Introduction

**Krassimira Ivanova, Milena Dobreva,
Peter Stanchev, George Totkov**

The access to cultural heritage artefacts in the digital space is already in the stage of maturity. The romantic time of pilot attempts to digitise several selected objects is long gone, as well as the debate if digitisation is done for preservation or for access. If we can take as a sign of the significance and the scale of effort the collection of statistical data on pan-European scale, this started with the NUMERIC[1] study in 2007-2008 and currently continues with the ENUMERATE[2] project. Europeana, the showcase digital library of Europe, grows quickly in the number of objects which can be discovered and accessed through it – currently about 25 million.

The European Commission Recommendation of 27 October 2011 on the digitization and online accessibility of cultural material and digital preservation (2011/711/EU) is very straight-forward in stating that: *"The Digital Agenda for Europe seeks to optimise the benefits of information technologies for economic growth, job creation and the quality of life of European citizens, as part of the Europe 2020 strategy. The digitisation and preservation of Europe's cultural memory which includes print (books, journals and newspapers), photographs, museum objects, archival documents, sound and audiovisual material, monuments and archaeological sites (hereinafter 'cultural material') is one of the key areas tackled by the Digital Agenda."* It further emphasizes that the member states should develop their planning and monitoring of the digitisation of books, journals, newspapers, photographs, museum objects, archival documents, sound and audiovisual material, monuments and archaeological sites, and should contribute to the further development of Europeana. According to this Recommendation, Bulgaria

---

[1] http://www.numeric.ws/
[2] http://www.enumerate.eu/

is currently present with 38,263 objects in Europeana and is expected to contribute further 267 000 objects until 2015. This is indeed a change of scale which can not be achieved with the efforts of a single institution, and will require not only to intensify digitization efforts, but also to integrate intelligent tools in the digitization as much as possible.

However, the advancement in the area of digital cultural heritage is not only in the numbers of digitised objects. In the last decades, the larger involvement of heritage institutions in digitisation activities motivated a broad range of research activities. Some of them addressed the changing landscape of humanities and arts and inspired the growth of the area of Digital Humanities. But the practical digitisation needs also required specific support from the information and computer sciences. Areas such as dealing with "big data", new metadata models, new methods for content-based analysis and retrieval are only few examples of information research which not only could be applied to a new domain but also had to cater for the specific needs of the digital heritage.

While in many European countries the discussions of the value and impact of digital collections are currently high on the agenda, Bulgaria still finds itself in a different situation where the key issues are to achieve a critical mass of digital content, and to integrate this critical mass better within the European efforts to provide access to the digital cultural heritage.

In the recent years, four national events, organised by Emil Stoyanov, MEP, member of the Committee on Culture and Education in European Parliament, three in Plovdiv and one in Brussels, were held. These events gave the opportunities to the experts from memory institutions, representatives of the government and scholars to discuss the achievements in digitisation in Bulgaria and the best way forward. These events had an essential role in building a professional community of those already experienced or just starting in digitisation.

One of the outcomes of this intensified cooperation of institutions is this book. This edited collection brings together the outcomes of national and international research projects aiming to develop specialised technologies and tools which would be of help in digitisation efforts.

➢ *The Projects that Supported the Creation of this Book*

This book presents the outcomes of several projects which created and delivered specialised information research methods in the digital heritage domain. Below, we give a brief explanation of the projects.

✓ Project **D002-308** *"MetaSpeed: Automated Metadata Generating for e-Documents Specifications and Standards" funded by the Bulgarian National Science Fund*

The main objective of the project was the creation and testing of technologies, methods and tools for automated specification documents of different electronic format (text, graphics, etc.), different content (cultural and historical artefacts, educational materials, scientific publications, etc.), and different location (local multimedia repositories, web pages, etc.).

To achieve the goal several research tasks were addressed:

– Standards for e-documents and tools for their automatic generation: to make critical analysis of existing standards in the areas of cultural heritage, scientific publications, e-learning, and geospatial data.

– Automatic generation of metadata from text documents: to create and study of special methods, algorithms and tools for extracting structured data of electronic text documents, especially of the Bulgarian language.

– Automatic generation of metadata from multimedia documents: supposed to be developed and used methods and tools for automated extraction of metadata for images, both context and content of objects.

– Creation and testing of digital repositories in different areas: cultural heritage, scientific publications, e-learning, and geographic information systems.

The project executives are the teams from Plovdiv University "Paisii Hilendarski" (coordinated by Prof. George Totkov), Institute of Mathematics and Informatics – Bulgarian Academy of Sciences (coordinated by Prof. Peter Stanchev), New Bulgarian University (coordinated by Assoc. Prof. Juliana Peneva), and Sofia Technical University (coordinated by Prof. Elena Shoykova).

✓ Project **RI-246686** *"OpenAIRE: Open Access Infrastructure for Research in Europe" within the FP7 Framework programme*

Creating and maintaining a strong network for European research cooperation is high on the EU agenda. Helping drive this effort is OpenAIRE, a project encouraging and supporting free online access to knowledge generated by researchers with grants from the Seventh Framework Programme (FP7) and the European Research Council. One of the most critical components of the OpenAIRE project is to give researchers, businesses and citizens free and open access to EU-funded research papers. The OpenAIRE infrastructure is also helping devise new methods of indexing, annotating, ordering and linking research results, as

well as automating processes. With these activities in mind, OpenAIRE will contribute to the development of fresh services on top of the information infrastructure which it offers.

The project coordinator for Bulgaria is Prof. Peter Stanchev from the Institute of Mathematics and Informatics – Bulgarian Academy of Sciences (IMI-BAS).

✓ *Project **R-1875** "Search in Art Image Collections Based on Colour Semantics" funded by the Hasselt University*

This is the doctoral research project of Krassimira Ivanova from IMI-BAS with advisor Prof. Koen Vanhoof from Hasselt University. The aims of the project were to make a comprehensive analysis of successful colour combinations examined and used by the artists during the centuries which send different impression, expression and construction messages to the viewers. Typical features are color contrasts, because one of the goals of the painting is to produce specific psychological effects in the observer, which are achieved with different arrangements of colors.

The main goals of this work were:

– to provide a detailed analysis of the colour theories, especially on existing interconnections in successful colour combinations;

– to formalize them in order to implement automated extraction from digitized artworks.

The extracted features were successfully used for similarity search with selected image by one or more of the extracted features; search of images that satisfied user queries featuring contrasts' characteristics; as well as for investigation on the possibilities to integrate such characteristics within specialized resource discovery (searching for distinctive feature of movements, artists or artists' periods).

✓ *Project **R-1876** "Intelligent systems' memory structuring using multidimensional numbered information spaces" funded by the Hasselt University*

This is the doctoral research project of Iliya Mitov from IMI-BAS with advisor Prof. Koen Vanhoof from Hasselt University.

The goals of this thesis were two-fold:

– to introduce a parameter-free class association rule algorithm, which focuses primarily on the confidence of the association rules and only in a later stage on the support of the rules. We expect that this approach will ensure implementing high-quality recognition especially within unbalanced and multi-class datasets. The nature of such a classifier is more oriented to having characteristic rules;

− to show the advantages of using multidimensional numbered information spaces for developing memory structuring in data mining processes on the example of implementation of the proposed class association rule algorithms.

The main results of this project lay as one of the basic algorithms in data mining environment PaGaNe. The system incorporates different types of statistical analysis methods, discretization algorithms, association rule miner, as well as classification algorithms.

➢ *The Book Content*

The book consists of five chapters. Here we give a brief explanation of them.

The first chapter "Digitization of Cultural Heritage – Standards, Institutions, Initiatives" provides an introduction to the area of digitisation. The main pillars of process of creating, preserving and accessing of cultural heritage in digital space are observed. The importance of metadata in the process of accessing to information is outlined. The metadata schemas and standards used in cultural heritage are discussed. In order to reach digital objects in virtual space they are organized in digital libraries. Contemporary digital libraries are trying to deliver richer and better functionality, which usually is user oriented and depending on current IT trend. Additionally, the chapter is focused on 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. In recent years, the main focus in the creation of digital resources shifts from "system-centred" to "user-centred" since most of the issues around this content are related to making it accessible and usable for the real users. So, the user studies and involving the users on early stages of design and planning the functionality of the product which is being developed stands on leading position.

Chapter 2 "REGATTA – Regional Aggregator of Heterogeneous Cultural Artefacts" describes the prototype of REGATTA (REGional Aggregator of heTerogeneous culTural Artefacts), aimed to present different types of collections, including museum collections, archaeological sites, immovable heritage from Ancient, Mediaeval and National Enlightenment periods in Bulgaria. The chosen approach supports the idea of preserving the valuable national monuments in the European area of culture keeping their identity and specificity. 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. Currently

the first application of the REGATTA is applied for Plovdiv Ethnographic Museum where artefacts from fund "Crafts" are presented and possibilities to make virtual tours in the museum are discussed.

Chapter 3 "Automated Metadata Extraction from Art Images" focuses on content based image retrieval. 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 field of image retrieval has to overcome a major challenge: it needs to accommodate the obvious difference between the human vision system, which has evolved genetically over millenniums, and the digital technologies, which are limited within pixels capture and analysis. And the challenges are even bigger because when focus on analysis of the aesthetic and semantic content of art images. The content-based image retrieval is a technology that helps to organize digital images based on their content. This way, a variety of features of different level of conceptualization can be extracted. Typically, a content-based image retrieval system consists of three components: feature design, indexing, and retrieval. The *feature design* component extracts the visual feature(s) information from the images in the image database. The *indexing* component organizes the visual feature information to speed up the query or processing. The *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.

Chapter 4 "APICAS – Content-Based Image Retrieval in Art Image Collections Utilizing Colour Semantics" offers a succinct review of colour theory from different points of view. The rationale for that is the strong connection of any work on art paintings with the complex area of colour perception. Physiological grounds of this phenomenon are taken as a starting point for focusing the search within art painting images. A brief historical overview of attempts to define colour interconnections and mutual colour influences is made. Visual low-level features, which represent colour distribution in art images, were chosen as a ground for constructing higher-level concepts. The 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, was made. The formal description of defined harmonies and contrasts was established. A method for extracting local features that capture local colour and texture information, based on tiling the image and applying vector quantization of MPEG-7 descriptors, calculated for the tiles of the image, has been described and implemented. A program system APICAS ("Art Painting Image Colour Aesthetics and Semantics") was developed in order to supply an appropriate 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.

The final chapter "Automatic Metadata Generation and Digital Cultural Heritage" examines how automatic metadata generation methods can be used in the field of cultural heritage.

An approach for indirect spatial data extraction by learning restricted finite state automata is proposed. The realized system InDES was tested over extraction of spatial metadata from websites and shows promised results. It give assurance that such approach can be used for metadata extraction from objects descriptions and this way to be used in the process of migration from older representations of the objects in the case when the descriptions are in non-structured form.

One approach for association rule mining, which uses the possibilities of the multidimensional numbered information spaces as a storage structures is build. The algorithm ArmSquare is realized 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.

Based to similar techniques, but in the field of categorization are class-association rules algorithms. The created algorithm PGN, which is a kind of such algorithms, is also realized in PaGaNe. It was implemented in the field of analyzing semantic attributes, extracted from art images using content-based image retrieval. Within the frame of the data management and the access of the aggregator the classifier PGN can be used for enforcing information discovery.

➢ *Contributors of the Book*

This book brings together the efforts of multiple authors and their experiences in a number of domains.

Krassimira Ivanova is a Senior Assistant in the Information System Department at Institute of Mathematics and Informatics – Bulgarian Academy of Sciences. Since 2011 she is a head of the laboratory "Digitisation of Scientific and Cultural Heritage". Her main topics of interests are Data Mining, Image Retrieval, Multimedia Semantics, Data Bases, and Information Systems applied in the areas of Digitization of Cultural Heritage, Disaster Risk Management, and Analysis of Economical Processes. She has published about 70 journal articles and conference peer-reviewed papers.

Milena Dobreva is a Senior Lecturer in library, archival and information studies at the University of Malta. She was the principal investigator of EC, JISC and UNESCO funded projects in the areas of user experiences, digital cultural heritage and digital preservation. She has worked at IMI-BAS since 1990 where she earned her PhD degree in Informatics and Applied Mathematics and was the founding head of the first Digitization Centre in Bulgaria (2004). She also served as a chair of the Bulgarian national committee of the Memory of the World programme of UNESCO. Currently, she is a regular project evaluator for the EC and a number of national research bodies. She contributed to 5 books, was the lead author 5 textbooks on informatics for the secondary school, and published over 40 journal articles, technical reports and conference papers. Milena was awarded an Academic Award for young researchers (Bulgarian Academy of Sciences, 1998) and a honorary medal for contribution to the development of the relationships between Bulgaria and UNESCO (2006).

Peter Stanchev is a Professor at Kettering University, Flint, Michigan, USA and Professor and Head of Information System Department at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria. He has published 2 books, more than 200 chapters in monographs, journal and conference peer-reviewed papers, more than 200 conference papers and seminars, and have more than 600 citations. His research interests are in the field of Multimedia Systems, Database Systems, Multimedia Semantics, and Medical Systems. Serving also on many database and multimedia conference program committees, he is currently editor of chief and on the editorial boards of several journals.

George Totkov is a Professor and a Head of the Computer Science Department at Plovdiv University. Currently he is a Vice-rector on information infrastructure, quality systems and accreditation, and distance learning at Plovdiv University. His main topics of interests are computational linguistics, e-Learning, conceptual modelling, information systems, applied mathematics, etc. Prof. Totkov has been a principal investigator of over 40 national and international projects in computer science, e-learning and applications of IT in education. He authored over 200 scientific publications and 10 textbooks for the secondary and higher education.

Kalina Sotirova is a Senior Assistant and PhD student in the Information System Department at the Institute of Mathematics and Informatics – Bulgarian Academy of Sciences. Her research interests are digitization of and online access to Cultural heritage, computer graphics, EduTainment, digital storytelling, initiatives for communicating-science.

She has published over 10 scientific papers and is contributing to foster the creation of a national digitization strategy.

Juliana Peneva is an Associate Professor at New Bulgarian University, Computer Science Department and at the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences. Her research interests include information modelling, database systems, software engineering, and e-Learning. Juliana Peneva has participated in over 15 research projects and authored over 60 publications.

Stanislav Ivanov is an Associate Professor and Head of the Computer Science Department at New Bulgarian University. His research interests include object-oriented programming, computer aided design, computer graphics and e-Learning. Stanislav Ivanov took part in over 20 research projects and has more than 80 publications.

Rositza Doneva is an Associate Professor at the Department of Computer Sciences, Plovdiv University (PU) and Manager of the Regional Distance Educational Centre at PU. She has been a principal investigator or member of over 20 national and international projects in computer science, electronic and distance learning and applications of IT in education. Her main scientific interests are in the fields of electronic and distance learning, object-oriented programming, information modelling, etc. Rositza Doneva is the author of over 70 scientific publications and 10 handbooks on computer science and information technologies for the secondary and higher education.

Emil Hadjikolev is a Senior Assistant at the University of Plovdiv. His main topics of interests are information systems, digitization of cultural heritage, business process modelling, databases, programming, etc. Emil Hadjikolev authored over 10 scientific publications.

George Vragov is a Senior Assistant in the Information Systems Department, Institute of Mathematics and Informatics – Bulgarian Academy of Sciences. He works at the branch of the Institute in Plovdiv. He has an extensive experience in the field of information modelling and research and their applications to different domains. His professional and scientific research interests include web-based systems for content management, digital libraries and digitization of cultural heritage.

Elena Somova is an Associate Professor in the Computer Science Department at the University of Plovdiv. She earned her PhD degree in the field of e-learning in 2003. Her main topics of interests are distance education, e-Learning management systems, process modelling, and metadata extraction and generation in different fields including cultural

heritage. Her scientific results are published in 35 papers and 6 books. Elena Somova participated in more than 20 international, national and university projects.

Evgenia Velikova is an Associate Professor, currently Vice-Dean for the Bachelor's programs at the Faculty of Mathematics and Informatics of Sofia University "Saint Kliment Ohridsky". Her main research interests are in the field of algebra and its applications, such as coding theory and cryptography.

Iliya Mitov is a Senior Assistant in the Information System Department at Institute of Mathematics and Informatics – Bulgarian Academy of Sciences. His main topics of interests are data mining, databases, and information systems applied in the areas of digitization of cultural heritage, disaster risk management, and analysis of economic processes. He has published about 70 journal articles and conference peer-reviewed papers.

Koen Vanhoof is a Full Professor in Business Informatics at the University of Hasselt, Belgium. His major research interests lie in the areas of Data Mining, Statistics, Knowledge Engineering and Modelling, Computational Intelligence Methods, Decision Support Systems and Soft Computing Applications to Information Management, Marketing and Finance, Mobility and Traffic Safety. He has authored and/or co-authored over 60 peer-reviewed journal articles and about 7 book chapters and 60 conference papers. Currently he is Vice-Dean Research at the Faculty of Applied Economics and project leader of the Business Informatics research group at University of Hasselt.

Benoit Depaire is an Assistant Professor Business Informatics at the University of Hasselt, Belgium. His research interests focuses on Data Mining, Data Analytics, Data Modelling and Statistics within the field of Business Studies. Recently, he expanded his research to the domain of Business Process Modelling and Business Process Mining. He has authored and/or co-authored over 15 peer-reviewed journal and conference papers.

Dimitar Blagoev is an Assistant Lecturer at the University of Plovdiv and a Senior Assistant in the Information System Department at Institute of Mathematics and Informatics (Bulgarian Academy of Sciences). His main topics of interests are Computational Linguistics, Business Process Modelling, Data Mining and Image Retrieval.