

# Information Technologies for Presentation of Bulgarian Folk Songs with Music, Notes and Text in a Digital Library

Lozanka Peycheva<sup>1</sup>, Nikolay Kirov<sup>2</sup>, Maria Nisheva-Pavlova<sup>3</sup>

<sup>1</sup> Institute of Folklore, Bulgarian Academy of Sciences  
lozanka.peycheva@gmail.com

<sup>2</sup> Department Informatics, New Bulgarian University  
and Institute of Mathematics and Informatics, Bulgarian Academy of Sciences  
nkirov@nbu.bg

<sup>3</sup> Faculty of Mathematics and Informatics, Sofia University  
marian@fmi.uni-sofia.bg

**Abstract.** The paper presents the main aspects of an ongoing project aimed at the development of technologies for digitization of Bulgarian folk music and building a heterogeneous digital library with Bulgarian folk songs presented with their music, notes and text. The digitization of sound recordings and notations of folk music will facilitate the access to them, will provide their preservation in another quality and will present the opportunity for these recordings and notations to be used by a greater number of users through web-based access. The digital library will provide both digital preservation of the sound recordings and notations of Bulgarian folk songs and a possibility for new interpretations of the archaic Bulgarian folklore heritage. *Information technologies for presentation of Bulgarian folk songs* is an interdisciplinary project, which imposes a specific research approach, combining the plurality and dialogue of a multitude of viewpoints, positions and perspectives.

**Keywords:** Digitization, Digital Preservation, Digital Library, Bulgarian Folk Songs

## 1 Introduction

In the recent years informatics and information technologies have played an important role in a number of successful projects oriented towards the digital preservation of manuscripts, printed documents, photographs, audio materials etc. that are regarded as significant scientific and cultural heritage. At the same time the number of originally created in digital format documents, publications, books etc., which are important for the development of the public sphere and the industry have increased at high rates.

The availability of an enormous quantity of different materials on the web makes them difficult to access and thus it is necessary to organize them in the form of the so called digital libraries. In the past decade a great number of good results have been



achieved in the sphere of creation of digital libraries and the development of tools for access to their content. In particular, the project for an European digital library [1] was successfully implemented and in 2009 an initial version of a World digital library was launched [2].

At the same time the creators of digital libraries and tools for access to their content are still faced with numerous challenges. An example of such a challenge is the speed of fulfillment of user queries. Another challenge is the provision of maximum accurate and thorough responses to questions and queries of the users. As another serious problem can be viewed the necessity for definition and implementation of appropriate tools for search of a number of various by type and content materials accessible in different digital formats. In connection to that, the problems of the development of suitable methods and software tools for supporting the experts in different fields in the interpretations of the results from the search in big digital libraries is becoming an urgent matter.

The paper discusses the main activities planned within the scope of a project titled "Information Technologies for Presentation of Bulgarian Folk Songs with Music, Notes and Text in a Digital Library" which sets as its task to carry out a series of researches oriented towards the development of technology and an according supporting software tools for the creation and the adequate usage of heterogeneous institutional digital libraries. The practical tests and the pilot implementation will be performed on the basis of a collection of Bulgarian folk songs presented by notes, lyrics and music.

So far the most considerable contribution to the process of digitization, preservation and presentation of folk materials has been done by a project titled "Creation and presentation of a Bulgarian folk heritage collection", that was developed by experts from the Institute of Folklore and the Institute of Mathematics and Informatics of BAS. In the last years new information technologies arise which improve and simplify digitization process (compare with [3]), and increase the potential for supporting more complex and flexible data bases and digital libraries.

## **2 Necessity of Digitization of Manuscripts (Music Notations and Lyrics) and Recordings of Bulgarian Folk Music**

Bulgarian folk music is a valuable resource of cultural memory and is among the main characteristics of the national identity of Bulgarian people. Throughout the years the Bulgarian researchers of musical folklore have written down in lyrics and notes hundreds of thousands of musical folk samples (songs and instrumental melodies). Part of these music notations have been published, another part is preserved as manuscripts in specialized institutional or personal archives. The major part from the available music notations with Bulgarian folk music however is on paper. This, on the one hand, makes the access to them difficult and, on the other, bears the risk of their loss (especially, in the case where the notations haven't been published and exist in only one copy). Today's information technologies give us the possibility to digitize the existing collections of folk music records. This will facilitate the access to them,

will provide their preservation in another quality and will present the opportunity these records to be used by a greater number of users through web-based access.

The lyrics of the folk songs are also a valuable informational archive – they carry an affluent information for the lifestyle of the Bulgarian people during the centuries and for their mentality as well (religious beliefs, notions, cultural practices). The problem of the lexical composition of the Bulgarian folk song and its internal lexicological structure is still not solved. It is clear that the folk song lexis is completely different from the contemporary (and not only) colloquial and literary lexis, but this difference is still far from defined. The creation of the different types of indicators will be a starting point of many new scientific researches in different fields: linguistics, ethnology, ethnomusicology etc.

We propose a modern informational processing of recorded Bulgarian folk songs. The sound recordings of the songs in their authenticity were made mainly in the 1960-70s and are available only on old magnetic tapes. The authentic and unique field recordings collected in the course of decades from a number of Bulgarian folklorists (and other researchers) are preserved in the Musical Folk Archive in the Institute for Art Studies – the most significant collection of metadata for musical folklore in Bulgaria – mainly on magnetic tapes, the oldest of which date back to 1954. This type of preservation is not long-lasting and is in urgent need of digitization which will preserve the tapes for the future generations with the purpose of maintaining the cultural heritage of Bulgaria. The recorded tape is being damaged every time it is played – stretched, crumpled and demagnetized and often torn. This leads to deterioration of the material quality and to loss of particular fragments, or even whole samples and tapes. Thus unique copies are being lost forever. Contemporary information and communication technologies give the opportunity and the means to digitize these old music recordings, which is the only way for their preservation from a modern point of view and for making them accessible to scientific teams, educational institutions and the public. The process of digitization of old tape recordings is mainly focused on three technologies: transfer, mastering and conversion from Audio to MP3 files.

### **3 Some Achievements in Digitization of Folk Music**

Currently on Bulgarian internet sites one can find music notes for Bulgarian folklore that are only scanned versions of the old manuscripts or parts of books. At the same time, there are a lot of musical archives all over the world, digitized with contemporary means and technologies. The emergence of an archive with Bulgarian folk music undoubtedly will provide for the incorporation of the Bulgarian cultural heritage to the world accomplishments in that sphere.

Special attention should be paid to the “Multimedia database for Bulgarian folk music” of the Institute of Art Studies of BAS, which consists of more than 15 000 records of song lyrics, analyses of ethno-musicologists, audio recordings, music notes and photographs and videos of Bulgarian folk music. Every record has a description – 23 fields of metadata and a few lines of the song lyrics. 1702 of the records also

consist of a brief musical recording of the original performance – “audio samples”, and 4633 – of a scanned copy of the piece of paper from the researcher.

In the summer of 2008 in Bucharest (Romania) a new European project was presented (with the participation of four countries – Switzerland, Germany, Hungary and Romania) on the development of information technologies for digitization of ethno-musicological archives (these are archives that preserve unique musical documents with great aesthetic and scientific value and serve the audio-visual memory and cultural variety). The project EthnoARC [4] provides a valuable experience, which reveals how it is possible to systematically present in linked archives the otherwise heterogeneous ethnomusicological archives in the cities of Berlin, Budapest, Geneva and Bucharest with the help of information technologies. This project is a practical confirmation of the increased necessity for the development of specific software tools for constructing of metadata in order to have an adequate, easy and convenient access to the unique musical folk documents preserved in the archives of the Institute of Art Studies and the Institute of Folklore in Bulgaria.

#### **4 Digitization of Manuscripts and Books and Creation of a Database with Bulgarian Folk Songs**

The digitization of notations is a specific and labour-consuming work. For the present moment the authors have around 500 digitized songs from the archive manuscripts of Prof. T. Dzhidzhev [5,6], which consists of field recordings of over 1000 folk songs from the Tracia region. Because the paper is getting old, it is important to preserve these songs in due time through their digitization in order to avoid the potential loss of information over time.

LilyPond [7] is a system for digitization of music records under GNU license. It is accompanied by a detailed documentation and an active mailing-list. The system is dynamically developing and its capacity is great – in practice it can be used for writing various notations – from orchestrations for big orchestras to Gregorian records of clerical music and special notations for bagpipe. The music records can easily be turned into PDF and PS files suitable for printing, EPS and JPG formats for inclusion in books and publications, as well as XML formats for communication with other software and databases. The syntax LilyPond is based on print preparation systems TeX and LaTeX. The programming language is C++, the software is structured in libraries, which are managed through Scheme scripts. There is quite a complex parser. The system is expandable for various specific needs with Scheme programming language, which is one of the two main dialects of the Lisp language.

*Technology for coding of note records from original manuscripts with field recordings of Bulgarian folk songs*

Despite the great LilyPond capacities, there are specific symbols and other elements necessary for notation of the Bulgarian national music. For example, the variations in some tones, bars and song parts, as well as signs for increasing or decreasing of the tone pitch cannot be written down with the available tools from the system. In line with the rules of the open code these elements can be proposed to the developers for inclusion in LilyPond and thus can become accessible for everyone.

*Technology for scanning of printed books with BFS and OCR of the notation in them*

This task encompasses the organization of technology for scanning of printed materials, mainly books, that contain Bulgarian national songs. After the scanning it is necessary to divide notes from lyrics and an identification (OCR) to be made on notes and lyrics separately. The difficulty in recognizing the notes is connected to the specific characteristics of BFS and obstacles will emerge for the lyrics also, because of the different dialects in the Bulgarian language. Besides, many of the words are written according to their phonetic sounding and not by the rules of the contemporary Bulgarian language.

*Creation, support and completion of database for preservation of notes, lyrics and music recordings of BFS*

The modern way of preservation of data is connected to their organization in structured (relational) database. The DB of Bulgarian national songs, which will be created by the project team, will consist of three basic components:

1. Notations of songs in LilyPond format;
2. Lyrics of songs together with additional information for the researcher, place and date of the research, name of the researcher, notes from the researcher;
3. Musical (MP3) files with an authentic performance of the song and musical files generated from notations with LilyPond system.

## **5 Information Technologies for Building and Using Digital Libraries**

The aim of the research activities in this direction is to develop and adapt appropriate computer science methods, approaches, algorithms and tools for solving various problems concerning the analysis of folklore materials – methods for access to databases with music records and rhymed lyrics, analysis and effective (in particular, semantics oriented) search of specific regularities in the music and words of the songs etc. We expect to obtain some useful results in two main aspects: on the one hand, giving proper information technologies and software tools in order to assist experts of folklore in their professional activities; and on the other hand, to bring about new research topics in the field of Computer Science and Information Technologies, originating from the particular application.

*Methods and tools for building heterogeneous digital libraries with folklore music*

The objective of this task is to suggest a functional model of digital library with folklore music. Methods for building the particular components of such library and software tools for their integration will be developed as well. The experience and some former results of team members from the Faculty of Mathematics and Informatics of Sofia University [8] will be used for the purpose.

*Semantics oriented search in digital libraries with folk music*

This task has the aim to satisfy the necessity for the development and testing of a proper methodology for application of advanced information technologies (e.g., Semantic Web technologies) in building specific software tools for semantics oriented

search in the provided digital library with folklore music. These tools will support an adequate and flexible access to the complete palette of resources stored in the library.

*Knowledge extraction from databases with Bulgarian folk songs*

An attempt at formal conceptualization of a properly selected set of terms in the area of musical folklore will be made and a specific method for data mining and knowledge extraction from the database with Bulgarian folk songs will be proposed.

## 6 Conclusions and Future Work

The proposed database (the repository of the digital library) consists of the songs' notations in LilyPond format [7], the full lyrics of the songs, musical files, created by LilyPond and musical recordings of authentic performances (as far as such exist in the archives). This way the metadata will be extractable from the notes and from the lyrics of the song, for example bar measure, musical form, tone range, tone scale. Besides this, with the help of software tools scientific analyses will be able to be performed based on the songs notations and lyrics in the sphere of folkloristics, ethnomusicology and linguistics.

Our work is multidisciplinary, with the information technologies having the leading role for obtaining scientific results in connection with the solution of different problems arising during the process of creation of heterogeneous institutional digital library and the development of tools for the realization of different types of search in such a library, including semantics oriented search. Special attention will be paid to research dedicated to the creation of new methods and means for generating hypothesis and extraction of knowledge in the sphere of folkloristics. The scientific results in the sphere of ethnomusicology and linguistics would not be possible without sufficient collected information and software tools for their processing. For example, one ambitious task in folkloristics is to research the relationship between the lyrics of the song and the musical phrase.

The complete and up-to-date digitization – notes, lyrics, music – will complete and further develop what the experts from BAS have already achieved in terms of preserving the Bulgarian folklore. The application of European standards in the process will provide for the incorporation of the Bulgarian cultural heritage to the world's achievements in that area.

**Acknowledgments.** The project *Information technologies for presentation of Bulgarian folk songs with music, notes and text in a digital library* has been supported by Grant of the Bulgarian National Science Foundation under number DTK 02/54.

## References

1. European Digital Library Project, <http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/>
2. World Digital Library, <http://www.wdl.org/>

3. Kirov, N., Problems and Achievements in the Digitalization of Bulgarian Folklore Music, (Proc. Int. Congress MASSEE'2003, September 15-21, 2003, Borovets, Bulgaria), Review of the National Center for Digitization 4, Faculty of Mathematics, Belgrade (2004), pp. 78–83.
4. Linked European Archives for Ethnomusicological Research, <http://www.ethnoarc.org/>
5. Peycheva, L. Stroitel na suvremennata bulgarska muzikalna folkloristika. Todor Djidjev na 80 godini [Builder of the contemporary Bulgarian Musical Folkloristics. Todor Djidjev's 80-th Anniversary]. Bulgarski folklore [Bulgarian Folklore], 2007, ISSN 0323-9861, 33/2, pp. 123–136.
6. Vlaeva, I. Todor Djidjev – terenistut [Todor Djidjev, the Fieldworker]. Bulgarsko muzikoznanie [Bulgarian musicology], 2009, ISSN 0204-823X, 33/3-4, pp. 90–101.
7. LilyPond ... music notation for everyone, <http://lilypond.org/>
8. Nisheva-Pavlova, M., P. Pavlov. Building a Digital Library with Learning Materials. In: S. Mornati, T. Hedlund (Eds.), "Rethinking Electronic Publishing: Innovation in Communication Paradigms and Technologies. Proceedings of the 13<sup>th</sup> International Conference on Electronic Publishing (Milan, Italy, June 12-14, 2009)", Edizioni Nuova Cultura – Roma, 2009, ISBN 978-88-6134-326-6, pp. 471–483.