

The 7 Annual International Conference on Computer Science and  
Education in Computer Science, July 06-10 2011, Sofia, Bulgaria

## **DIGITAL LIBRARY AND SEARCH ENGINE OF BULGARIAN FOLKLORE SONGS**

Kiril Kirov<sup>1</sup>, Nikolay Kirov<sup>2</sup>

<sup>1</sup>Magrathea Co., [kiril@kirov.be](mailto:kiril@kirov.be)

<sup>2</sup>Computer Science Department, New Bulgarian University [nkirov@nbu.bg](mailto:nkirov@nbu.bg) *and*  
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

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 started in 2009, joining the efforts of various experts from three institutes of the Bulgarian Academy of Sciences –

- Institute of Mathematics and Informatics,
- Institute of Art Studies and
- Institute of Folklore

and two universities –

- Sofia University (Faculty of Mathematics and Informatics) and
- New Bulgarian University (Department Informatics).

in the field of applied informatics for digitization of musical notation and lyrics.

The project is partially supported by the National Science Fund of the Bulgarian Ministry of Education and Science.

## Project objectives

- Studying of the architectural principles of institutional and, in particular, the academic digital libraries, instruments and tools and creation of functional model of digital library with Bulgarian national music.
- Development and application of information technologies in the creation of digital libraries, instruments and tools for search and extraction of knowledge in a concrete scientific field.
- Creation of technological environment for digitization of notations specially adapted for Bulgarian folk songs. Creation and development of heterogeneous database with notations, lyrics and music.
- Digitization and preservation of Bulgarian cultural heritage, (under this project) over 1000 songs are already digitized and introduced into the database.
- Theoretical classification of the songs according to various criteria: technical, implicit-musical, implicit-textual, statistical, genre, cultural, etc.

Expected results and effects from project implementation, knowledge transfer potential, practical applicability of results

- First of all, a considerable collection will be created of over 4000 authentic folk songs that are priceless in value, because they were written down and recorded during the 60s and 70s of the XX century and it is not certain they can still be found.
- Second, the scientists from the Institute of Folklore and Institute of Art Studies of BAS that are engaged in the preservation and research of the folk songs will acquire (not only on theory, but also on practice by means of their participation in the project) the developed technology for digitization and publishing on the Internet of the newly recorded samples and of the already existing musical archives.
- Third, 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 Prof. 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 indexes 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.

- The research that is carried out under this project aims 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.
- Now our database contains notes (coded with LilyPond software), lyrics (in  $\text{\LaTeX}$ ) and music (MP3, digitized from old magnetic tapes)
- Now we will present a full text search engine in a collection of lyrics (text of songs) and coded notes (symbolic melody) as a part of digital library of Bulgarian folklore songs

## Example – original manuscript

TD-136, 1, 16

на хоро

Ехте ми пейте пиланца,  
дурде сти малки при мама.  
Аз га бях малка при мама

♩ ≈ 144

TD-136, 1, 16

Ех-те ми, пей-те, пи-ля-ца

дур-де сти мал-ки при ма-ма

когато почнем да мажем,  
аз маха тя го пемазва;  
когато тръгнем на нива  
всяка си земе детето,  
на мене дават стомните.

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

TD-136, 1, 16

на хоро

Ехте ми пейте пиланца,  
дурде сти малки при мама.  
Аз га бях малка при мама  
бяла бях като жадна,  
червена като ябълка,  
където ходи все пая,  
в градинка аляза и пая.  
А сега пиланца, а сега,  
даде ма майка ми, даде ма  
през девет седа в десето  
при девет млади итърви  
и девет зълви се моми,  
едно ми младо деверче  
и една ми стара свекърва.  
Когато почна лят да меса,  
аз меса тя го премиса;  
когато почна да пера,  
аз пера тя го препира;  
когато почнем да мажем,  
аз маха тя го пемазва;  
когато тръгнем на нива  
всяка си земе детето,  
на мене дават стомните.

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

Example – LilyPond coding:

```
\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
}

```



Example – lyrics:

```
%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 кадъна, /2  
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  
\krag  
и девет зълви [в]се моми,
```

LilyPond MIDI file:  
tdd\_136\_1\_16.mid

```
едно ми младо деверче  
и една ми стара свекърва.  
Когато почна [х]ляп да мяса,  
аз мяса, тя го премиса;  
когато почна да пера,  
аз пера, тя го препира;  
когато почнем да мажем,  
аз мажа, тя го премазва;  
когато тръгнем на нива,  
всяка си земе детето,  
на мене дават стомните.  
\end{multicols}
```

```
\singer{Пял: Стоян Гьоргеv Гайдаров}  
\village{с. Победа}  
\area{Ямболско}  
\bornYear{1895}
```

Original performance:  
TD\_136\_1\_16.mp3

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 мяса,  
аз мяса 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 Гьоргеv Гайдаров от с. Победа, Ямболско, род. 1895 г.  
Записал и нотирал Тодор Джиджев.

The basic data files can be defined in four types:

- $\text{\LaTeX}$  lyrics files.

The lyrics of the songs in this library are written in  $\text{\LaTeX}$  typesetting system. In addition to the song text, each  $\text{\LaTeX}$  file provides meta-data information encoded in the text. This meta-data is in the form of different  $\text{\LaTeX}$  commands, and could be used both in compilation of the source and generating the index.

- LilyPond notes files.

The notes of the songs are written in LilyPond music typesetting system.

- MP3 digitized authentic performance files.

These performances are digitized from magnetic tape libraries of the Archives of Bulgarian Academy of Sciences. They had been recorded in different rural parts of Bulgaria during the 60s and 70s.

- JPEG digitized handwritten texts.

The handwritten note-books included in this library were made by the experts, who worked with the authentic performers and then analyzed the collected data. They are valuable source of information about the circumstances and different traditions associated with the performance of each particular song.

5 Ха хого (in Beaudou)

Раде ле, Раде ле,  
Рада си исна ба нде  
Ха ити маха ка збожа.  
Ма Раде ле, Раде ле,  
Мама си Рада не туча. 2  
Раде ле, Раде ле,  
Седи ми, Раде, не ходи  
Ха ити маха ка збожа.

Рада си мама не суча,  
Стара ми Рада, отиде.  
От де са Рада видели  
и ити (теди) махаевски момчета,  
Те си не Рада думаха: 2  
- Гловата ~~за~~ збожа злиме. 2  
Рада момчета думаха: 2  
- Седите, каждам, дробите,  
Дробит по дробит сорице,  
Който майка си не суча.  
Кому си мама викаше:  
"седи ми, Раде, не ходи."

- The system uses Ferret – a high-performance, full-featured text search engine library written for Ruby (C-implementation)
- The input files for indexing are lyrics and notes.
- The compilation process could be started by bin/lilypond and bin/latex

- Рада – *A simple one word search* for Рада – a popular given name in Bulgarian. This search should return a result with all songs that contain the word.

[Разширено търсене](#)

☒ в текст по ключови думи
 ☐ в текст, семантично
 ☐ в нотен запис

Търсене

Търсене в 1071 български народни песни

- `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.
- `content:"ожадня стоян за вода"` – *A whole phrase search*. Should return songs with containing 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 meta-data field, which describes that area.
- `notes:a2\\fermata` – *A notation search*. This search should return all of the songs, which contains a half note “la” with “fermata” (an element of musical notation) in their LilyPond coding.

This system uses an integrated Ruby web server stack to service and present data in web form to its users. This stack includes the following Ruby Gems:

- Thin – A web server
- Rack – A web server interface
- Sinatra – Web development framework
- HAML – A web page template system

Search Result Table:

- The search result table contains the songs that match a given search query
- By default the search result table is sorted by the relevance index given by the search engine
- The context of the given match is also displayed in the search result table
- In addition to that the user could sort the table by any field

Резултати от търсене за: **Рада**

Код на песента	Съвпадение	Контекст	Текст ▼	Lilypond	PDF	EPS	Изпълнение	MIDI	Изображения
				нотен запис	нотен запис	нотен запис		музика	
<a href="#">td_112_1_05</a>	0.51	...Димо на <b>Рада</b> думаше %Седенкарска %Джиджев:седенкарски \begin{multicols}{2} Димо на <b>Рада</b> думаше: -- Ради мо, любе, Ради мо, знаеш ли, Раде, помниш ли, мама й меджия...	<a href="#">.txt</a>	<a href="#">.ly</a>	<a href="#">.pdf</a>	<a href="#">.eps</a>	<a href="#">.mp3</a>	<a href="#">.midi</a>	<a href="#">.jpg</a>

<a href="#">td_181_1_30</a>	0.54	...Първан и <b>Рада</b> %На хоро %(Пролетно) \begin{multicols}{2} Любят сай Първан и <b>Рада</b> та.....са за женени. Той прати Първан за <b>Рада</b> . <b>Рада</b> майка си думаше: -- Мамо ле мамо...	<a href="#">.txt</a>	<a href="#">.ly</a>	<a href="#">.pdf</a>	<a href="#">.eps</a>	<a href="#">.mp3</a>	<a href="#">.midi</a>	
-----------------------------	------	--	----------------------	---------------------	----------------------	----------------------	----------------------	-----------------------	--

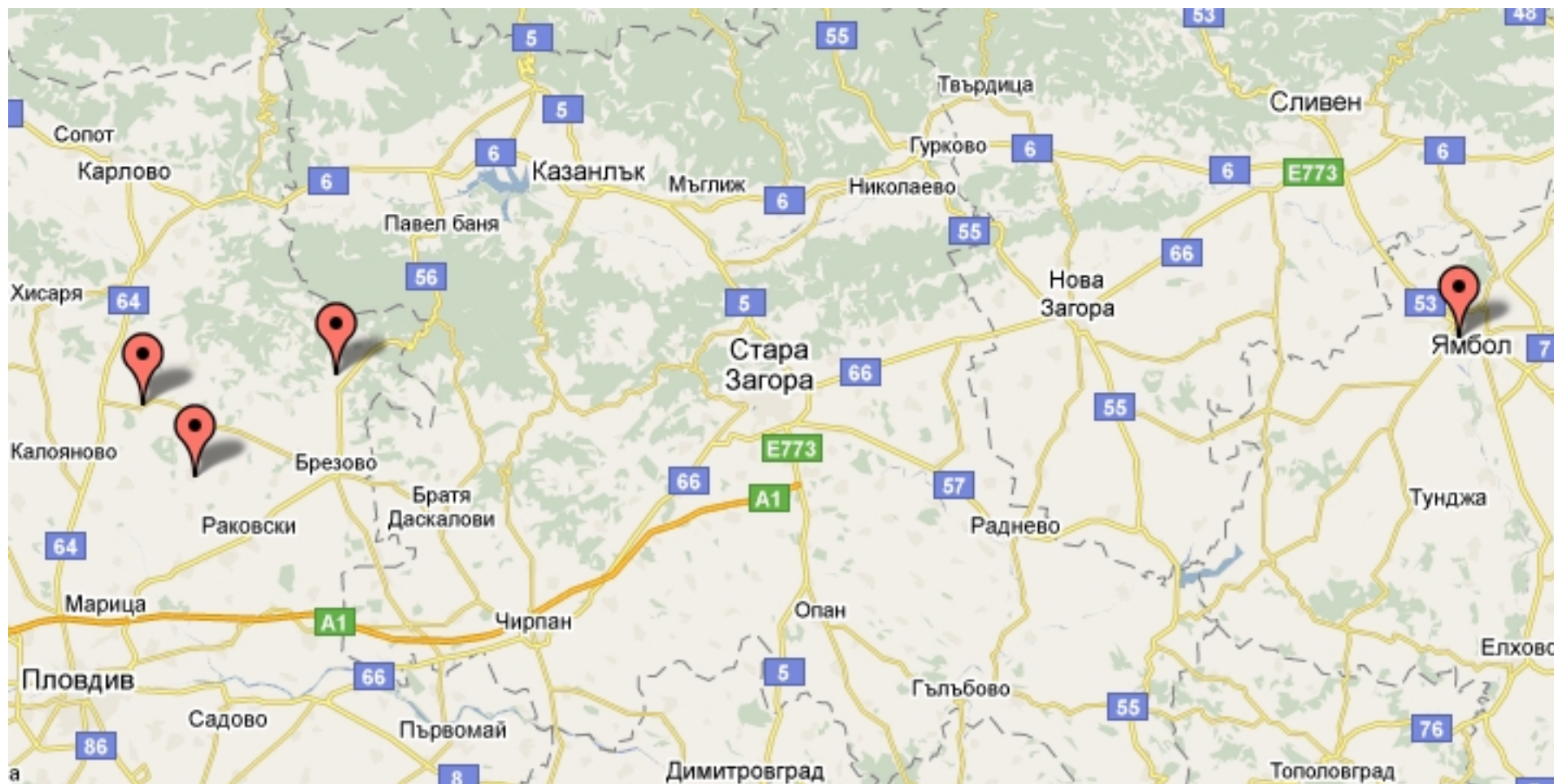
61 намерени песни

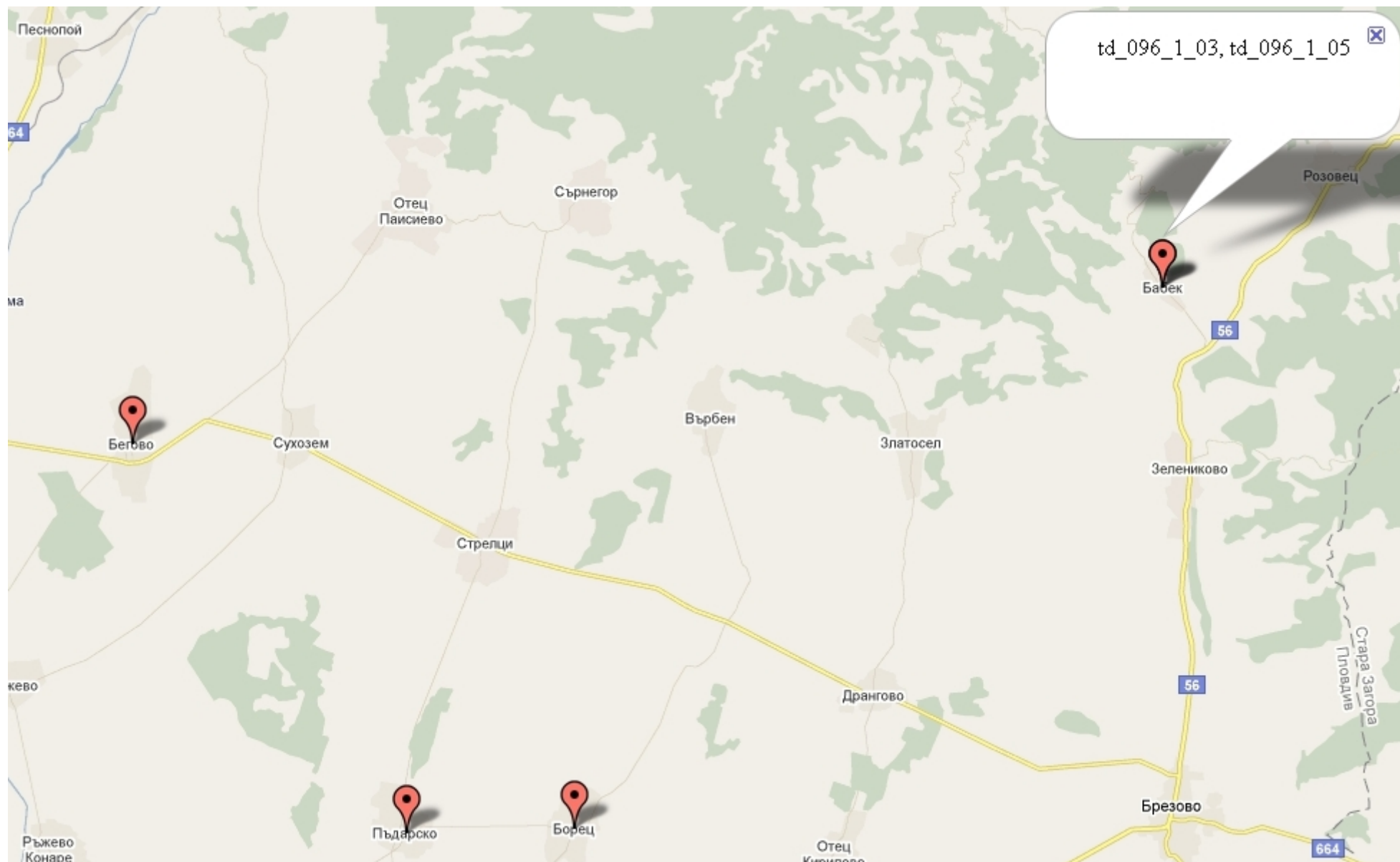
Търсене в резултатите

Карта



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.





The result of searching in the LilyPond codes for `a2\fermata`.

Резултати от търсене за: `a2\fermata`

Код на песента	Съвпадение	Контекст	Lilypond					MIDI музика	Изображения
			Текст	нотен запис	нотен запис	нотен запис	Изпълнение		
<a href="#">td_139_1_10</a>	0.71	... <b>a2\fermata</b> s2 \bar "" s2 \break b8([ a]) a4    b4( c16[ b a8])   b4 a   b4( c16[ b a8])   a8([ b]) g4   a2^\rtoe   a4 a   <b>a2\fermata</b> s2..... <b>a2\fermata</b> s2 \bar "" s2 \break g4 a4    b4( c16[ b a8])   b8([ a]) a4   b2\mordent   b8([ a]) a8([ g])   a2^\rtoe   a4 a   <b>a2\fermata</b> ...	<a href="#">.txt</a>	<a href="#">.ly</a>	<a href="#">.pdf</a>	<a href="#">.eps</a>	<a href="#">.mp3</a>	<a href="#">.midi</a>	<a href="#">.jpg</a>
<a href="#">td_095_1_10</a>	0.57	...c" { \tempo 4 = 152 \time 2/4 g4 a8\noBeam b   c4 d\mordent   c8\noBeam c c8([ b])   \afterGrace <b>a2\fermata</b> ( { c16[ d]})   b8 c4.   c8\prall([ b]) a4   <b>a2\fermata</b> \bar " ." } \addlyrics { Раз- бу- ля- ла "съ~й," ми- ла ма- мо, ху- ба- ва Ган- ка...	<a href="#">.txt</a>	<a href="#">.ly</a>	<a href="#">.pdf</a>	<a href="#">.eps</a>	<a href="#">.mp3</a>	<a href="#">.midi</a>	<a href="#">.jpg</a> <a href="#">.jpg</a>

<u>td_134_2_30</u>	0.19	%{ laz_134_2_30 %} \include "td-preamble.ly" \score { \relative c" { \tempo 4 = 168 \time 2/4 \varA a8\startTextSpan( b) c.....varEF \time 4/16 c16\startTextSpan([ b a b\stopTextSpan)]   \time 2/4 <b>a2~   a\fermata</b>   a4^\toe a^\rtoe   a2 \bar " ." s4 \fixB cis, \fixC...	<u>.txt</u>	<u>.ly</u>	<u>.pdf</u>	<u>.eps</u>	<u>.mp3</u>	<u>.midi</u>	<u>.jpg</u>
<u>td_138_2_04</u>	0.19	%{ TD_138_2_04 %} \include "td-preamble.ly" \score { \relative c" { \tempo 4 = 112 \time 2/4 a4 d   \varA d8.....c8\stopTextSpan) \bar ""   \varD a4\startTextSpan a   <b>a2\stopTextSpan\fermata</b> \bar " ." } \addlyrics { Как се Ма- рий- ка...	<u>.txt</u>	<u>.ly</u>	<u>.pdf</u>	<u>.eps</u>	<u>.mp3</u>	<u>.midi</u>	<u>.jpg</u>

82 намерени песни

[Demonstration] [td\_139\_2\_03]

**Thank you for your attention.**

<http://nikolay.kirov.be/2010/folk/F0LK.html>

*Acknowledgement.* This work is supported by Grant of the Bulgarian National Science Foundation under number DTK-02-54/2009.

08.07.2011