Nikolay Kirov
New Bulgarian University

DIGITIZATION OF BULGARIAN FOLK SONGS WITH MUSIC, NOTES AND TEXT

Abstract. A digitization project for Bulgarian folk songs *Information technologies for presentation of Bulgarian folk songs with music, notes and text in a digital library* was started last year, joining the efforts of various experts from three institutes of the Bulgarian Academy of Sciences, Sofia University and New Bulgarian University. The research that is carried out under this project is aimed at the development of a technology and corresponding supporting software tools for the creation and usage of heterogeneous institutional digital libraries. The tools will satisfy the needs of the researchers for information technologies in the fields of ethnology, ethnomusicology and folkloristic. In the project frame a technological environment for digitization of notations is created, specially adapted for Bulgarian folk songs. Now a database with notes, lyrics and music is under development. An initial digitization and preservation of the Bulgarian cultural heritage will be carried out by means of digitization and insertion into the system of over 1000 songs that were recorded and written down during the 60s and 70s of XX century.

Keywords: Digitization, Bulgarian folk songs, digital library.

The Project

A digitization project for Bulgarian folk songs *Information technologies for presentation of Bulgarian folk songs with music, notes and text in a digital library* was started last year, joining the efforts of various experts from:

- Institute of Mathematics and Informatics, Bulgarian Academy of Sciences,
- Institute of Art Studies, Bulgarian Academy of Sciences,
- Institute of Folklore, Bulgarian Academy of Sciences,
- Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski",
- Department of Informatics, New Bulgarian University.

The research that is carried out under this project is aimed at the development of a technology and corresponding supporting software tools for the creation and usage of heterogeneous institutional digital libraries. The tools will satisfy the needs of the researchers for information technologies in the fields of ethnology, ethnomusicology and folkloristic.

In the project frame a technological environment for digitization of notations is created, specially adapted for Bulgarian folk songs. Now a database with notes (coded with LilyPond software [5]), lyrics and music (digitized from old magnetic tapes) is under development. An initial digitization and preservation of the Bulgarian cultural heritage will be carried out by means of digitization and insertion into the data base of over 1000 songs that were recorded and written down during the 60s and 70s of XX century.

The main activities under the project are distributed in eight working packages (WP). The majority of WP tasks are under development in the current, already funded project but others (marked by *) will be worked out in the next stage.

WP1: User requirements, market surveys and standards

1.1 Identification of potential users and their requirements for web-access and usage of the digital library with Bulgarian folk songs (BFS).

1.2 Market surveys for existing technological solutions and software tools for project tasks and objectives.

1.3 Metadata for folk songs.

A suitable multitude of metadata for the components of the folk songs (lyrics, note records, musical recordings) will be determined taking into account the internationally recognized classification schemes and the national traditions.

1.4 Integration and adaptation of existing standards.

Digitization of the data model will be performed, based on the integration and possible adaptation of the existing European and world standards (see [7]).

WP2: Digitization of manuscripts and books and creation of a database with BFS (see [4] for details).

2.1 Technology for coding of note records from original manuscripts with field recordings.

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

2.3 Creation, support and completion of database for preservation of notes, lyrics and music recordings (see Mutopia Project [8]).

WP3: Project website and Internet access to the database

3.1 Project site with the description of database, technologies for database completion and tools for searching and knowledge extraction.

The website will be provided with a description of the project, public documents on its implementation, description of the database of BFS, description of the developed technological environment for the introduction of musical notes with the specialized for BFS tools. Instruments (tools) for working with the database will be described.

3.2 Web-based access and visualization of the note records and the song lyrics (see [8]).

3.3 Web-based access to metadata and musical objects of the database.

The metadata in the database will be provided with public access via Internet. Every user will be able to get (according to the international standards) XML files with metadata.

3.4* Visualization of geographic objects connected to geographic elements in the database.

The database consists of names of settlements (mainly villages) in Bulgaria, where the songs have been recorded. These geographical data will be presented on the map of the country. On this map parts of the database that will come as a result of a search with set criteria, will be presented.

WP4: Research in the field of information technologies for building and using of digital libraries with BFS (see [4])

4.1 Methods and tools for building heterogeneous digital libraries.

4.2 Semantics oriented search in digital libraries.

4.3 Knowledge extraction from databases.

4.4* Software tools for special purpose (in particular, semantics oriented) search and knowledge extraction from a database.

WP5: Scientific research in ethnomusicology (see [3] for details)

5.1* Completion and diversification of the collection of BFS with newly "on the field" recorded samples.

5.2 Development of a scheme for classification of folk songs with respect to technical data (settlement, researcher, informant).

5.3 Classification of the songs with respect to their implicit musical characteristics: tone scale; tone range; metro-rhythmical features (bar measures).

5.4* Classification of the songs according to their cultural functions.

WP6: Scientific research of verbal folklore

6.1 Creation of a set of phonetic variations, which any sound can undergo according to the Bulgarian lexicology and dialectology.

The task is aimed at ensuring the creation of an "intelligent" search with sufficient information, which finds the wanted word despite the change of a sound in the root of the word (for example: юнак, юначен, юнаци) and despite the dialect specifics in writing. The most potential changes in every sound will be taken into account according to the data from the Bulgarian lexicology and dialectology. A review of the different solutions that folklorists use for fixing the differences between dialects will be carried out.

6.2 Creation of frequency dictionaries over various sets of data: from one collection; from one region; from one settlement; from a particular year of recording, etc.).

6.3 Creation of frequency dictionaries – concordances.

The concordance is a variation of the frequency dictionary that is aimed at detaching the context, in which the given word is being used. Here is one imaginary example: we search for the word *wine* and in the lyrics it is most often found near the words: *rakia, pour, table, Marko, three-year-old*, etc. There are also unexpected appearances such as *wine-river*, which represent a deep fundamental connection in meaning.

6.4* Classification of songs by implicit lyrics, genre and statistic characteristics.

This task will allow for theoretical observations to be made on the question whether any fundamental relationship between different groups of songs from the different classifications exists. In this way statistically we can present and illustrate what kind of songs and in what context are performed (for example: epic – at the table, mythical – on Christmas time, etc.), if there is any regularity in this and thus if there is any fundamental relationship between results from different classification strategies (for example genre – bar) etc.

6.5* Theoretical aspects of giving a new meaning to the notion "genre" in folklore.

Observations of the stable relationship between lyrics (content) and context (time and place of performance) can give an entirely new meaning of the notion "genre" in folklore and lead to a formulation of a new definition in terms of the two categories' communicative intention and cultural function (see 5.4).

WP7: Information technologies for digitization of music recordings (see [2] for details)

Field recordings of the most significant music folklore collection in Bulgaria, situated in the Music Folk Archive of the Institute of Art Studies, BAS, are preserved on magnetic tapes. This way of preservation is not long-lasting and urgently needs restoration through digitization - currently it is the only way to preserve and make the access to them easier for scientific research.

7.1 Expert activities (preliminary and basic).

7.2 Transfer (analogue-digital conversion).

7.3 Mastering.

7.4* Conversion of audio information to MIDI impulses.

7.5* Theoretical analysis of the differences between the subjective notation (made by a human) and the authentic tone picture, and of the basic principles for the production of software for folk music notation.

WP8: Dissemination of the project results

8.1 Dissemination of the results on the Internet.

8.2 Dissemination of the results among scientific institutions in the country and abroad.

8.3* Organization of a training seminar for working with the LilyPond software for digitization of notes.

8.4* Dissemination of the results among the educational institutions and the public.

8.5 Publications.

The most important publication will be the printing of a book with collection of notated folk songs that were gathered during the 60s and 70s by Todor Dzhidzhev. The book will be completed with a CD of authentic musical recordings of songs. The ambition of the team is to provide the book with indices for the main criteria for classification of folk songs and to turn it into a model to be followed in the preparation of similar collections.

Example

The image of an original page from a manuscript which contained a song, is presented here:

The code of notes in LilyPond is:

```
\include "td-preamble.ly"
\score {
\relative c' {
 \tempo 4 = 144
 \time 2/4
d4 f | \acciaccatura f8 g2 | d4 f | g2 |
\override Glissando #'style = #'zigzag
\varA
 \acciaccatura  a8\startTextSpan  \noteFi  g4\glissando  f16(  ees  c8)
\stopTextSpan | d4 f | \break
\varB
 \set Score.measureLength = #(ly:make-moment 3 4)
 s4\startTextSpan g2\stopTextSpan |
 \set Score.measureLength = #(ly:make-moment 2 4)
f8( e^"(v)") d4 d2 d4 d
\bar "|."  s8 \fixB gis,4 \fixC s8 \bar "|."
\endm
\varA
 g'8\startTextSpan\prall( f) f16([ e c8)] \stopTextSpan \bar "||"
\varB
 g'4.\startTextSpan( a8)  \stopTextSpan  \bar "||"
\varB
 g4.\startTextSpan( f8)  \stopTextSpan \bar "|."
}
\addlyrics { Еж- те ми, пей- те, пи- лян- ца дур- де сти мал- ки при
ма- ма }
\layout {
  indent = #0
  line-width = 190\mm
  ragged-right=##f
}
\midi {
    \context {
        \Score tempoWholesPerMinute = #(ly:make-moment 144 4)
            }
        }
    }
\header {
  opus = "ТД-136,1,16"
  tagline = ##f
}
```

The text of the song and description from the manuscript is prepared in LaTeX format:

```
%На хоро
\begin{multicols}{2}
Ежте ми пейте пилянца,
дурде сти малки при мама.
Аз га бях малка при мама
бяла бих като кадъна,
червена като ябълка,
където ходя все пея,
в градинка вляза и пея.
А сега пиленца, а сега,
даде ма майка ми, даде ма
през девет села в десето
при девет млади итърви
и девет зълви се моми,
едно ми младо деверче
```

```
         и една ми стара свекърва.
         Когато почна ляп да меса,
         аз меса тя го премиса;
         когато почна да пера,
         аз пера тя го препира;
         когато почнем да межем,
         аз мажа тя го премазва;
         когати тръгнем на нива
         всяка си земе детето,
         на мене дават стомните.
         \end{multicols}
         \singer{Пял Стоян Гьоргев Гайдаров}
         \village{с. Победа}
         \area{Ямболско}
         \bornYear(1895)
```

and is stored in a plane text format file. The system LilyPond produces NKirov_folk.mid file with piano sound of the song notes. The original performance, recorded "on the field" and digitized, is stored in file NKirov_folk.mp3. The sheet music and lyrics of this song can be seen here:

Ежте ми пейте пилянца
(На хоро)
ТД-136,1,16

Еж- те ми, пей- те, пи- лян- ца дур- де

сти мал- ки при ма- ма

Ежте ми пейте пилянца,
дурде сти малки при мама.
Аз га бях малка при мама
бяла бих като кадъна,
червена като ябълка,
където ходя все пея,
в градинка вляза и пея.
А сега пиленца, а сега,
даде ма майка ми, даде ма
през девет села в десето
при девет млади итърви
и девет зълви се моми,

едно ми младо деверче
и една ми стара свекърва.
Когато почна ляп да меса,
аз меса тя го премиса;
когато почна да пера,
аз пера тя го препира;
когато почнем да межем,
аз мажа тя го премазва;
когати тръгнем на нива
всяка си земе детето,
на мене дават стомните.

Пял Стоян Гьоргев Гайдаров от с. Победа, Ямболско, род. 1895 г.
Записал и нотирал Тодор Джиджев.

Conclusions and Perspectives

The activities of explanation and advertising of the database by the team members will give the Bulgarian musicologists the opportunity to take an advantage from the created technology even after the end of the project. In this way they will be able not only to more easily notate the recorded folk songs, but also to make their field recordings more popular by placing them on the Internet. Having in mind the fact that the designed database is both

software and a media, it will turn into a platform for the preservation and publication of the large amounts of music archives.

The enrichment of the collections is beneficial for all the concerned parties and presents a long-term perspective for the transformation of the designed database into a specific musical and cultural atlas (similar to the "dialectical atlas") that will show the "geography", e.g. the distribution of a given song, given musical features and the clear outline of the music folk dialects. These opportunities will undoubtedly inspire the creation of scientific researches based on the resources in the designed database, which, on the other hand, will advertise it and will provoke the adding of new data.

There is a reasonable possibility for international collaboration for digitization of authentic folk songs from

- neighbor countries (Romania, Serbia, Macedonia, Greece, Turkey);
- Balkan countries (SEEDI countries);
- Cyrillic alphabet countries (Serbia, Macedonia, Ukraine, Belarus, Russia).

An extended collection of digitized authentic folk songs from several countries can generate many investigations about common roots of popular customs and culture for the people in these countries.

Current activities in the project frame can be found on
http://www.math.bas.bg/~nkirov/2010/folk/FOLK_en.html

References

[1] Nikolay Kirov, 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), 78-83.
[2] Aleks Nushev, Digitalization, Restoration, and Audio Mastering Practices at the Instituteof Art Studies in Sofia, Bulgaria, Review of the National Center for Digitization,
[3] Lozanka Peycheva, Grigor Grigorov, How to Digitalize Folklore Song Archives? Review of the National Center for Digitization
[4] Lozanka Peycheva, Nikolay Kirov, Maria Nisheva-Pavlova, Information technologies for presentation of Bulgarian folk songs with music, notes and text in a digital library, Proc. of Fourth International Conference "Information Systems & GRID Technologies", 28-29 May 2010, Sofia, Bulgaria.
[5] LilyPond ... music notation for everyone, http://lilypond.org/
[6] Linked European Archives for Ethnomusicological Research, http://www.ethnoarc.org/
[7] The Music Encoding Initiative … …modeling music notation with XML, http://music-encoding.org/
[8] The Mutopia Project: Free sheet music for everyone, http://www.mutopiaproject.org/

nkirov@nbu.bg