

Development of New Solutions in the Field of Digitization and Digital Presentation of the National Folklore Heritage

Galina Bogdanova, Radoslav Pavlov

Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences, Sofia, Bulgaria
galina@math.bas.bg, radko@cc.bas.bg

Abstract. This article briefly reviews the software developments for digital presentation and preservation of Bulgarian folklore treasure created within the project “Knowledge Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage” by teams of the Institute of Mathematics and Informatics.

Keywords: Multimedia, Knowledge Technologies, Digitized Objects, Digital Archive, Digital Libraries, Information Portal, Cultural Heritage

1 Introduction

The digitization of cultural heritage has been a priority in Bulgaria in recent years. This paper discusses applied scientific research of folklore artifacts, presenting investigations done within the national project „Knowledge Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage” (FolkKnow)¹ by teams from the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences (IMI-BAS). Among the partners in the project are the Institute of Ethnology and Folklore Studies with Ethnographic Museum at the Bulgarian Academy of Sciences (IEFEM-BAS) and the St Cyril and St Methodius University of Veliko Tânovo (VTU).

FolkKnow aims to build a multimedia digital archive “Bulgarian Folklore Heritage”, digital library and virtual information portal with folk multimedia objects and selected collections of the fund of the National centre for non-material cultural heritage, IEFEM-BAS.

The project comprises four interrelated research modules:

Module 1: Development and exhibition of fund “Bulgarian Folklore Heritage” (BFH), № IO-03-01/2006 (IEFEM-BAS, V. Mateeva);

Module 2: Development, annotation and protection of a digital fund, IO-03-02/2006 (IMI-BAS, G. Bogdanova);

¹ Contract IO-03/2006, Information Society Programme (leader: Assoc. Prof. Dr Galina Bogdanova), partly funding this article.

Module 3: Development of digital libraries and information portal with virtual expositions „Bulgarian Folklore Heritage", IO-03-03/2006 (IMI-BAS, R. Pavlov);

Module 4: Socially oriented applications of the BFH in the education, scientific researches and cultural tourism, IO-03-04/2006 (VTU, G. Todorov).

FolkKnow is an interdisciplinary research project aimed at the discovery and development of new solutions for digitization and virtual presentation of artifacts and collections of Bulgarian folklore heritage. The target group of users covers professionals and scientists, non-professionals, connoisseurs and viewers of the domain. The digital libraries and Semantic web technologies are used.

2 FolkKnow Research and Developments

2.1 Creating, Annotating and Protecting a Digital Archive

Using modern methods, tools and techniques for digitization of multimedia objects, folklore artifacts were studied, collections were selected and a digital archive “Bulgarian Folklore Heritage” was built. The most appropriate technologies and knowledge-based methods for describing intangible digital resources were researched and implemented and tools for security and for optimizing the access to multimedia data archives were designed [1-6].

Techniques and methods for digitizing diverse folk multimedia objects (text, images, audio and video) were used. Semantic and structural analysis of the folk materials and a selection of collections were carried out. Standards for digitization were examined and used as a basis for setting requirements for the parameters of files, aiming at preserving optimal information on the artifacts. A Matrix for describing folk artifacts was built.

An organization of metadata was implemented in order to index the content of digital resources and the digitization process. Metadata were prepared on the digitization and display of basic and specific parameters for indexing the multimedia content. A technology was chosen for adding metadata attached to digital resources. Folklore objects (text, audio, video, photo) were digitized and described. A digital archive of objects with added metadata and security was organized.

The parameters of file formats used for WEB and the methods used for digitization (see Table 1) were studied and presented in detail in [1-6].

Table 1. Parameters of files

| Parameters | Parameters of the file formats with priority on information details | Parameters of the file formats for resources used for the WEB |
|-------------------|---|---|
| PHOTO: | | |
| File size: | From 2 MB to 50 MB | Less than 1 MB |
| Resolution: | From 1200x800 to 6000x4000 px | 800x600 px |
| Resolution: | From 600 dpi to 2000 dpi | 600 dpi |
| Number of colors: | More than 16 bit color | More than 16 bit color |

| | | |
|---|---------------------------------------|---------------------------|
| File formats: | BMP, TIFF, PNG or JPG | PNG or JPG |
| VIDEO: Parameters for digital video recording | | |
| System for streaming video | PAL | |
| Number of frames per second | 25 | 300x200 px |
| Resolution of picture | 720x576 px | |
| Ratio of picture | 4:3 | |
| Resolution of picture | From 2000 Mbps to 4000 Mbps | |
| Ratio of picture | DVD-Video, DV-AVI, MPG or AVI | Stream, MPEG-4, FLV |
| Bitrate | Up to 500 MB | Up to 20 MB |
| File formats: | | |
| File size: | | |
| AUDIO: | | |
| File formats: | WAV | MP3 |
| Audio format: | PCM | |
| Sample rate: | 44 100 Hz | 44 100 Hz |
| Bit resolution: | 16 bit | |
| Bitrate: | | From 128 kbps to 320 kbps |
| File size: | Up to 100 MB | Up to 10 MB |
| TEXT: | | |
| Table of characters: | Unicode, Windows Cyrillic 1251, UTF-8 | |
| File formats: | TXT, DOC, PDF, RTF, HTML | |

2.2 Bulgarian Folklore Digital Library

With the use of modern methods, tools and techniques for storage and digital representation of knowledge on intangible cultural heritage in Module 3 (see [5, 8–10]) the following research results were achieved:

- Building semantic-oriented representations of selected subject areas of the Bulgarian folklore heritage with the approach and means of the Semantic Web technology. Ensuring an opportunity for further expandability and new applications of the ontological structure.
- Multimedia presentation, metadata and semantic annotation of the objects of the BFN digital archive for the purposes of the project.
- Developing an architecture and implementation of a digital library containing digital multimedia folklore objects annotated according to the created ontological structure.
- Introducing folklore objects/documents in a digital library.

The created folklore digital library aims at a new modern representation, in a virtual form, of significant phenomena of the Bulgarian folklore heritage and provides their innovative exposure in the global information space. For the development a service-based software architecture was used, ensuring rich functionality and a cost-

effective environment tailored according to the target user group. Its main services are:

- Content creation: add (annotate and semantic indexing), store, edit, preview, delete, group, and manage multimedia digital objects (images, text, sound, video) and collections;
- Metadata description, metadata management;
- Preview, browse, navigate digital objects, collections and their descriptions;
- Information retrieval and mining;
- Search (simple, extended, semantic, context-based, etc.), filtering, selecting and grouping objects;
- User modeling/Personalized and adaptive access to the content
- Protection and preservation of the digital content;
- Administration and tracking services, etc.

To achieve the results, extensive research was carried out of the modern tools, standards and languages for creating semantic descriptions and indexing, of the semantic structure and properties of the target area. The basic ontological structure constructed reflected maximally the characteristics and specifics of the subject area and described objects. The semantic tagging of objects was at first conducted as a test using the standard tools of the Semantic Web outside the environment of the digital library developed in parallel. At the next stage the ontology served as a basis for the annotating tree template of the constructed application “Bulgarian Folklore Digital Library” (<http://folkknow.cc.bas.bg/>). The services provided by the environment were created by a flexible concept that allows continuous update and development of the functionality in line with the current trends in the construction of digital content management systems, Web 2.0, etc. Numerous experiments were carried out with multimedia collections of objects from the target domain.

2.3 Preservation, Protection and Access to Data and Additional Information Sources

In the project FolkKnow methods for protecting information and providing additional sources and bibliographic support of the digital archive and library [1-6] were tested and implemented, including:

- New steganographic methods of software protection of data.
- Specialized steganographic software for protecting MP3 files [1]. Text is embedded into MP3 files by Padding Byte Stuffing, whereby two functions are offered: adding text and text retrieval. The text is pre-coded by the lzw algorithm to save space.
- Steganographic software for protecting images from unauthorized distribution. MATLAB modules for protection by watermark are included, using protection techniques with expanding spectrum, capabilities for embedding and extracting a watermark in static images, noise-protection encoding before embedding using a developed scheme [4, 6].

- Software information compression methods (for sound files).
- A specialized frequency dictionary—hierarchical data structure/tree with included tables for administering sections/categories, allowing growth in volume and depth.
- A system for research publications on folklore topics, allowing storing, indexing and searching data on folklore rubrics according to the Description Matrix, with consideration of their relationship to the created digital folklore archive FolkKnow and with constructed levels of access and security.

3 Information Artery for Bulgarian Folklore. Socially Oriented Applications

During the final stage of the project an information artery for Bulgarian folklore and traditional culture was set up, with ensured access to folk artifacts through the digital library, a function module for grouping objects into collections, information modules for dynamic presentation of events, current projects, publications, etc., in the domain of digital representation and preservation of cultural heritage in Bulgaria (<http://artery.bg73.net/home>).

The main components of the Bulgarian folklore artery are:

- Bulgarian Folklore Digital Library, providing access to:
 - A module for adding and editing folklore objects. The library expects as an in-put two types of objects: simple folklore objects and complex folklore objects.
 - A module for viewing the content of folklore objects (according to their base type and rubric to which they belong or by different descriptive characteristics).
 - A module for searching by: signature and archive number; keywords of the following categories: name, language, annotation, type of the folklore object/rubric; file type; record information (simultaneously or one by one): by situation, by reporter name, by recorder name, by record date and by recording location; extended search – it provides the option for searching through all the object characteristics;
 - A module for managing the user data;
 - A module for monitoring the user's actions, which keeps track of the following: a) Actions related to working with the system; b) Actions related to the object manipulation; c) Actions related to the content viewing; d) Other administrative actions.
 - A module for file format conversion;
 - A module for generation of XML copies of the objects in the system.
- Virtual Expositions of the Bulgarian Folklore: This section shows groups of Bulgarian folklore objects, separated as collections with interactive preview and social networking possibilities (i.e. user tagging, commenting, liking, following & book-marking, rating, etc.). The source of the objects is the Bulgarian Folklore Digital Library, where the expositions' creation and management is executed by folklore specialists.

- eLearning Corner: This section realizes the relation between the Bulgarian Folklore Digital Library and its distant learning application.
- Actual Folklore News: This section aims to contain current and actual news from the Bulgarian folklore and cultural heritage domain. It will also keep all older news in an archive. Every news record could be tagged, commented or liked by the artery users.
- Folklore Discussion Forum: This section provides possibility to joint and communicate different users (i.e. specialist and non-specialist), dedicated to Bulgarian folklore and traditional culture. The access to the forum will be achieved through registration and authorization. The users have to accept terms and conditions to use the forum.
- Folklore and Cultural Heritage Conferences and Science Forums: This section provides actual information for the conferences, science forums and initiatives for digitization and virtual presentation of the folklore and cultural heritage domain.
- Bulgarian Folklore in the Social Networks: The Bulgarian folklore artery will actually attend in the social network like Facebook and Twitter, through implementation of Web 2.0 tools and services. This section has a key role in the wide popularization and advertising of the Bulgarian folklore artery in the global information space.

Investigations were carried out to create socially oriented applications in education (including Interactive distance learning/self-learning), research and cultural tourism. [7]. An analysis of the needs, properties and profile of the target user group was performed, as well as a survey of the requirements and structuring of the components into virtual expositions from the point of view of their use in informal learning, an analysis of the folklore content and objects for cultural tourism, etc.

4 FolkKnow Results and Their Dissemination

Within the project research was carried out, whose results are summarized as follows:

- Building an annotated and protected digital archive BFN, a digital library and information artery with socially oriented applications; selection, preparation and generation of digitized samples; development of requirements and standards for digitization, annotation schemes and metadata on artifacts; a matrix for describing folk-artifacts, a matrix for describing digitized files, etc.
- Steganographic methods for protecting information, methods for software compression of information (for audio files), a specialized frequency dictionary and a publication system were developed for optimizing the digitization of folklore objects, storing, protection and access to data.
- Training specialists on the topic of the project: three PhD and six MSc theses defended (supervisors G. Bogdanova and R. Pavlov); an MSc and trainee programme in the area of digitization and digital library were implemented.
- Organization of joint seminars, meetings, round tables, agreements to promote the results of FolkKnow and search for new joint project implementations. A regular

scientific interdisciplinary seminar “Information Society” with mobile hosting and virtual lectures was created and grew into a national research network with participants from over 20 schools, institutes, libraries, museums, etc. (<http://ioseminar.co.cc/>). A First and Second international conferences “Digital Presentation and Preservation” (DiPP2011) (11-14.09.2011, Veliko Târnovo) and DiPP2012 (18-21.09.2012) were organized.

The dissemination of the results was done through reports and publications at prestigious national and international scientific conferences and fora, in scholarly publications, the Internet and through specialized training sessions.

5 Conclusions

The fundamental and applied research carried out within the project FolkKnow is of national importance for the digitization of the wealth of the heritage, the preservation of endangered artifacts and documents and ensuring broad and efficient access and durability of the information for BFN. The created applications support the archiving and virtual presentation of valuable collections of Bulgarian folklore, allowing their further exposure in the global network. The results achieved by FolkKnow are applicable to the construction of other similar environments and are oriented towards the creation of a single network for digitization and digital preservation of intangible cultural heritage in Europe and worldwide. The need for continuing and for wide dissemination of the project results led to their inclusion and further development in numerous national and international programmes and projects.

References

1. Bogdanova, G., Zahariev, G.: Steganographic Methods for Embedding Information into MP3 Audio Files, *New Challenges in the Technologies of Programming, Computational Algorithms and Learning in Higher Education*, Apriltsi'2009, 85–93 (2010)
2. Mateeva V., Stanoeva, I.: Classification Scheme of the Typological Catalogue at the Folklore Institute. *Bulgarian Folklore*. № 2–3, 96–109 (2001)
3. Mateeva, V.: Folklore Heritage and Digital Archive. *Bulgarian Folklore*, №3–4, 1–219 (2010)
4. Bogdanova, G., Georgieva, Ts.: Using Error-correcting Dependencies for Collaborative Filtering. *Data and Knowledge*, 66, 3, Elsevier (2008)
5. Bogdanova, G, Pavlov, R., Todorov, G., Mateeva V.: Knowledge technologies for creation of digital presentation and significant repositories of folklore heritage). – *Advances in Bulgarian science*, Sofia, 7–14 (2006)
6. Bogdanova, G., Todorov, T., Georgieva, Ts.: Algorithms for security and analization of experimental multimedia archive. *Math. Balcanica, New Series*, 21, (3–4), 225–232 (2007)
7. Doshkova-Todorova J., Todorova, M., Valcheva, D.: Social-oriented Applications of the Folklore Heritage for the needs of the Non-formal Learning and Cultural Tourism, In: *First International Conference on Digital Preservation and Presentation of Cultural and Scientific Heritage*, Veliko Tarnovo, Bulgaria, 51–59 (2011)

8. Luchev, D., Paneva, D., Rangochev, K.: Use of Knowledge Technologies for Presentation of Bulgarian Folklore Herigage Semantics, International Journal „Information Technologies and Knowledge”, 2(4), 307–313 (2008)
9. Paneva-Marinoва, D., Pavlov, R., Goynov, M.: Business Modeling of the Application Architecture of the Bulgarian Folklore Artery, In: First International Conference “Digital Preservation and Presentation of Cultural and Scientific Heritage, Veliko Tarnovo, Bulgaria, 43–50 (2011)
10. Paneva, D., Rangochev, K., Luchev, D.: Ontological Model of the Knowledge in Folklore Digital Library, In: Fifth HUBUSKA Open Workshop "Knowledge Technologies and Applications", Kosice, Slovakia, 47–57 (2007)