CULTURE HERITAGE DIGITAL REPOSITORIES.

RESEARCH QUESTIONS

Peter L. Stanchev
Kettering University
Flint, Michigan, 48504 USA
e-mail: pstanche@kettering.edu
WWW URL: www.kettering.edu/~pstanche

This discussion is about innovative solutions for assembling multimedia digital repositories for collaborative use in specific contexts and communities and enhancing scholarly understanding and experiences of digital cultural heritage. Several aspects are stress such as the dynamic aggregation of cross-media resources across existing institutional digital libraries and repositories. Research questions about the scalability, interoperability and distributed architectures, aggregation, and semantic search tools are addressed.

Keywords: Culture Heritage, Digital Repositories

1. CULTURE HERITAGE DIGITAL REPOSITORIES

The construction of European Knowledge Society is based on transparent, free and ubiquitous access to knowledge. In its strategic initiative i2010 *Digital Libraries*, COM (2005) 465 final, Brussels, 30/09/2005, the European Commission foresees the creation of digital libraries which will make results of research, multimedia resources and the multilingual culture of Europe accessible to all. This initiative also presents the European strategy for digitalisation, on-line access and digital preservation of the cultural heritage. The document appeals to the EU member states to develop digitalisation tools on a wide scale, so as to speed up the process of integrating the European cultural heritage into the European information space. The Commission also recommends actions for solving problems such as copyright, access rights, the systematic preservation of digital content, etc. In 2008 the European Commission presented the document "Europe's cultural heritage at the click of a mouse: progress on the digitisation and online accessibility of cultural material and digital preservation across the EU" (COM(2008) 513), Brussels, 11/08/2008), which stated the directions in the digitalisation of Europe's cultural and historic heritage.

In its Fourth Challenge *Digital Libraries and Content*, the 7[th]FP ICT Work Programme 2009-2010 clearly defines the topicality of the subject matter and the principal directions of research for the European Research Area: to create new possibilities of using and sharing Europe's rich cultural and scientific resources and new services enriching the perception and understanding of these resources - aggregating and annotating objects available in digital libraries, 3D visualisation and access to virtually recreated cultural artefacts.

The state of European research in the area of multimedia digital libraries, conceptual and technological solutions, specifications, proposals and recommendations are set out to a sufficient extent in the research project **DELOS** – *A Network of Excellence on Digital Libraries* and are available on the project's site (http://www.delos.info). The state of the research and the recommendations in the area of digitalisation of cultural and historic heritage and the tasks of EU member states are set out in the European project **MINERVAPLUS** *Ministerial Network for Valorising Activities in Digitization* (http://www.minervaeurope.org). The methodology for digitalisation and digital presentation of cultural heritage, as well as, best practices in the area, are developed in the European project **MICHAEL** *Multilingual Inventory of Cultural Heritage in Europe* (http://www.michael-culture.org).

The cultural digital repositories can employ similar techniques as in generic systems in order to solve standard questions for searching objects. The cultural heritage objects are rich in content, describing events, monuments, places, people etc. and distributed across different locations. The users can formulate queries using different modalities such as free text, similarity matching, or metadata.

In 2004 David Mattison, named as a master of the online archive universe, published a series of lectures in the Searcher magazine focusing on state-of-the-art of available Web resources current techniques for object retrieval, and finishing with mentioning of national collections that document the art history of Western civilization from medieval times through the 19th century. The creators of art collections and guides usually are academic, librarian, commercial, and private art museums and galleries, amateur and professional art historians, artist sites, commercial image agencies, auction houses (usually on a temporary basis), etc.

Some of the projects, such as *Vasari* (1989-1992) and *Marc* (1995-1996), are focused on digital acquisition, storage and handling of colorimetric high-definition images of paintings (up to 2GB per image) for different galleries and museums in Western Europe. The *Crisatel* project (2001-2004) developed equipment for the direct fast capture of paintings, with a new ultra-high definition multi-spectral scanner in order to make spectrometric analysis of varnish layers to allow the effect of an aged varnish to be subtracted from an image of a painting. Project *FingArtPrint* (2005-2008) is aimed to make combination of 3D surface scanning and multispectral imaging in order to create a unique data record of the object which can be compared to check its authenticity.

NARCISSE (1990-1992) was one of the first projects, which has created with a very high digitized image bank, supervised by a multilingual text database. The objective of the project *Artiste* (2000-2002) was to develop and prove the value of an integrated art analysis and navigation environment aimed at supporting the work of professional users in the fine arts.

In recent years several projects and initiatives are focused on harmonizing activities carried out in digitization of cultural and scientific content, for creating a common platform to cultural heritage. Such project is *MINERVA+*, sponsored by 6FP, which intends to enlarge the existing thematic network of European Ministries in this direction. Since 2005 the Netherlands Organization for Scientific Research supports the research program *CATCH*

(Continuous Access to Cultural Heritage) that finances teams which focus on improving the cross-fertilization between scientific research and cultural heritage.

The latest and biggest project in this domain is *Europeana* (http://www.europeana.eu/portal/), funded by the European Commission and the member states. The idea of *Europeana* was born in 2005, when the European Commission announced its strategy to promote and support the creation of a European digital library. The European Commission's goal for *Europeana* is to make European information resources easier to use in an online environment. Till now more than 9 millions digital items are available.

Some examples of repositories with heritage objects are: *Artcyclopedia* (http://www.artcyclopedia.com/) is an online database of museum-quality fine art founded by Canadian John Malyon; *Artchive* (http://www.artchive.com/) is a virtual art gallery website, ordinate by Mark Harden; *OCAIW* (Orazio Centaro's Art Images on the Web) (http://www.ocaiw.com/) is an educational and non-profit site for art-lovers, teachers, students, artists and collectors; *WebMuseum* (http://www.ibiblio.org/wm/) is one of the earliest examples of a virtual museum; *Web Gallery of Art* (http://www.wga.hu/) is a virtual museum and searchable database of European painting and sculpture of the Romanesque, Gothic, Renaissance, Baroque, Neoclassicism, Romanticism periods (1000-1850), currently containing over 23 200 reproductions; *Backtoclassics* (http://www.Backtoclassics.com/) is a new virtual art gallery (since 2009), created by Italian division of Microsystems MS Lab, which gives a look at the creations of artists past and present, where paintings are classified not only by movements and artists, but also thematically (for instance Rembrands' paintings are grouped into following series: Portraits; Biblical Themes; Various Paintings; Self-Portraits; Etchings; Drawings; Landscapes); *Olga's Gallery* (http://www.abcgallery.com/) is one of the largest online painting museums, containing works and biographies of most of the world's best known artists.

Currently a lot of museums have online galleries, supplying access to the museum's collections. The search engines of some of them are based only on metadata information using categories such as artist, title of work, subject, chronology and reference number (for instance http://www.museodelprado.es/ en/the-collection/online-gallery/). In other sites attempts to implement CBIR (Content-Based Image Retrieval) techniques during the search process are included (such as http://www.hermitagemuseum.org).

2. RESEARCH QUESTIONS

The newly available image capture techniques, inexpensive storage, and widely available dissemination methods have made digital objects a convenient and easily available. This increased availability of objects is accompanied by a need for solutions to the problems in indexing and retrieval. In the multimedia retrieval systems, an important parameter to measure user-system interaction level is the complexity of queries supported by the system. The queries are using different modalities, such as: direct filling the values of the desired features; giving the image or sketch as example; keywords or free-text, and their combination.

Relevance feedback is very important step in image retrieval, because it defines the goals and the means to achieve them. Relevance feedback provides a compromise between a fully automated, unsupervised system and one based on subjective user needs. It is a query modification technique which attempts to capture the user's precise needs through iterative feedback and query refinement.

Today interoperability is considered a key-step to move from isolated digital repositories towards a common information space that allow users to browse through different resources within a single integrated environment. Some problems are connected with the scientific study and development of innovative technological solutions for assembling multimedia digital libraries for collaborative use in the context of cultural heritage, supporting their semantic interoperability and developing new services for dynamic aggregation of their resources, improvement of access, personification, intelligent curation of content, as well as content protection and ensuring intellectual property rights. The interoperability can be divided in the following categories:

- *Organizational interoperability:* refers to cooperation between and within digital library organizations, business goals and process modeling;

- *Semantic interoperability:* refers to understanding the meaning of information in digital libraries;

- *Technical interoperability:* refers to interconnection, presentation and exchange of digital objects within digital library, accessibility and security issues.

By his nature, the objects, presented in the digital space can be:

- *Digitally born art-objects*. These new kind of masterpieces live in digital media, which is theirs primary environment;
- *Digital images of analogue art-objects*. Here a digitized work of art is not the work itself but an image (instance) of this work;
- *Objects, which are connected with other art-objects.* For instance, essay over some painting period of the artists, or sketch, representing main characteristics of painting, sculpture or architecture building.

The standard royalties are expanded with the new ones, concerning life-cycle of digital objects in Web space. One attempt for establishing legal framework for open multimedia supply and consumption to be used by all participants in the chain is made in MPEG-21. It defines a "Rights Expression Language" standard as means of sharing digital rights/permissions/restrictions for digital content from content creator to content consumer. As an XML-based standard, MPEG-21 is designed to communicate machine-readable license information and do so in a "ubiquitous, unambiguous and secure" manner. This open framework provide content creators, producers, distributors, representatives and service opportunities of the existing free market. This also benefit consumers by providing access to a wide variety of content in an interoperable way. MPEG-21 is based on two basic concepts:

- A basic *unit* of distribution and transaction - a digital item - which is "object" in the multimedia framework (e.g., video art, music album);

- The concept of *user* interacting with him, which is "subject" in this framework.

Not much time passed before the idea of "Web 3.0" appeared. Amit Agarwal states that Web 3.0 is about semantic web (or the meaning of data), personalization (e.g. iGoogle), intelligent search and behavioral advertising among other things. While Web 2.0 uses the Internet to make connections between people, Web 3.0 will use the Internet to make connections with information. The intelligent browsers will analyze the complex requests of the user, given in natural language, search the Internet for all possible answers, and then organize the results for him. The adaptation to user specifics and aptitudes will be based on capturing the historical information thorough searching the Web. Many of the experts believe that the Web 3.0 browser will act like a personal assistant. The computer and the environment became as artificial subjects, which will pretend to communicate in real manner as real humans. Of course, the problems of applying rights policies in such new atmosphere are crucial.

SOME   RESEARCH QUESTIONS:

- Design the basic architectures, development of multimedia digital repositories for cultural heritage;

- Design distributed architecture for assembling multimedia digital libraries for collaborative use in the cultural heritage domain and specification of innovative services (maintenance of a personal work space, improvement of search and browsing, intelligent curation of content, content protection and ensuring Intellectual Property Rights);

- Specification of the semantic elements for identification of multimedia objects in assembling digital libraries for collaborative use in cultural heritage, digitization, annotation and semantic indexation;

- Design, development and testing Web 3.0 culture heritage digital repositories.