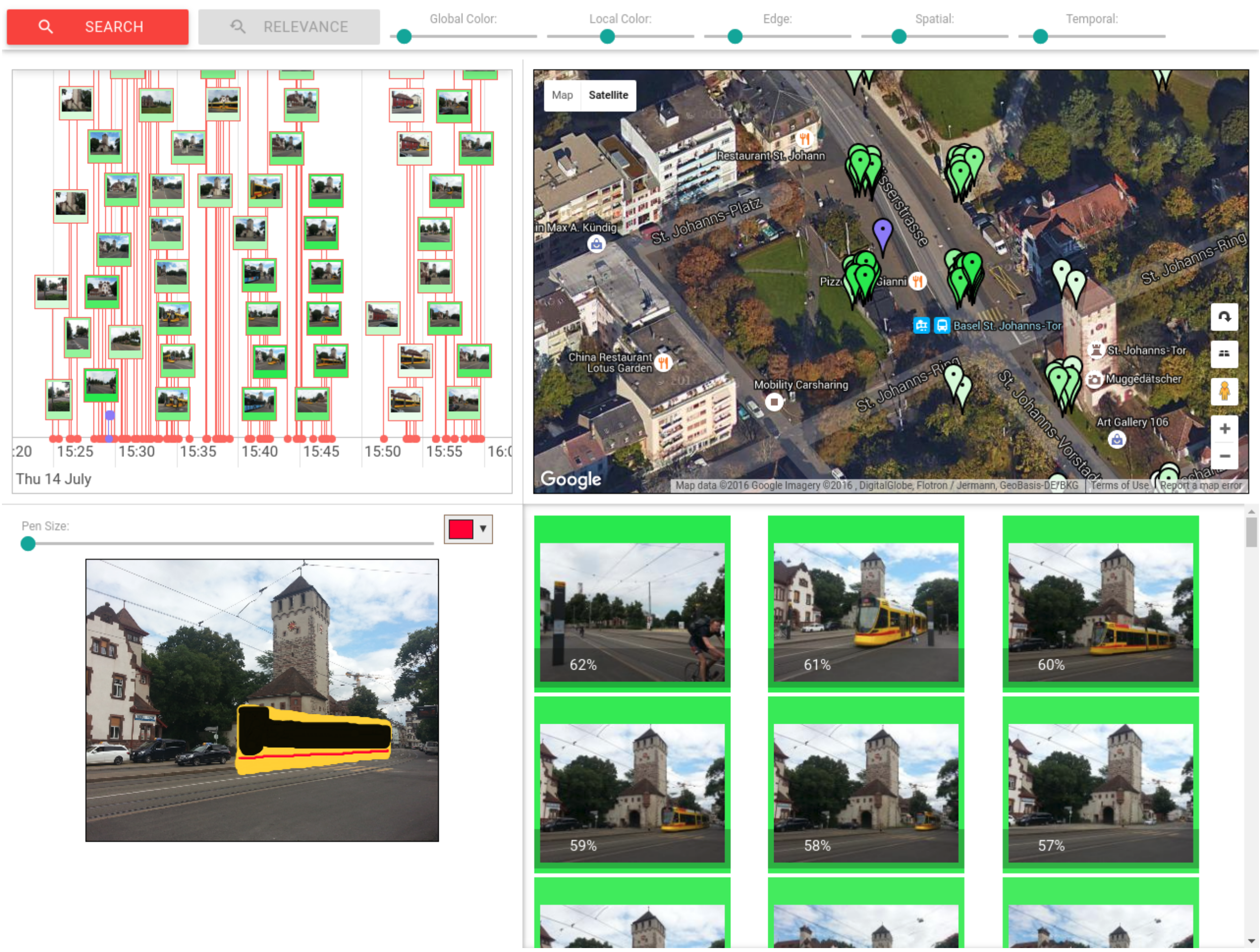


Figure 3: Screenshot of the spatio-temporal content-based retrieval front-end of *City-Stories* [5].

of vitrivr including Cineast, the extraction and retrieval engine, and ADAM$_{pro}$, the database engine.

In Cineast, we differentiate between an *on-line* and an *off-line* phase. The off-line phase includes the feature extraction and the storage of the resulting metadata. In the on-line phase, where the actual retrieval happens, the engine executes a given query and returns a list of documents ordered by similarity.

Cineast uses multiple features in combination to describe a document. In total, there are already 40 content descriptors for videos and images provided by Cineast. These descriptors extract mainly color, edge, and motion information (where applicable). Building on top of that, we extend Cineast with the following higher level descriptors:

- *Spatial and temporal similarity features* by using geolocalization and timestamp metadata provided by the content (usually via the capturing devices).
- A *descriptor using semantic entities* that are provided by the entity recognition described in what follows in Section 3.2.
- *Semantic concept features* provided by a deep neural network.

3.2 Entity Recognition and Linking

In order to integrate the archive data available to us with other data sources such as knowledge bases, we apply named-entity recognition, candidate selection, and entity linking techniques on available text data. Named-entity recognition is the task of identifying mentions of entities, which can take various surface forms, in text. Once an entity mention has been extracted, the corresponding URI in the knowledge base has to be found. Selecting the URI corresponding to a mention of an entity in text is called *entity linking*.

By linking to a knowledge base, specifically to DBpedia[2] [6], we enhance our data with relevant information on the entity in the knowledge base and can link data from further sources to the same entities for integration.

In order to identify entities associated with media items, we use textual media metadata and extract entities from associated titles and descriptions. Consequently, when displaying media items, additional information from the knowledge base can be retrieved and displayed. Knowledge bases tend to offer various attributes for entities, many of which will not be relevant to the viewer. Thus,

[2] http://dbpedia.org

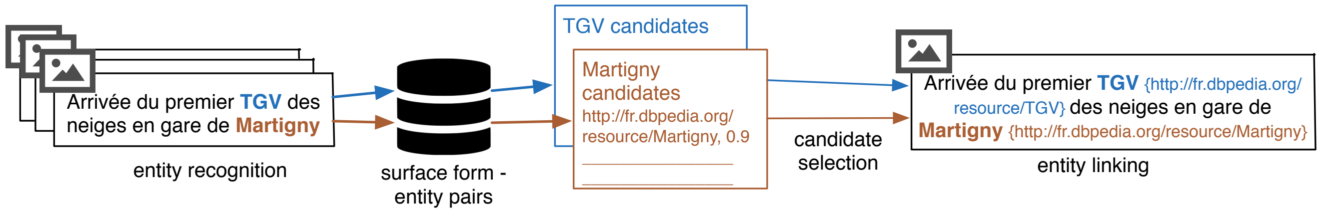


Figure 4: Entity recognition and linking process.

when choosing which information to display, we also rank the attributes by their importance to the viewer (for example, when viewing information on a city, the population and the founding year will likely be more relevant to a tourist than the ZIP codes in this city).

Furthermore, we are able to leverage the relationships between media items by extracting the relationships between entities in a graph-structured knowledge base. This also allows to transitively extract related information, e.g., for entities with little available data. Oftentimes, in such graph structures, we can rely on higher-level categories to provide general information on an entity for which specific information may not be available.

3.3 Crowdsourcing

The crowdsourcing component provides the possibility of enriching content with new data provided by different types of users and enhancing the integrated digital collection data.

Collected data is not always complete, i.e., it could be noisy and/or miss particular information. In a tourist application, for instance, information provided by users could miss the location where a point of cultural interest can be found, to whom it belongs, or even the title or a description. Likewise, data might be incorrect or conflicting as a result of the integration of collected datasets coming from different sources. Hence, these challenges are grouped into two categories: *conflicting information* and *missing information*, both of which are addressed in *City-Stories* using crowdsourcing.

Crowdsourcing is able to build an open, connected, and smart cultural heritage with involved consumers and providers [7]. The crowdsourcing service built into *City-Stories* enables volunteers to engage in order to share new data, as well as complete the existing data and improve data quality.

In order to optimize the assignment of tasks to the appropriate crowds, we make use of *push* crowdsourcing [8]. In contrast to standard *pull* crowdsourcing, where workers pull the tasks at random, push crowdsourcing is oriented towards modeling tasks based on users' profiles and pushed to them. At first, users specify their interests and topics they have knowledge on. Then, by matching users' profiles with the available HITs, *City-Stories* recommends tasks to the best matched users based on their interests and skills.

In paid micro crowdsourcing scenarios, money as incentive is the main motivator for workers to contribute and can also be used for quality control (e.g., constrain a payment on the quality of the work that has been done). However, it is important to distinguish malicious users from workers that do not have the intention of fraud but may affect the overall system due to misunderstandings or lack of knowledge and experience with the platform [9, 10]. As the latter scenario is more likely to happen in our case, we apply two quality control mechanisms to evaluate the quality of volunteers' work in *City-Stories*: *play cards* and a *weighted majority voting with reputation system*.

4 *CITY-STORIES* SYSTEM

In what follows, we describe details on the implementation of the different components of *City-Stories* and the content that has been integrated in a first prototype.

4.1 Spatio-temporal Multimedia Browser

When querying the retrieval engine via the *City-Stories* UI, for each given feature, Cineast performs an extraction on the given query document resulting in a feature vector. Each feature vector is passed to the ADAM_{pro} database backend to perform a k-nearest neighbor (k-NN) similarity search to find a list of similar documents. ADAM_{pro} has been shown to scale to collection sizes of up to 50 million entries and feature vectors with up to 500 dimensions [4]. After receiving a list of documents ranked by similarity for each feature, the Cineast retrieval engine merges these results into one result list (depicted in Figure 2).

In particular, for the spatial and temporal similarity search we use a nearest neighbor query on the two-dimensional geolocation data and the one-dimensional timestamp data, respectively (see Figure 3).

To search for similar documents based on semantic properties, we provide two different kinds of features:

- Based on entity recognition and linking, each document is characterized by a *list of semantic entities*. Using this list of semantic entities we calculate a pairwise distance to estimate the similarity between two documents.
- The similarity based on *semantic concepts* utilizes an AlexNet convolutional network [11]. Using the output of the last fully connected layer, fc7, we obtain a 4096-dimensional feature vector for a given image that can be used in a k-NN search. We provide two different features by training the network on different datasets. The training data from the Places2 dataset [12] equates to a feature focusing on scene and environment similarity. Compared to Places2, the data from the MS COCO Detection challenge [13] provides a feature focusing on object similarity inside a scene.

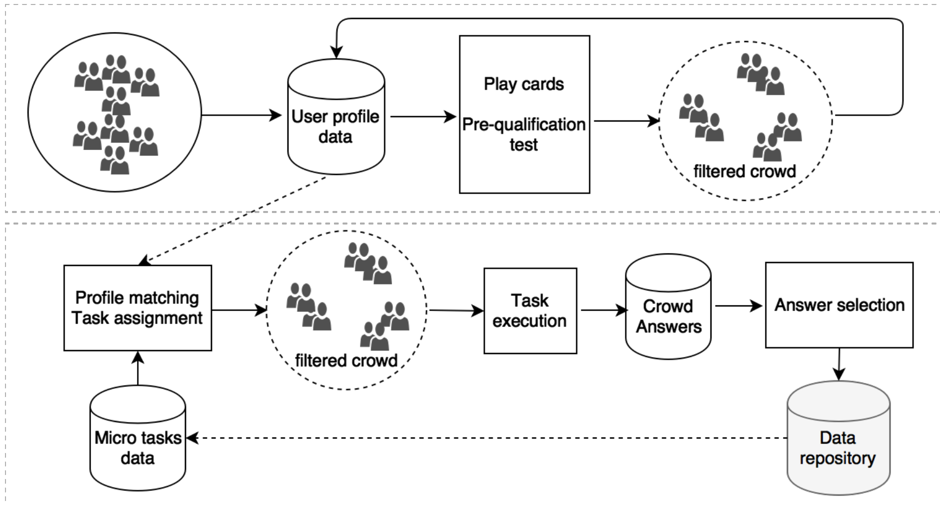


Figure 5: Crowdsourcing quality control and task management process.

4.2 Entity Recognition and Linking

The data integration component of the *City-Stories* system focuses on leveraging textual data to integrate different sources and schemata. In the implementation, we use metadata provided with the media items in the DigitalValais[3] dataset, specifically, title and description corresponding to each item, to extract and link related entities.

The implementation consists of three steps (see Figure 4):

(1) *Named-entity recognition*: Using the Stanford Named Entity Recognizer (NER) [14], we extract likely mentions of entities from the full text, which may be in different languages (German, French or any other local language).
(2) *Candidate selection*: This step consists in choosing a set of candidate entities that may correspond to the extracted surface form. We choose a set of candidate entities for each extracted mention in the text.
(3) *Entity linking*: We rank the candidates and choose the best matching entity for each mention, then add the DBpedia URI to the media item in the *City-Stories* database.

Candidates are selected from a database of pairs of surface forms and entities (given as DBpedia URIs). This database is created by processing all Wikipedia articles and extracting hyperlinks where the hyperlink text corresponds to the observed surface form (i.e., the link text) and the link location corresponds to the entity this surface form links to (i.e., the linked article). The frequency of observing a specific *(surface form, entity)* pair yields a prior probability that a particular surface form corresponds to the linked entity. We rank the candidates using this prior score and select the top candidate to link this mention to an entity in the knowledge base.

Having done the linking, we are able to create a graph of the media items where an edge is present between two items if they contain the same entity and are thus related. Knowing the DBpedia URI of an entity found in the metadata of a media item, we extract relevant attributes for this entity by ranking the attributes with their frequency of appearing in close proximity to this type of entity.

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

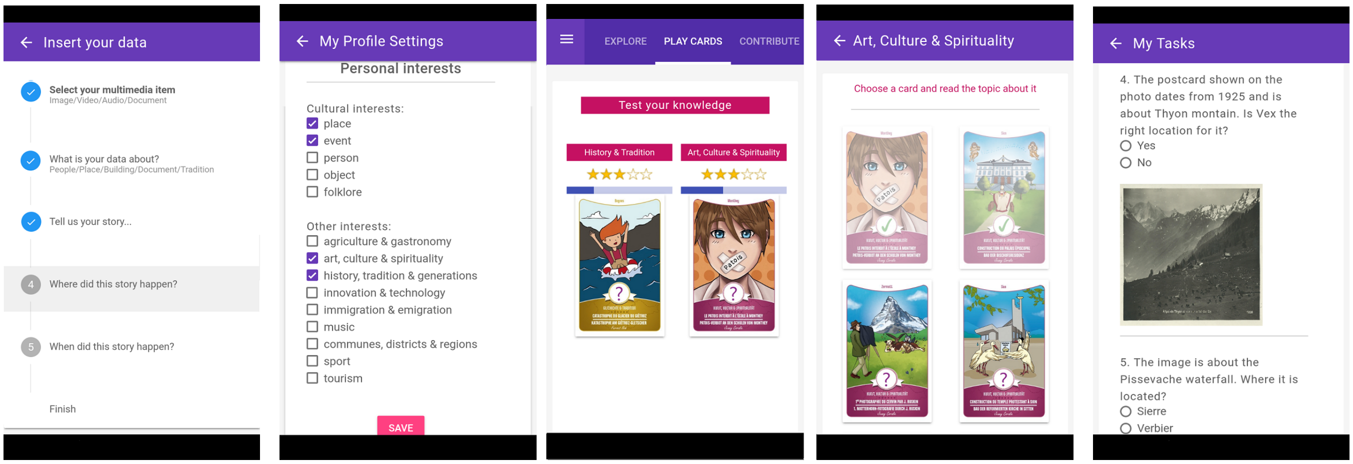
4.3 Crowdsourcing

In parallel to collecting data from institutions such as audio/visual archives, Mediatheque[4], and DigitalValais, we emphasize the importance of data sharing from people that have valuable data and information about cultural heritage in private collections. This part of the system enables users to participate and contribute to cultural heritage. Once shared, users' data is integrated to the data repository. In order to maintain a high level of quality of the crowdsourced data, we apply the following control methods (shown in Figure 5):

Play cards is a test measure used to qualify or disqualify users for solving tasks of a certain category. We use 195 playing cards grouped in 13 different categories, where categories represent subject areas of the crowdsourcing tasks. Each card contains both a question and its answer (not visible to the user). At first users provide information on their topics of interest which intersect with card categories. Then they are forwarded to the test phase. To get qualified, they have to correctly answer at least 70% of the questions matched to their interests. Upon two consecutive failures, a user is no longer considered for tasks on these specific topics. Providing the option to choose topics they like or have knowledge on avoids false positives, i.e., eliminating users due to lack of knowledge on randomly assigned questions coming from a pool of predefined questions. Moreover, implementing this measure in a game fashion increases the interactivity and the interest of the users. Additionally, users can test their knowledge on topics covered by the cards and at the same time expand their knowledge with additionally provided information by these cards.

A weighted majority voting with reputation system combines majority voting (MV) with users' reputation scores. MV [15] as a quality control mechanism assigns the same task to multiple users and aggregates the results to choose the right answer. On the other hand, the reputation strategy allows to track users' performance

[4]http://www.mediatheque.ch/



(a) Data sharing (b) Play cards (c) Annotation task sample

Figure 6: Screenshot of the crowdsourcing front-end of *City-Stories*

during crowdsourcing tasks and is complementary to majority voting. The users' results from the play cards qualification tests are assigned as initial reputation scores. These scores are later used as weights during the aggregation of the answers, i.e. an answer from a user with high reputation has higher weight. After each aggregation phase, the users' reputation scores are updated by considering their outcomes on that task.

A screenshot of the crowdsourcing frontend of *City-Stories* showing how data is shared, the play cards for quality control, and sample annotations is depicted in Figure 6.

5 RELATED WORK

In general, existing retrieval systems applied to the cultural heritage domain are either metadata-based or content-based, and few of them implement both [16]. Metadata-based systems focus on keyword-based queries and linking of digital objects with external data sources, whereas content-based systems focus on query-by-example, query-by-specification and browsing.

EUscreen[5] is a project related to multimedia archives that focuses on the collection, integration, and publication of audio-visual content. Oomen et al. [17, 18] created a European television archive using data from many different TV broadcasters. It provides an interface with keyword search. Media in Context[6] is a platform for cross-media extraction (via pipelined extractors), analysis, metadata publishing, and querying. Within this project lies Sensefy [19], a multimedia search and information retrieval system, that provides metadata keyword search and object linking with real world entities and concepts. Otegi et al. [20] introduced Personalized PageRank, a tool for generating personalized recommendations in a cultural heritage collection. They use metadata, session logs from users, and Wikipedia as an external sources to elicit recommendations. INVENiT [21] is a semantic search system used for cultural heritage collections, making use of links between objects and terms provided by structured vocabularies. Moreover, users can contribute by annotating collection objects. Their system lacks a trust assessment which denotes a key issue for data quality. SCULPTEUR [22] is a multimedia retrieval system for searching digital collections in museums. It features content-based image retrieval as well as semantic retrieval using metadata and a semantic layer. An extension of this work [23] provides a hybrid model for cultural heritage collections, combining the two retrieval methods for image search.

Named-entity recognition is based largely on natural language processing techniques and is required as a step prior to performing candidate selection and named-entity disambiguation. The Stanford NER [14] is trained by combining a constraint model with a sequence model for the purpose of extracting information from text, including named entities. Prokofyev et al. [24] employ n-gram based features for NER and demonstrate their accuracy in idiosyncratic domains, which could also be applied to the domain of historical archive data. SANAPHOR [25] introduces the use of a type system on top of recognized entities. The relatedness of types is then used to identify co-references referring to the same entity, and to link these identified mentions of an entity to a DBpedia URI. DBpedia Spotlight [26] provides the entire pipeline from NER over candidate selection, i.e., retrieving candidate entities that may correspond to the extracted surface form, to entity linking. Their work has focused on the implementation of a usable system for multi-lingual entity extraction and linking.

Crowdsourcing has shown to be an effective solution for problems that are difficult to solve for computers and problems that require human intelligence [27]. Its popularity has grown due to online platforms such as Amazon Mechanical Turk[7] and CrowdFlower[8], which allow crowds to participate in solving paid microtasks. Concerning quality control, a broadly used quality checker is the gold questions technique [28], a test measure to qualify users for solving tasks. However, this method alone leads to the elimination of honest workers who lack some knowledge. Aggregation

[5]http://euscreen.eu
[6]http://mico-project.eu
[7]https://www.mturk.com/mturk/
[8]https://www.crowdflower.com/

methods also known as voting strategy aim to avoid biased workers. Majority voting [15] is a redundancy mechanism that is widely applied to prevent spammers and lazy workers.

6 CONCLUSION

In this paper, we have presented *City-Stories*, a novel system that combines content-based similarity search on both historical and contemporary multimedia data with spatio-temporal queries and that exploits semantic analysis of the content for entity recognition and linking. We have deployed *City-Stories* in a tourist application where cultural heritage content from archives and memory institutions is complemented with content contributed by users via a crowdsourcing approach.

In our future work, we aim to increase the size of the collections available in *City-Stories* for several selected tourist locations to show the generic applicability of the *City-Stories* approach. We also intend to perform user studies at these locations to assess the usability of the system and the effectiveness of the integrated content and entity retrieval approach. Moreover, we plan to further exploit the synergies obtained from the combination of all retrieval modes supported in *City-Stories*, for instance by proactively making recommendations during the retrieval process and by providing additional information from external sources. Finally, we aim to extend the user experience by providing within the user interface an overlay function that allows to superimpose the camera view of a smartphone showing the current view of a place of interest with historical content in order to better show the development of a particular place. When multiple visual objects of a place are available from different periods of time (taken from the same or at least a similar perspective), this will lead to a "history browser" which can be used to steer the overlay with a slider on the timeline, to select the object chosen for the overlay, and to gradually visualize the development.

ACKNOWLEDGMENT

This work was partly funded by the *Hasler Foundation* in the context of the project *City-Stories*. We would like to thank the cantonal archives and the "Mediathèque" of the canton of Valais and the team of Digital Valais project for delivering a data testbed.

REFERENCES

[1] J. Ren, Y. Zhang, K. Zhang, and X. Shen, "Exploiting mobile crowdsourcing for pervasive cloud services: challenges and solutions," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 98–105, March 2015.

[2] L. Rossetto, I. Giangreco, C. Tănase, and H. Schuldt, "vitrivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections," in *Proceedings of the 2016 ACM Conference on Multimedia Conference (ACM MM 2016)*, Amsterdam, The Netherlands, Oct. 2016, pp. 1183–1186.

[3] L. Rossetto, I. Giangreco, and H. Schuldt, "Cineast: A Multi-feature Sketch-Based Video Retrieval Engine," in *Proc. of the International Symposium on Multimedia (ISM 2014).* Taichung, Taiwan: IEEE Computer Society, Dec. 2014, pp. 18–23.

[4] I. Giangreco and H. Schuldt, "ADAM$_{pro}$: Database Support for Big Multimedia Retrieval," *Datenbank-Spektrum*, vol. 16, no. 1, pp. 17–26, 2016.

[5] L. Beck and H. Schuldt, "City-Stories: A Spatio-Temporal Mobile Multimedia Search System," in *Proc. of the IEEE International Symposium on Multimedia*, San Jose, CA, USA, Dec. 2016, pp. 193–196.

[6] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia," *Semantic Web Journal*, vol. 6, no. 2, pp. 167–195, 2015.

[7] J. Oomen and L. Aroyo, "Crowdsourcing in the cultural heritage domain: opportunities and challenges," in *Proceedings of the 5th International Conference on Communities and Technologies (C&T'11).* Brisbane, Australia: ACM, 2011, pp. 138–149.

[8] D. E. Difallah, G. Demartini, and P. Cudré-Mauroux, "Pick-a-crowd: Tell Me What You Like, and I'll Tell You What to Do," in *Proc. of the 22nd Int'l World Wide Web Conf. (WWW'13).* Rio de Janeiro, Brazil: ACM, May 2013, pp. 367–374.

[9] H. Li, B. Yu, and D. Zhou, "Error Rate Bounds in Crowdsourcing Models," *ArXiv e-prints*, Jul. 2013.

[10] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality Management on Amazon Mechanical Turk," in *Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP '10).* Washington DC, USA: ACM, 2010, pp. 64–67.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Curran Associates, 2012, pp. 1097–1105.

[12] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, and A. Oliva, "Places: An image database for deep scene understanding," *arXiv preprint arXiv:1610.02055*, 2016.

[13] T. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, in *Proceedings of the 13th European Conference on Computer Vision (ECCV 2014)*, ser. Lecture Notes in Computer Science, vol. 8693. Zurich, Switzerland: Springer, Sep. 2014, pp. 740–755.

[14] J. R. Finkel, T. Grenager, and C. D. Manning, "Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling," in *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, Michigan, USA, 2005, pp. 363–370.

[15] A. Sorokin and D. A. Forsyth, "Utility data annotation with Amazon Mechanical Turk," in *CVPR Workshops*, Anchorage, AK, USA, June 2008, pp. 1–8.

[16] C.-F. Tsai, "A review of image retrieval methods for digital cultural heritage resources," *Online Information Review*, vol. 31, no. 2, pp. 185–198, 2007.

[17] J. Oomen and V. Tzouvaras, "Publishing Europe's Television Heritage on the Web: The EUscreen Project," in *Proc. of the 1st International Workshop on Multimedia for Cultural Heritage (MM4CH 2011)*, ser. Communications in Computer & Information Science, vol. 247. Modena, Italy: Springer, 2012, pp. 136–142.

[18] J. Oomen, E. Verbruggen, V. Tzouvaras, and K. Hyyppä, "Television heritage linked and visualized: The EUscreen virtual exhibitions and the Linked Open Data pilot," in *Proceedings of the 2013 Digital Heritage International Congress (DigitalHeritage 2013)*, vol. 2. Marseille, France: IEEE, 2013.

[19] C. Perera and D. Jayakody, "Cross media entity and concept driven search," in *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems (SEMANTiCS'16) and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS'16)*, ser. CEUR Workshop Proceedings, vol. 1695. Leipzig, Germany: CEUR-WS.org, Sep. 2016.

[20] A. Otegi, E. Agirre, and P. Clough, "Personalised PageRank for Making Recommendations in Digital Cultural Heritage Collections," in *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '14).* London, United Kingdom: IEEE Computer Society, Sep. 2014, pp. 49–52.

[21] C. Dijkshoorn, J. van Ossenbruggen, L. Aroyo, and G. Schreiber, "INVENiT: Exploring Cultural Heritage Collections While Adding Annotations," in *Proceedings of the 3rd International Conference on Intelligent Exploration of Semantic Data (IESD 2014)*, ser. CEUR Workshop Proceedings, vol. 1279. Riva del Garda, Italy: CEUR-WS.org, Oct. 2014, pp. 95–99.

[22] S. Goodall, P. H. Lewis, K. Martinez, P. A. S. Sinclair, F. Giorgini, M. Addis, M. J. Boniface, C. Lahanier, and J. Stevenson, "SCULPTEUR: multimedia retrieval for museums," in *Proceedings of the 3rd International Conference on Image and Video Retrieval (CIVR 2004)*, ser. Lecture Notes in Computer Science, vol. 3115. Dublin, Ireland: Springer, Jul. 2004, pp. 638–646.

[23] S. Vrochidis, C. Doulaverakis, A. Gounaris, E. Nidelkou, L. Makris, and I. Kompatsiaris, "A hybrid ontology and visual-based retrieval model for cultural heritage multimedia collections," *International Journal of Metadata, Semantics and Ontologies (IJMSO)*, vol. 3, no. 3, pp. 167–182, Feb. 2008.

[24] R. Prokofyev, G. Demartini, and P. Cudré-Mauroux, "Effective named entity recognition for idiosyncratic web collections," in *Proceedings of the 23rd International Conference on World Wide Web (WWW '14).* Seoul, Republic of Korea: ACM, Apr. 2014, pp. 397–408.

[25] R. Prokofyev, A. Tonon, M. Luggen, L. Vouilloz, D. E. Difallah, and P. Cudré-Mauroux, "SANAPHOR: ontology-based coreference resolution," in *Proceedings of the 14th International Conference on The Semantic Web (ISWC 2015)*, vol. 9366. Bethlehem, PA, USA: Springer, Oct. 2015, pp. 458–473.

[26] J. Daiber, M. Jakob, C. Hokamp, and P. N. Mendes, "Improving Efficiency and Accuracy in Multilingual Entity Extraction," in *Proceedings of the 9th International Conference on Semantic Systems.* Graz, Austria: ACM, Sep. 2013, pp. 121–124.

[27] G. Li, J. Wang, Y. Zheng, and M. J. Franklin, "Crowdsourced Data Management: A Survey," *IEEE TKDE*, vol. 28, no. 9, pp. 2296–2319, Sept 2016.

[28] D. Oleson, A. Sorokin, G. P. Laughlin, V. Hester, J. Le, and L. Biewald, "Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing," in *Proceedings of the AAAI Workshop on Human Computation (AAAIWS'11)*, San Francisco, CA, USA, Aug. 2011, pp. 43–48.