METADATA GENERATION FOR CULTURAL HERITAGE: *CREATIVE HISTORIES – THE JOSEFSPLATZ EXPERIENCE*

Brigitte Krenn[1], Gregor Sieber[1], Hans Petschar[2]

Abstract

Creative Histories is a Cultural Heritage application using a full 3D-navigation model for PC and mobile phones to navigate historical sites across different epochs, thus creating a new conceptual model for presenting and navigating spatio-temporally anchored cultural information. In this contribution we present the system's metadata generation component. The major characteristics of which are: automatic extraction of presentation objects from legacy documents, storage of presentation objects in RDF/XML format along media type and semantic dimension, and online generation of interactive tours adapted to requirements of the device, user interests, user position in the spatial and temporal model, and user history.

1. Introduction

In this paper we present a metadata generation component that is part of the so-called *Creative Histories – The Josefsplatz Experience*, a research project named after the application scenario employed for system development. More information on the project can be found under http://www.vrvis.at/research/projects/josefsplatz/. The Josefsplatz is a site of extraordinary cultural importance in Vienna, and accordingly an area of high relevance for tourists.

The Josefsplatz is a city-centre public square in the city of Vienna, centred on a full-sized statue and monument of Joseph II, Holy Roman Emperor mounted on a horse (1795-1806, by sculptor Franz Anton Zauner). The square is regarded as one of the most beautiful examples of urban architecture enclosed by various buildings once part of Hofburg Imperial Palace complex, notably the national library (*Österreichische Nationalbibliothek*). The Library is a fine example of late baroque architecture on a grand scale, designed by Fischer von Erlach the Elder and Fischer von Erlach the Younger (1723 – 1726). The State Hall of the Library (Prunksaal) holds today 200.000 rare books and is open to the public. In 1767 two side wings were added designed by Nikolaus Picassi, the right one covering the Redoutensäle (badly damaged by a grand fire in 1992), the left one integrating the Augustinerkirche. Across the street the Palais Pallavicini and the Palais Pallfy complete the square.

---

[1] Österreichsches Forschungsintitut für Artificial Intelligence (OFAI), Freyung 6, 1010 Wien; {brigitte.krenn,gregor.sieber}@ofai.at

[2] Österreichsche Nationalbibliothek, Josefsplatz 1, 1010 Wien; hans.petschar@onb.ac.at

Creative Histories allows a complex 3D model of an (urban) environment (the Josefsplatz) to be reconstructed from contemporary and historical photos, historical pictures and paintings. This 4-dimensional information space (i.e. the 3D geometry in its temporal variation) is available to the user both on the PC and on mobile devices. In other words, the system enables the user either to immerse in a virtual 4D world through navigation on the PC, or to physically visit the respective site and let her mobile phone take her on a time travel through a historically changing environment. In addition to the visualisation of the architectural space, the user is provided with adaptive meta-information fused into the 4D environment. Thus the system creates a 5D information space. In our approach to generation of multi-modal, adaptive, interactive information presentation, the 3D-model of the architectural environment and its temporal dimension is the pivot of the multimedia presentation of the meta information. The user navigates in a multi-dimensional environment where different strategies for presentation of metadata apply, depending on the previous context, the availability of data, and the kind of interaction device. The latter sets technical restrictions, such as available screen size and resolution, computational power, and bandwidth, and also influences the interaction modalities. For instance, due to a small screen on the mobile presentations employing (synthesized) speech will occur more frequently than presentations incorporating text.

Creative Histories stands in a broad context of systems and projects related to Cultural Heritage. It brings together techniques from 3D reconstruction and modelling, as well as data representation and storage, and from the semantic web. The ARCHEOGUIDE system [7], for instance, is comparable to *Creative Histories* in so far as it aims at a 3D representation of historical sites (ancient Olympia in Greece) and its audio-visual presentation to the visitor. While ARCHEOGUIDE requires a portable computer and Head Mounted Display, our system is designed and implemented for standard devices such as PC and mobile phones. A data retrieval concept for mobile devices similar to that employed in *Creative Histories* can be found in the CHI system [2]. In both systems the data is stored in RDF triples, thus allowing queries on semantic concepts. Other examples of presentation generation based on RDF data are given in [3], [1]. While the focus of the former lies on semantic inferencing to identify metadata for presentation, our approach focuses on automatic, user adaptive tour generation from minimal presentation units. Thus *Creative Histories* further differs from systems such as ARCHEOGUIDE and CHI where tours are typically predefined.

To set the context for the presentation of our proposals for metadata extraction on the one hand, and for generation of multimedia presentations fused into the architectural model on the other hand, we give a brief illustration of the Josefsplatz scenario in section 2. How the data are acquired from legacy documents, and how the information is stored in a database will be discussed in section 3. The model for tour generation will be presented in section 4.

2. The Josefsplatz Application

An interactive visit of the Josefsplatz and the Austrian National Library State Hall can be seen as an opportunity to travel back in history. Starting from the present view of the Josefsplatz a user experiences the history of the library, the changes in architecture, music and literature etc. related to the place. Especially the State Hall provides a sensual experience of the world of Baroque. A visitor seeks her own interactive way on her tour or follows guided tours presenting the history of the architecture, the statues, the iconoghraphic program of the paintings and the imperial perspective of the Habsburg emperor Charles VI, who ruled from 1711 to 1740 and whose statue in the center of the State Hall welcomes thousands of visitors and tourists.

The screenshots in Figure 1 show some aspects of the architectural model. From left to right you find a bird’s eye view of the Josefsplatz area, a view on the Josefsplatz and on the State Hall. The

user can freely navigate in the model. Cf. [6] for a description of the reconstruction technology applied.

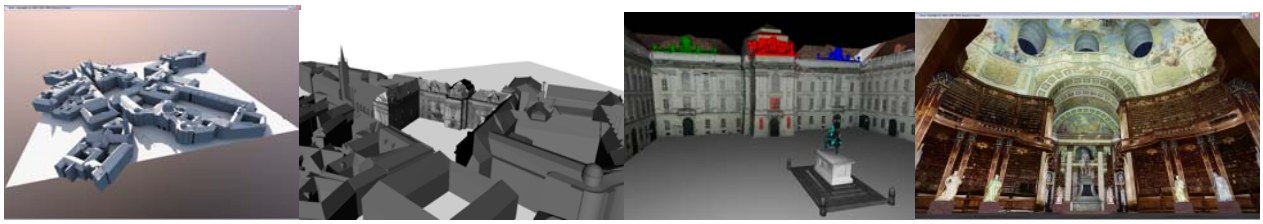

Figure 1: Screen shots from the architectural model

The mobile application enables users to walk freely on the Josefsplatz and in the State Hall using the mobile as a pointer to areas of interest and as a window into history. For a description of the mobile Josefsplatz viewer see [5]. The mobile interface is illustrated in Figure 2. The pictures from left to right show the introduction page, the menu from which the user selects categories of interest on her tour, and a view of the 3D model of the Josefsplatz.

Figure 2: Mobile application

3. Metadata: Creation and Storage

As a first step the available data for the application scenario are surveyed. The data relevant for 3D architectural modelling and visualization includes construction maps of the relevant buildings in their different appearance over time, historical and current photographs, painted views from different viewpoints, etc. The textual and multimedia data of the content to be anchored in the spatial and time dimensions are collected and organized along thematic tiers. An XML-based framework is specified and implemented that allows for a unified representation of all the relevant aspects of the data. To facilitate ontology extension and querying the metadata are stored in RDF format.

Metadata are created in two stages: 1) information is automatically extracted from the legacy documents that have been identified as relevant for the application scenario by a human expert; 2) the thus extracted data are further processed to be suitable for use in PC and mobile applications. The initial data were taken from various digital sources and holdings of the Austrian National Library. Text and HTML documents from the Library-website (www.onb.ac.at), documentations of virtual exhibitions as well as the vast material of a multimedia encyclopedia presenting the treasures and the history of the Austrian National Library have been selected and prepared for integration.

3.1 Metadata Extraction from Legacy Documents

For the extraction of metadata, structural information present in the documents (e.g. HTML mark-up) is used. In addition, some user defined mark-up in the source is taken into account, e.g. citations. The differentiation of quotes from other text is of interest especially for speech synthesis, where different voices are employed for citation and running text.

In the document parser, the structural information is mapped to corresponding slots in the database. The parsing of documents is dependent on recurring structure in the source documents and has to be manually adjusted to changes in document structure or to different documents. Content items that are grouped in the source documents, such as a text with images and sound files, are stored as individual presentation objects and connected by links in the database. This preserves information present in the source while allowing, through further processing, the discovery of new links and creation of new groupings with other objects in the database. Textual data is stored sentencewise for better results in speech synthesis and easier manipulation during the generation of tours. As a result of metadata extraction from legacy documents the database is filled with minimal units for presentation. The presentation objects are further automatically enriched with information relevant for presentation on the PC and on the mobile.

3.2 Enriching the presentation objects

Especially for presentation on the mobile phone, a number of additional information is required to be stored together with the presentation object to ensure the adaptiveness to the device. For instance, due to bandwidth restrictions it is crucial to know the file sizes and the scalability range of sound files, videos and pictures so that the tour generation component can decide whether an object is suitable for presentation on the mobile phone. For photos camera parameters are stored for perspectively correct positioning in the 3D environment.

To overcome the severe restrictions on the amount of text presented on the mobile, textual presentation objects are synthesized employing state of the art text-to-speech (TTS) synthesis tools. We have both experimented with a unit selection and diphone synthesis tool.[3] While the results of unit selection sound more naturally than diphone synthesis, the advantage of diphone synthesis is that the sound quality stays fairly constant, because of the higher flexibility for the manipulation of sound inherent to the technology, whereas unit selection strongly deteriorates in quality when similar units are not available in the sound database. For CH applications this is an issue, particularly as long as unit selection databases are typically geared towards the lexicon of contemporary common language and not towards the domains relevant for CH. For both technologies, unit selection and diphone synthesis, manual creation of additional domain-specific lexica is necessary. These lexica contain grapheme to phoneme transcriptions of those words that were not pronounced properly employing the TTS internal letter to sound conversion. For the current version of our Josefsplatz scenario a domain lexicon of about 100 words has been created. Even though both TTS tools employ SAMPA[4] and it is fairly easy to convert the lexicon formats, it is not possible to use the same lexicon with the two tools. Due to differences in the grapheme to

[3] As for unit selection synthesis we have had access to a demo version of the Loquendo TTS (German and English, http://www.loquendo.com). For diphone synthesis we have employed the open source MARY system (German, http://mary.dfki.de, [4]).

[4] http://www.phon.ucl.ac.uk/home/sampa/index.html

phoneme components, different words need to be covered in the lexica. Moreover the two synthesizers interpret the phonetic transcriptions differently. For example this may be due to different duration control and sound inventory.

3.3 Data Scheme and Storage

In the current version (Figure 3), the data scheme is based on structuring criteria that are imposed by the thematic tiers identified as relevant for the Josefsplatz, and the possibilities of user navigation in the 4D environment. This leads to the following core dimensions according to which each presentation object is structured: category and relevance, spatial anchor in the 3D architectural model, and temporal anchor in the 4D model.

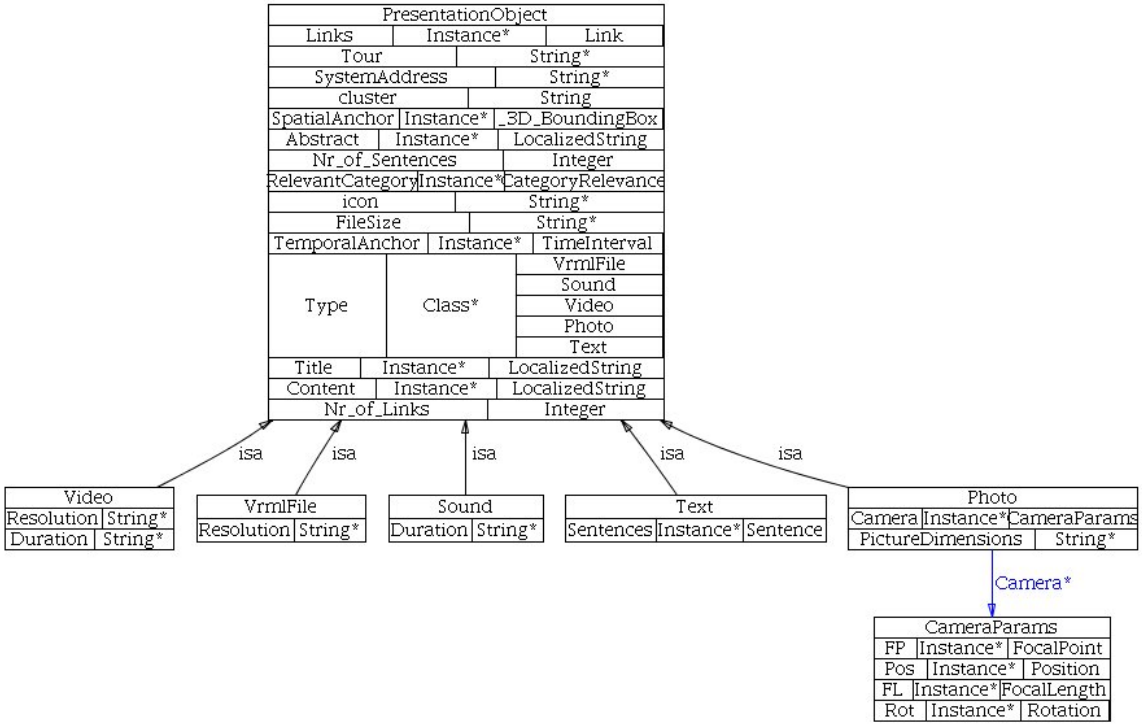

Figure 3: Data scheme for presentation objects

Examples for categories are history of the National Library, history of the Hofmusikkapelle (royal orchestra), music, architecture, society, the Prunksaal (State Hall) and the such. In other words, the categories are application specific. Relevance on the contrary is considered as a general dimension that is introduced to enable differentiated presentations of meta-information suitable for laymen and tourists on the one hand, up to researchers on the other hand. Currently our data covers the tourist-level domain with data from the Austrian National Library for other target groups to be added in the future. The bounding box identifies the spatial anchor in the 3D model, i.e. depending on the position and orientation of the user in the virtual space, different areas are in focus. Moreover, each presentation object is of a certain media type (video, sound, picture, …) and has its media type specific descriptions, such as resolution, duration, camera parameters, etc. Currently we distinguish two kinds of links: a) links that group content items that were already related in the legacy document; b) links that establish semantic relations, either between presentation objects or between presentation objects and nodes in ontologies.

For ease of modelling and visualisation the open source ontology and knowledgebase editor Protégé[5] is utilized for creating and editing the metadata scheme. Protégé has also been selected, because it supports data export in RDF format. For RDF schema inferencing and querying the data instances are stored in Sesame[6] which too is an open-source semantic web tool.

4. Tour Generation

In Figure 4 we illustrate the overall architecture of the tour generation component. A tour unfolds stepwise by generating and presenting minimal presentation units one after the other. Relevant presentation objects are selected according to

1. input parameters specified via the user interface, such as device (PC or mobile), category, relevance, temporal and spatial anchor, and the region (computed by means of user position and orientation) in the 3D model the user focuses on. In the PC application the current view in the 3D model determines the area of user attention. On the mobile device the user location and orientation are determined by GPS data, a tilt sensor and a magnetic compass, cf. [5]. A point or interval in time is selected via a slider in the interface. Zooming on the timeline controls the level of detail.

2. the user history storing information on which parts of the 4D model the user has already visited in the course of a session and which presentation objects have already been presented.

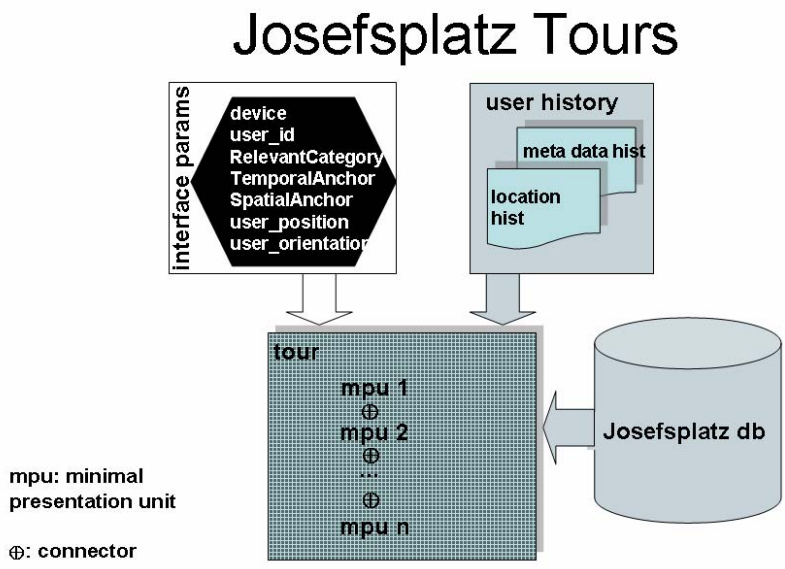

Figure 4: Tour generation: general architecture

[5] http://protege.stanford.edu/

[6] http://www.openrdf.org

Objects are grouped into minimal presentation units by analysis of their link structure, (semantic) similarity of content, i.e. whether they belong to the same or related categories, and temporal co-occurrence. In the simplest case, a minimal presentation unit consists of a single presentation object only. Another example for a minimal presentation unit is a text object and the pictures and possibly some pieces of music linked to it. Text objects in the Josefsplatz database are also stored as synthesized speech. Such a presentation unit is generated as a standard HTML page interlinking text, pictures and music, and as a multimedia SMIL[7] presentation with temporal alignment of voice, pictures and sound. Different templates may be applied depending on the topic and the media types in a unit. These may contain transition elements such as small pieces of synthesized speech to keep the presentation coherent. Timing information is extracted from the Mary TTS output and links in the text source. Link position and clause structure determine the temporal positioning of pictures. The HTML pages are created dynamically by use of extended stylesheet language transformations (XSLT). This approach allows presentations to be adjusted not only to their content and the display device but also to accessibility and usability demands.

For the serialisation of minimal units certain properties such as temporal linearity, spatial distance, temporal coverage, size of the minimal units, distance to the centre of user attention/orientation, or similarity of topic may be considered more important. For example it may be desirable to break up temporal linearity if the majority of units are of one topic and only a smaller fraction of the presentation units is of a second topic. As a consequence two topic blocks are created which are in themselves temporally serialised. In practice, the starting point of the presentation is selected according to the focus of the user orientation (cursor position). For computing the tour the properties of the presentation units are mapped into feature vectors with weights according to the importance of properties. The next move from a given point in the presentation is determined by calculating its nearest neighbour in Euclidean vector space. Depending on the level of detail requested by a user, units of a presentation may be classified as optional. To ensure cohesion of the tour another class of connector elements is added between the moves. These connectors are selected according to the types of minimal units. They may be prefabricated text templates, selected pictures, sound files, etc. Moves may also be serialized without the use of intermediate material. User interaction at each transition allows for detailed viewing of presented material, selection of optional units, navigating backwards or forwards in the presentation or returning to the 3D model view. The SMIL presentation is translated into a format suitable for the 3D engine and playback is integrated into the 3D model. The user is walked through the model, always being taken to the best viewing position on the presented objects.

5. Conclusion

Creative Histories brings a new dimension and quality to the emerging issue for digitisation of Cultural Heritage and its broad access via mobile devices. Metadata is extracted semi-automatically from legacy documents and stored in a database. Different strategies for presentation of multimedia content apply, depending on the user-defined context represented via the interface parameters category, timeline and relevance, the metadata available, and the capabilities of the interaction devices (PC, mobile phone). Tours are automatically generated by serialising groups of presentation objects according to properties such as temporal linearity, spatial distance, topic etc. mapped into weighted feature vectors. From a given starting point, next moves are calculated step by step from nearest neighbour to nearest neighbour in vector space. Each part of a tour is presented as a

[7] www.w3.org/TR/REC-smil/

synchronized multi-media presentation adapted to the capabilities of the output device, and making use of state of the art speech synthesis tools.

Summing up, we create a new cultural experience for the user where it is possible to navigate in a four-dimensional space, i.e. the 3D spatial environment and its architectural changes over time, and where the user is presented with meta-information (tours) grounded in the spatio-temporal environment. The system demonstrates a novel user interface and a novel experience of CH, accessible for a wide mobile audience. It is extendible and supports the interaction of the public with cultural content and cultural institutions.

Acknowledgments

The work has been partially supported by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) under the project “Creative Histories. The Josefsplatz Experience”. OFAI is supported by the Austrian Federal Ministry for Education, Science and Culture and by the Austrian Federal Ministry for Transport, Innovation and Technology. Furthermore we wish to thank Peter Fröhlich from FTW for his support in speech synthesis.

6. References

[1] Little, S., Geurts J., Hunter J. "Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inferencing", Lecture Notes in Computer Science, Volume 2458 / 2002, Springer, Berlin, pp. 158 – 175.

[2] Neumann M. Spatially Navigating the Semantic Web for User Adapted Presentations of Cultural Heritage Information in Mobile Environments. In Cruz, I.F., Kashyap, V., Decker, S. Eckstein, R. (eds). Proceedings of SWDB’ 03. The first International Workshop on Semantic Web and Databases, Berlin, September 7-8, 2003, pp. 9-14.

[3] Rutledge, L., Alberink, M., Brussee, R., Pokraev, S., van Dieten, W., and Veenstra, M. Finding the Story - Broader Applicability of Semantics and Discourse for Hypermedia Generation. In: Proceedings of the 14th ACM Conference on Hypertext and Hypermedia Nottingham, UK, ACM Press, August 23-27, 2003, pp. 67-76.

[4] Schröder, M., Trouvain, J. The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. *International Journal of Speech Technology*, 6, 2003, pp. 365-377.

[5] Simon, R., Kunczier, H., Anegg, H. "Towards Orientation-Aware Location Based Mobile Services." In Gartner, G. (ed). Proceedigs of 3rd Symposium on Location Based Services and Telecartography, Vienna, Austria, November 28-30, 2005. Geowissenschaftliche Mitteilungen, Nr. 74, pp. 53-57.

[6] Sormann, M., Zach, Ch., Zebedin, L., Konrad Karner, K. HIGH QUALITY 3D RECONSTRUCTION OF COMPLEX CULTURAL OBJECTS. In (online) Proceedings of CIPA International Symposium. Torino, 26. Sept. – 2. Oct. 2005. http://cipa.icomos.org/fileadmin/papers/Torino2005/952.pdf (viewed 7.6.2006)

[7] Vlahakis, V. Ioannidis, N., Karigiannis, J., Tsotros, M., Gounaris, M., Stricker, D., Gleue, T., Dähne, P., Almeida, L. Archeoguide: An Augmented Reality Guide for Archaeological Sites. In: IEEE Computer Graphics and Applications 22 (2002), 5, pp. 52-60.