Bulgarian Folk Songs in a Digital Library

Lozanka Peycheva[1], Nikolay Kirov[2,3]

[1] Institute for Ethnology and Folklore Studies with Ethnographic Museum, Bulgarian Academy of Sciences, Moskovska Str. 6A, 1000 Sofia, Bulgaria,
lozanka.peycheva@gmail.com
[2] New Bulgarian University, Montevideo Str. 21, 1618 Sofia, Bulgaria,
nkirov@nbu.bg
[3] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 8, 1113 Sofia, Bulgaria

Abstract. The paper presents the main results 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. An initial digitization and preservation of the Bulgarian cultural heritage starts by means of digitization and insertion into the library of over 1000 songs that were recorded and written down during the 60s and 70s of XX century. Also we present a full text search engine in a collection of lyrics (text of songs) and coded notes (symbolic melody). Some perspectives for future projects are also discussed.

Keywords: digitization, Bulgarian folk songs, digital library, search engine

1. Introduction

A digitization project for Bulgarian folk songs "Information technologies for presentation of Bulgarian folk songs with music, notes and text in a digital library" has been initiated in 2009 by several folklorists and computer science specialists from five institutions:

- Institute of Mathematics and Informatics, Bulgarian Academy of Sciences (BAS);
- Institute of Art Studies, BAS;
- Institute for Ethnology and Folklore Studies with Ethnographic Museum, BAS;
- Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ochridski";
- Department Informatics, New Bulgarian University.

The project has been partially supported by Bulgarian NSF (see [1]). The research aims at the development of a technology and corresponding supporting software tools for the creation and usage of heterogeneous institutional digital libraries with folk songs. The tools will satisfy the needs of the researchers in the fields of ethnology, ethnomusicology and folkloristics.

The team activities referring the project are described on the project website [7].

2.   The Collection

2.1   Description

The sources for our collection are the archives of National Center for Intangible Cultural Heritage at the Institute for Ethnology and Folklore Studies with Ethnographic Museum, and of the Institute of Art Studies. They are a part of the scientific heritage of Professor Todor Dzidzev - over 1000 songs that were recorded and written down during the 60s and 70s of XX century. The archive consists of:

- About 2000 typewritten pages with handwritten (musical) notes (scores), and text of songs (Fig.1);

Fig. 1. A typewritten page and handwritten notes of a song

- 27 old magnetic types with original performances of songs;
- 13 note-books and pads with handwritten texts of the songs and additional notes, written in the time of research expeditions, when the songs had been recorded (Fig.2).
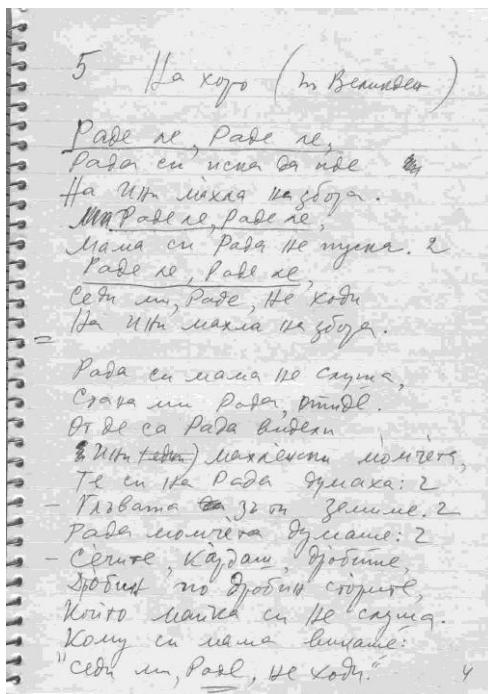
Fig. 2. A handwritten page of a note-book

All materials are digitized and are included in a collection. The digital library of Bulgarian folk songs is already created adding tools for accessing the data. (A digital library is a library in which collections are stored in digital formats and accessible by computers).

```
\score·{¶
\relative·c'·{¶
\tempo·4·=·144¶
\time·2/4¶
d4·f·|·\acciaccatura·f8·g2·|·d4·f·|·g2·|¶
\varA¶
·\acciaccatura·a8\startTextSpan·g4\glissando·f16(·ees·c8)·\stopTextSpan·|·d4·f·|¶
\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·}¶
```

Fig. 3. A fragment from a source code in LilyPond notation

The main object in the library is a song. The song consists of four items: notes, lyrics (the text of the song), music, and manuscripts. The notes are presented in a form of LilyPond coding [6] (plain text files, Fig.3) and print view (pdf and eps files, Fig.4). Now we have about 1100 songs with notes and text.

Fig. 4. Print view of a song notes

The texts of the songs can be extracted from the library in form of LaTeX (text files with metadata, Fig.5) and in printed form (pdf format, Fig.6).

```
%Ежте·ми,·пейте,·пилянца¶
%На·хоро¶
%Джиджев:хороводни¶
\begin{multicols}{2}¶
Ежте·ми,·пейте,·пилянца,¶
дурде·сти·малки·при·мама.¶
Аз,·га·бях·малка·при·мама,¶
бяла·бих·като·кадъна,·(2)¶
червена·като·ябълка:¶
където·ходя,·все·пея,¶
в·градинка·вляза·и·пея!¶
.....................¶
когато·тръгнем·на·нива,¶
всяка·си·земе·детето,¶
на·мене·дават·стомните!¶
\end{multicols}¶
\singer{Пял:··Стоян·Гьоргев·Гайдаров}¶
\village{с.·Победа}¶
\area{Ямболско}¶
\year(1895)¶
```

Fig. 5. Text of a song in LaTeX format

The music records consist of authentic performances of folk songs, digitized from old magnetic types. They are presented in MP3 format. Now 1403 songs are in the library, some of them are not written down in notes. Also MIDI files, produced by LilyPond system (computer interpretation of the notes) are included into the library.

The handwritten note-books from the research expeditions are presented in a form of page images (jpg format, Fig.2).

2.2   Metadata

Based on the expertise of the scientific team of the project the following metadata components of recorded and notated folk songs were identified and defined:

- title
- context implementation
- genre of the song on a functional basis
- name of the singer
- birthplace of the singer
- municipality, which belongs to the birth place of the singer
- name of the recorder
- place of the recording
- time of the recording

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.

Fig. 6. Print view of a song text

3.   Search Engine

The presented here search engine can be used as web application for keyword-based searching in our library. The folklore songs are provided as an index of digital content – lyrics, notes and images. This engine could be used by professionals in the field of folklore research to look for common motives, characters and similarities between different songs. These could be songs from different parts of Bulgaria, variants of the same song or simply common keywords.

The technical details about the search engine implementation can be found in [3].

3.1   Search Query

The search engine provides a Google-like web interface that would be used for searching in the library (Fig. 7). It uses a search phrase (query) which could contain data as well as metadata. Here is a short list of possible searches:

- Рада – A simple one word search for "Рада" - a popular given name in Bulgarian. This search engine returns a result with all songs which contain that word. Note that "Рада" could be the name of the singer or the name of the folklore hero for whom the song was made.

Fig. 7. Interface for searching

- code:ba_002_2_04 – A code search. Every folklore song in the library has a unique code. The "code" is a separate field in the index table, so we specify a field using the shown above syntax.
- content:"ожадня стоян за вода" – A whole phrase search. The engine returns songs which contains the given words in the given order. "content" is a keyword for the lyrics field.
- ст*ян AND area\{ямболско\} – A wildcard and boolean search. In the different folklore songs the name "Стоян" is sometimes spelled "Стуян", so we want to match both of them. We also want to search only in the "ямболско" municipality, so we specify a metadata field, which describes that area.
- notes:fermata – Search in LilyPond coding. This search returns all of the songs, which contains a "fermata" (an element of musical notation) in their LilyPond coding.

3.2   Search Result Table

The search result table contains the songs that match a given search query (Fig. 8). Each song is represented by a row in that table, which contains all the data and metadata in the library about that song. Every song is identified by its unique code. By default the search result table is sorted by the relevance index given by the search engine. So the best matches are shown first. In addition to that the user could sort the table by any field. The context of the given match is also displayed in the search result table, so the user could see the specific stanzas for example, that contain the given word. The user could also hear the authentic performance online, by clicking on the given MP3 link, for the specific song in the search results. A compiled, MIDI version of the LilyPond source file could also be heard. That could be used as a reference between written notes and the authentic performance.

3.3   Google Maps Visualization

In every step of the search process, a link is provided that could visualize the resulting songs in Google Maps (Fig.9). The system extracts the relevant meta-data from the index and forms a series of Google Maps queries, that should return the exact location (or locations) associated with each given song. These queries are formed as strings containing the name of town or village, where the song was performed or associated with, and the municipality in which that town or village is located. Since Google uses keyword based search, that pair should be enough to distinguish between names of villages located in different municipalities (which is a common occurrence in Bulgaria). Google Maps queries return a GPS coordinates, on which each song location is visualized. By using this technique a user could figure out how a given song motive is spread across rural areas of Bulgaria. It can be used also to track the locations associated with different singers.

Резултати от търсене за: **Ежте**

Код на песента	Съвпадение	Контекст	Текст	Lilypond нотен запис	PDF нотен запис	EPS нотен запис	Изпълнение	MIDI музика	Изображения
td_136_1_16	0.56	...**Ежте** ми, пейте, пилянца %На хоро %Джиджев:хороводни \begin{multicols}{2} **Ежте** ми, пейте, пилянца, дурде сти малки при мама, Аз, га бях малка при мама, бяла бях като...	.txt	.ly	.pdf	.eps	.mp3	.midi	
td_139_2_10	0.50	%Йелка момаре дойдаха %Седенкарска (на летни попрелки) %Джиджев:седенкарски \begin......майка \'{и} посрещна и на момаре думала: -- **Ежте** и пийте, момаре, и си със здраве...	.txt	.ly	.pdf	.eps	.mp3	.midi	.jpg .jpg
td_170_2_07	0.35	%Посъбра Гълъб, Гълъбо %На хоро %Джиджев:хороводни %%този и следващите записи са......думаше: %%"Гълъб" вм. "и на" - според записа -- **Ежте** и пийте, дружина, и са в акъла...	.txt	.ly	.pdf	.eps	.mp3	.midi	
td_178_2_06	0.35	%Посъбра Гълъб, Гълъбо %На хоро (по Великден) %Джиджев:хороводни %%този и......да е от "револвери" и на дружина думаше: -- **Ежте** и пийте, дружина, и са в акъла...	.txt	.ly	.pdf	.eps	.mp3	.midi	

4 намерени песни

[Търсене в резултатите]

[Карта]

Fig. 8. Search table – the result of searching the word Ежте

3.4   Perspectives

In its current state the folklore songs search engine is experimental and can be used only by experts and folklore professionals. The web interface of the system could be improved and redesigned according to web standards. Such improvement would make

the system easier to use for professionals in the field as well as usable to the interested communities on the Internet. An essential improvement will be semantics-oriented search based on an ontology approach [5].
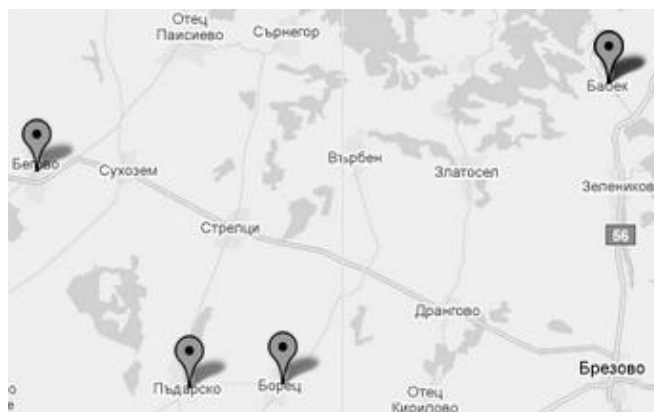

Fig. 9. Google Maps shows the result of searching the word Рада

4. Future Developments

Our project team has an ambition to extend the contents of the digital library including into it folklore songs, which are recorded by many prominent Bulgarian researchers of musical folklore. Vasil Stoin is one of them and we propose to start with the digitization of books, which are published in a result of Stoin's project in the 20s of XX century. The arguments for our choice are:

- Vasil Stoin collections (books and unpublished records) are a valuable treasury of folk music heritage of Bulgaria.
- Much of the notated songs published in these books have long been forgotten and not persist in the musical practices of Bulgarians. If they are published again and equipped with sound examples, these songs can be sung again by the modern and future fans of Bulgarian folk music.
- Only single specimens from the collections are saved today, wasted by the time and usage. Many of these songs are threatened with extinction.
- Issued at different times, the collections contain "scattered" over time folk music content, which if collected, will be discoverable and accessible to broader audiences.
- To our knowledge, Bulgaria is the only Balkan country that has implemented such a comprehensive project for collecting, deciphering and publishing folk music heritage.
- Stoin's folk music collections cover almost all ethnographic regions in Bulgaria. They provide a unique music style of the Bulgarian Dialect Atlas.

- The launch of the Stoin's project for collection and publication of Bulgarian folk songs had taken place due to the fact that this task was defined as a state priority by the Ministry of Education, funded by the Prime Minister Professor Alexander Tsankov in 1925 and by all subsequent ministers until completion of the project.

Our achievements are also a good base for international collaboration in the field of authentic folk songs digitization. One perspective is to take in mind the common features of Balkan folklore (Romania, Serbia, Macedonia, Greece, Turkey), and other is to attract Cyrillic alphabet countries (Serbia, Macedonia, Bosnia and Herzegovina, 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 of the people in these countries.

References

1. Kirov, N.: Digitization of Bulgarian folk songs with music, notes and text. Review of the National Center for Digitization 18, 35-41 (2011)
2. Peycheva, L., Kirov, N., Nisheva-Pavlova, M.: Information Technologies for Presentation of Bulgarian Folk Songs with Music, Notes and Text in a Digital Library. Proc. of Fourth Int. Conf. "Information Systems & Grid Technologies", Sofia, Bulgaria, May 28–29, 2010, 218-224 (2010)
3. Kirov, K., Kirov, N.: Digital library and search engine of Bulgarian folklore songs. Proc. 7 Ann. Int. Conf. on Comp. Sci. and Educ. in Comp. Sci., 245-254 (2011)
4. Peycheva, L., Grigorov, G.: How to Digitalize Folklore Song Archives? Review of the National Center for Digitization 18, 42-58 (2011)
5. Nisheva-Pavlova, M., Pavlov, P.: Search Engine in a Class of Academic Digital Libraries. Proc. 14th Intern. Conf. on Electronic Publishing 16-18 June 2010, Helsinki, Finland, 45-56 (2010)
6. LilyPond ... music notation for everyone, http://lilypond.org/
7. Information technologies for presentation of Bulgarian folk songs with music, notes and text in a digital library,
   http://www.math.bas.bg/or/nkirov/2010/folk/FOLK_en.html