ACCESS TO DIGITAL CULTURAL HERITAGE

Innovative Applications of Automated Metadata Generation

Kr. Ivanova, M. Dobreva, P. Stanchev, G. Totkov and other contributors
Authors (in order of appearance):
Krassimira Ivanova, Milena Dobreva, Peter Stanchev, George Totkov, Kalina Sotirova, Juliana Peneva, Stanislav Ivanov, RositzaDoneva, Emil Hadzhikolev, Georgi Vragov, Elena Somova, Evgenia Velikova, Iliya Mitov, Koen Vanhoof, Benoit Depaire, Dimitar Blagoev

Editors:
Krassimira Ivanova, Milena Dobreva, Peter Stanchev, Georgi Totkov
Support of the Work

- the Bulgarian National Science Fund
  - Project D002-308 "MetaSpeed: Automated Metadata Generating for e-Documents Specifications and Standards"
- the FP7 Framework programme
  - Project RI-246686 "OpenAIRE: Open Access Infrastructure for Research in Europe"
- the Hasselt University
  - Project R-1875 "Search in Art Image Collections Based on Colour Semantics"
  - Project R-1876 "Intelligent systems' memory structuring using multidimensional numbered information spaces"
The main purpose of this book is to provide an overview of the current trends in the field of digitization of cultural heritage as well as to present recent research done within the framework of the project D002-308 funded by Bulgarian National Science Fund. The main contributions of the work presented are in organizing digital content, metadata generation, and methods for enhancing resource discovery.
Chapter 1: Digitization of Cultural Heritage – Standards, Institutions, Initiatives

Chapter 2: REGATTA – Regional Aggregator of Heterogeneous Cultural Artifacts

Chapter 3: Automated Metadata Extraction from Art Images

Chapter 4: APICAS – Content-Based Image Retrieval in Art Image Collections Utilizing Colour Semantics

Chapter 5: Automatic Metadata Generation and Digital Cultural Heritage
The main pillars of process of keeping and accessing of cultural heritage in digital space: digitization, access and preservation

The metadata schemas and standards used in cultural heritage

The models of contemporary digital libraries

Some initiatives on world and European level that during the years enforce the process of digitization and organizing digital objects in the cultural heritage domain

Shifting the focus in the creation of digital resources from "system-centred" to "user-centred".
REGATTA is aimed to present different types of collections, including museum collections, archaeological sites, immovable heritage from Ancient, Mediaeval and National Enlightenment periods in Bulgaria.

It was designed following the standards of Europeana and characteristics specified in the Bulgarian regulation for creating and managing of museum funds.

The functional scheme, data model, and technological aspects in REGATTA are discussed.

The first application of the REGATTA - Plovdiv Ethnographic Museum where artifacts from fund "Crafts" are presented and possibilities to make virtual tours in the museum are shown.
Over the past decade, considerable progress has been made to make computers learn to understand, index and annotate pictures representing a wide range of concepts.

The content-based image retrieval (CBIR) is a technology that helps to organize digital images based on their content.

Typically, a CBIR system consists of three components:

- feature design - extracts the visual feature(s) information from the images in the image database
- indexing - organizes the visual feature information to speed up the query or processing
- retrieval engine - processes the user query and provides a user interface

During this process the central issue is to define a proper feature representation and similarity metrics.
Colour theory from different points of view - physiological, psychological, social

Proposed colour models during the years

Classification of harmonies and contrasts in accordance to Ittens' theory from the point of view of three main characteristics of the colour – hue, saturation and luminance

Formal description of defined harmonies and contrasts

Program system APICAS ("Art Painting Image Colour Aesthetics and Semantics") as environment for realizing proposed algorithms and for conducting experiments

A variety of experiments on the use of the system for different tasks (similarity search, user queries, predictive analysis) shows the vividness of proposed features as a step in the transition from Web 2.0 to Web 3.0.
Some automatic metadata generation methods that can be used in the field of cultural heritage:

- **indirect spatial data extraction by learning restricted finite state automata**: realized in system InDES and tested over extraction of spatial metadata from web sites. Such approach can be used for metadata extraction from objects descriptions and this way to be used in the process of migration from older object representations in the case when the descriptions are in non-structured form.

- **association rule mining, which uses the as a storage structures multidimensional numbered information spaces**: realized algorithm ArmSquare in data mining environment system PaGaNe. The possibilities of extracting frequent item-sets can be used for enforcing connections between metadata elements within created ontology.

- **associative classifier PGN**: also realized in PaGaNe. It was implemented in the field of analyzing semantic attributes, extracted from art images using CBIR. Within the frame of the data management and the access of the aggregator the classifier PGN can be used for enforcing information discovery.
The Connection between Pieces

Diagram showing the relationship between Cultural Heritage Holders, Digital Objects, Metadata Standards, Automated Metadata Extraction, Data Mining Methods, Preservation Planning, Ingest, Data Management, Archival Storage, Access, Administration, and queries/result sets/orders.
Thank you for your attention!

and

Enjoy the reading in 2012!

15.12.2011, IMI-BAS