



# Using AI to Access and Experience Cultural Heritage

Lynda Hardman  
and Jacco van Ossenbruggen,  
*Centrum Wiskunde & Informatica*

Lora Aroyo,  
*Vrije Universiteit Amsterdam*

Eero Hyvönen,  
*Helsinki University of Technology*

*Cultural heritage involves rich and highly heterogeneous collections that are challenging to archive and convey to the general public.*

**T**he digital age is transforming cultural heritage in methods of both creation and preservation. Whereas once we collected objects such as books, sculptures, statues, and paintings, we now also face the preservation and the archiving of digital artifacts. These might be digital representations of physical objects or purely digital creations that are culturally significant and worthy of preservation in their own right, such as interactive works of art, blogs, or even the World Wide Web itself. Intelligent systems can be used at different stages of creation, identification, preservation, authentication, and retrieval of these digital assets.

The interest generated by the First International Workshop on Cultural Heritage on the Semantic Web ([www.cs.vu.nl/~laroyo/CH-SW/organization.html](http://www.cs.vu.nl/~laroyo/CH-SW/organization.html)), held in Pusan, South Korea, in November 2007, inspired

this special issue. A large number of papers were submitted to the special issue: 33 in total, of which all but one were sent for review. They represented a large variety of topics in this comparatively narrow domain, and we are pleased that the final six papers selected for the special issue retain this diversity.

Cultural heritage institutions are excellent partners in research projects on curating and providing access to cultural assets because their mission is to share information with others. Funding for work beyond the required maintenance, registration of

## THE AUTHORS

**Lynda Hardman** is the head of the Interactive Information Access group at the Centrum Wiskunde & Informatica and a professor of multimedia interaction at the University of Amsterdam. Her research interests include exploring methods for presenting semantically annotated media information to end users, creating novel interfaces for information exploration and gathering tasks in semantically annotated media repositories, and investigating the underlying knowledge and document infrastructures required. Hardman received her PhD in computing science from the University of Amsterdam. She's a member of the British Computer Society and the ACM. Contact her at [lynda.hardman@cwi.nl](mailto:lynda.hardman@cwi.nl).

**Lora Aroyo** is an assistant professor at the Web and Media group at the Vrije Universiteit Amsterdam. Her research interests include personalized access to semantically enriched collections. She leads the Cultural Heritage Information Personalization (CHIP) project with the Rijksmuseum Amsterdam dealing with semantic recommendations for museum artworks and generating personalized museum tours. She's collaborating in the research and developed of the iFanzzy personalized TV-program recommendation system. She's also technical coordinator of the European Integrated Project NoTube on the integration of TV and Web data with the help of semantics. Aroyo received her PhD in intelligent educational systems from the University of Twente. She's a member of the Association for the Advancement of Artificial Intelligence and the ACM. Contact her at [l.m.aroyo@cs.vu.nl](mailto:l.m.aroyo@cs.vu.nl).

**Jacco van Osssenbruggen** is a senior researcher with the Interactive Information Access group at the Centrum Wiskunde & Informatica and an assistant professor at the Web and Media group at the Vrije Universiteit Amsterdam. His research interests include Semantic Web interfaces, multimedia on the Semantic Web, and the automatic generation of user-tailored hypermedia presentations. He cofounded the W3C Incubator Group on Multimedia Semantics, has organized several workshops at the International WWW Conference, and is currently active in the MultimediaN E-Culture project. Van Osssenbruggen received his PhD in computer science from the Vrije Universiteit Amsterdam. He is a member of the ACM. Contact him at [jacco.van.ossenbruggen@cwi.nl](mailto:jacco.van.ossenbruggen@cwi.nl).

**Eero Hyvönen** is a professor of semantic media technology in the Department of Media Technology at the Helsinki University of Technology and a research director in the Department of Computer Science at the University of Helsinki. He directs the Semantic Computing Research Group, which focuses on research of Semantic Web technologies and their applications. The group has created the semantic cultural heritage portals MuseumFinland and CultureSampo. Hyvönen received his PhD in computer science from the Helsinki University of Technology. Contact him at [eero.hyvonen@tkk.fi](mailto:eero.hyvonen@tkk.fi).

collections and digitization of objects is not always easy to obtain. The Netherlands has been very fortunate with longer term funding support, resulting in the high percentage of articles with Dutch-based authors in this special issue.

### Contributions

When cataloging artifacts, precise information on what the object is, and where and how it was created is necessary. Two papers investigate the use of intelligent systems to improve the accuracy of identification and classification of artifacts. Martin Kampel, Reinhold Huber-Mörk, and Maia Zaharieva present an article called "Image-Based Retrieval and Identification of Ancient Coins." Their sys-

tem applies a number of image analysis methods to determine descriptors such as a coin's outline from potentially low-quality images. Additional techniques correlate features on the faces of the coin, guided by orientation information from the process determining the outline. The results were tested on a collection of 240 different coins documented by the Fitzwilliam Museum.

In "Semantic Classification of Byzantine Icons," Paraskevi Tzouveli, Nikos Simou, Giorgios Stamou, and Stefanos Kollias explore different methods for identifying Byzantine icons, based on recognition of the sacred figure portrayed. The low variability of the image characteristics and the strict rules and iconographic patterns followed

by most artists enable successful application of the image analysis methods. The objects recognized, in turn, can be mapped to formal domain descriptions—for example, "young face" or "long hair"—in Semantic Web languages such as OWL. The authors applied their techniques to a set of 2,000 Byzantine images provided by the Mount Sinai Foundation. The images date from the 13th century, depicting around 50 different saints. The accuracy of the face detection module was 80 percent, where failure occurred mostly where the face area had been damaged.

As an example of creation, in "Automatic Generation of Chinese Calligraphic Writings with Style Imitation," Songhua Xu, Hao Jiang, Tao Jin, Francis C.M. Lau, and Yunhe Pan propose an algorithm that creates Chinese calligraphy by simulating the writing style of a calligraphist. They hope that systems such as theirs can help to rekindle interest in this important aspect of Chinese culture, particularly among young people with little appreciation for this ancient art. The system learns the style of a particular calligraphist using a stroke-based representation that takes the variability of the calligrapher into account. It can then generate new texts from the learned style. The system thus uses intelligent techniques to not only preserve the styles of different calligraphers but also create new artifacts.

Once the associated properties of artifacts are recorded in a database, chances are that they are not completely accurate. Antal van den Bosch, Marieke van Erp, and Caroline Sporleder present an approach to cleaning cultural heritage databases with the article "Making a Clean Sweep of Cultural Heritage." They present four case studies using databases from different cultural heritage institutions. Their method uses machine learning techniques to identify potential errors in the data. These are

conveyed to curators and researchers for authentication by human experts.

Machines can also generate metadata and add it to a database to improve subsequent retrieval. In “Knowledge-Based Linguistic Annotation of Digital Cultural Heritage Collections,” Tuukka Ruotsalo, Lora Aroyo, and Guus Schreiber produce annotations automatically for objects accompanied by a text description, a set of structured vocabularies, a metadata schema, and a training set of annotations. The authors focus on identifying the metadata schema roles that concepts play in the text—that is, Paris as subject matter rather than as place of creation. They evaluated their method using a data set with over 700 major exhibits from Rijksmuseum Amsterdam, annotated with a number of different vocabularies, including the Getty Thesaurus of Geographic Names.

Once a collection is classified in a database, users then need access to the information that interests them. Antoine Isaac, Shenghui Wang, Claus Zinn, Henk Matthezing, Lourens van der Meij, and Stefan Schlobach investigate how alignments among different thesauri can help improve access to collections and thesauri with the article “Evaluating Thesaurus Align-

ments for Semantic Interoperability in the Library Domain.” The authors explore common real-world problems in the National Library of the Netherlands. Two collections are indexed by separate thesauri with roughly the same coverage but different granularity. Each is maintained separately and does not provide access to the set of books described by the other. The authors investigate the improvements that four different thesaurus-mapping techniques can bring to search results.

**T**hese six articles represent only a portion of the richness and diversity of the cultural heritage field, and a sample of the breadth of techniques for improving the different stages of cultural heritage curation. But we hope they give some insights into the valuable work being carried out in this area. The close cooperation between “ivory tower” researchers and cultural heritage institutions indicates both a rich source of problems that still require solutions and a willingness from both sides to participate in a dialogue to evaluate state-of-the-art techniques. ■

