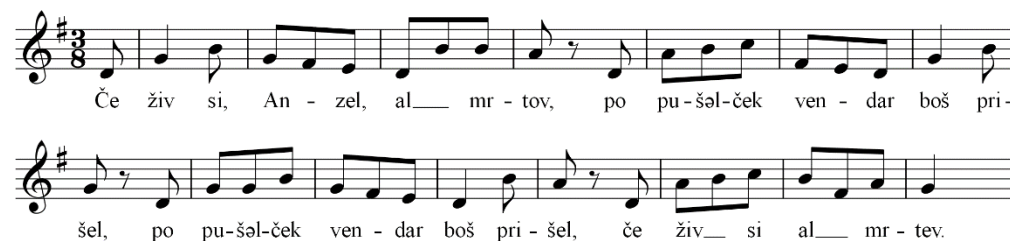


Računalniška analiza glasbenih gradiv

Matija Marolt, UL FRI

matija.marolt@fri.uni-lj.si



Laboratorij za računalniško grafiko in multimedije

lgm.fri.uni-lj.si



Raziskovalna področja:

- pridobivanje informacij iz glasbe
- (3D) grafika in vizualizacija
- poigritev učenja

Poučevanje:

- Računalniška grafika in tehnologija iger
- Napredna računalniška grafika
- Multimedijske tehnologije
- Tehnologija programske opreme
- Računalniško podprto oblikovanje (ALUO)

Pridobivanje informacij iz glasbe

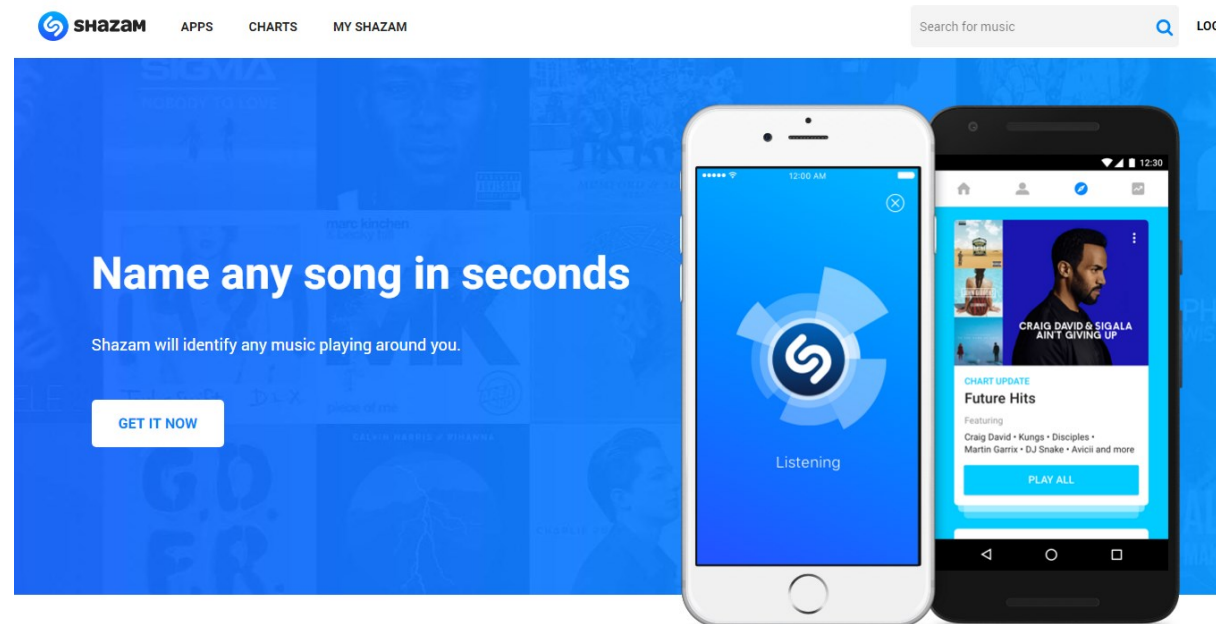
Pridobivanje informacij iz glasbe - *music information retrieval*

- Področja
 - **osnovni elementi:** melodija, harmonija, tempo, ritem, začetki not ...
 - **značilnosti:** inštrumenti, žanr, čustva, izvor ...
 - **iskanje:** identifikacija, podobnost, iskanje izvedb ...
 - **izvedba:** sledenje izvajalcu, izraznost, stil izvedbe, identifikacija skladatelja ...
 - **zbirke:** optično prepoznavanje, analiza korpusov, organizacija zbirk, vizualizacija, uporabniški vmesniki ...
- Primeri uporabe
 - **priporočilni** sistemi (značilnosti, podobnost)
 - **učenje** inštrumentov (elementi, sledenje)
 - **arhivi** (zbirke, značilnosti, iskanje)
 - **ustvarjanje** glasbe (iskanje, izvedba)



Primer: Shazam

- [Shazam](#): Ena prvih komercialno uspešnih storitev
- Identifikacija skladbe
 - *audio fingerprinting* – na podlagi primera najde originalni posnetek
 - enak posnetek mora biti v Shazam bazi
 - tolerira različna popačenja in šume



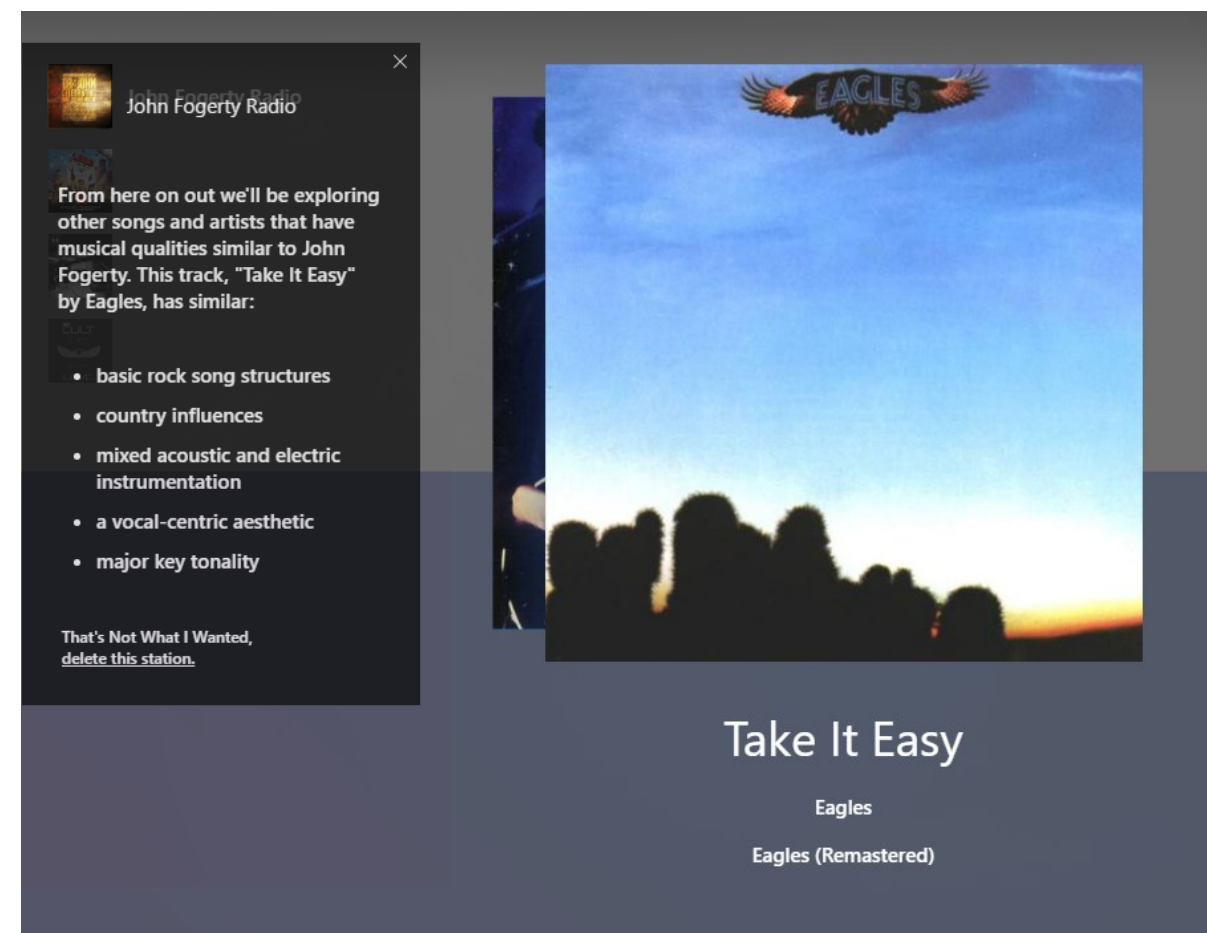
Primer: Pandora

- Pandora: vsaka pesem opisana s 400+ značilnostmi
 - *music genome*
 - sprva ročno označevanje, danes veliko opravijo algoritmi
- Primer: Atributi besedil

Abstract, Ambiguous, Angry, Clean, Erotic, Explicit, Great, Heartbreaking, Humorous, Famous Rap Artist, Heavy Erotic Content, Narrative, Offensive, Political, Positive, Upbeat, by Rap Icons, by Respected Rap Artists, Religious, Romantic, Sad, Storytelling, Street Talkin, Surreal, Upbeat, Violent
- Informirana priporočila

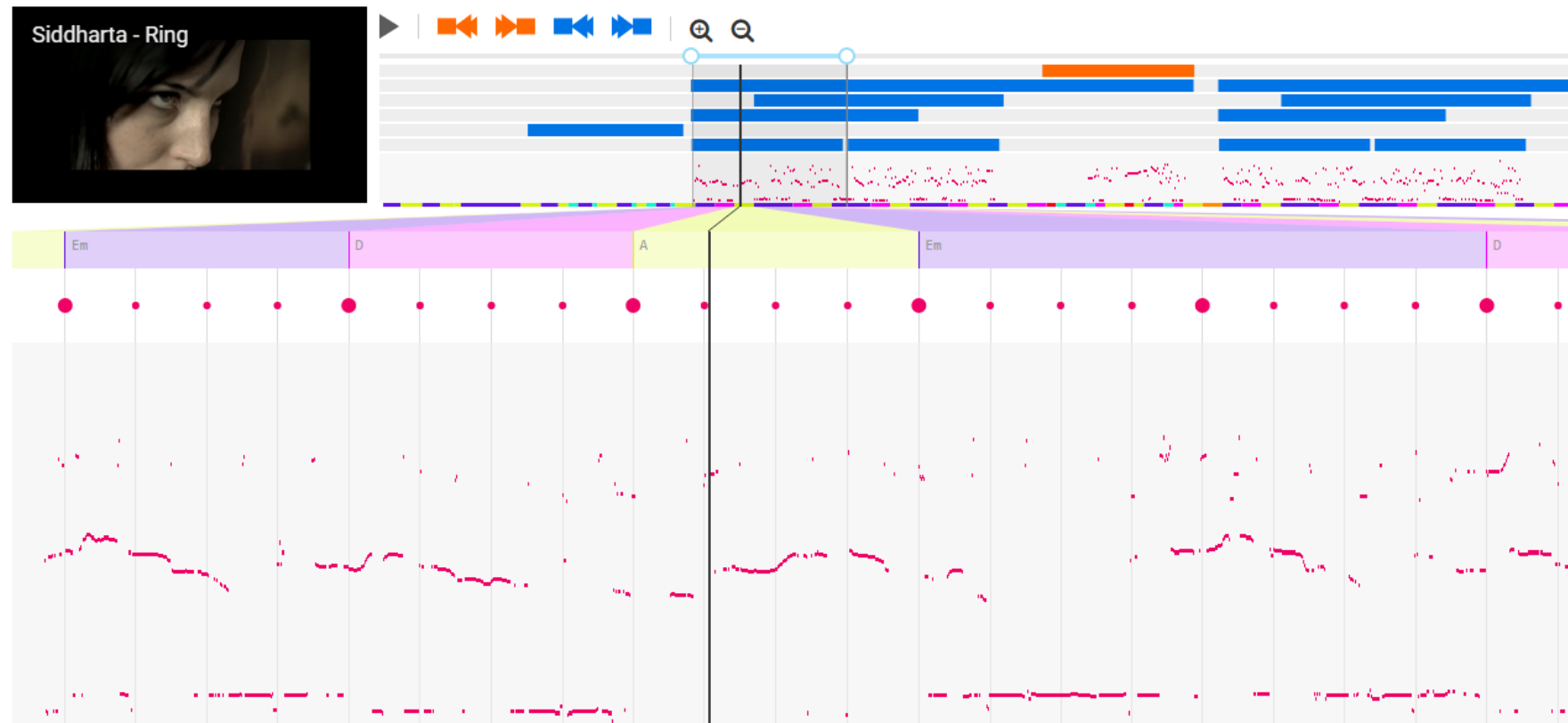
Pete Seeger Radio

To start things off, here's a song that's musically similar to Pete Seeger called "Aragon Mill" by Peggy Seeger, that features *folk roots, country influences, a subtle use of vocal harmony, acoustic sonority and major key tonality.*



Primer: Songle

- Vizualizacija avtomatsko pridobljenih opisov osnovnih elementov
 - struktura
 - harmonija
 - ritem in tempo
 - melodija in bas



Primer: ustvarjanje in učenje glasbe

- Učenje
 - [Yousician](#)
- Ustvarjanje
 - [Continuator](#)
 - [DeepBach](#)
 - [Flow Machines](#)
 - [folk-rnn](#)

"Ode to Joy" harmonized in the style learned from:



Analiza ljudske glasbe

Glasbenonarodopisni inštitut, ZRC SAZU

7439
7175

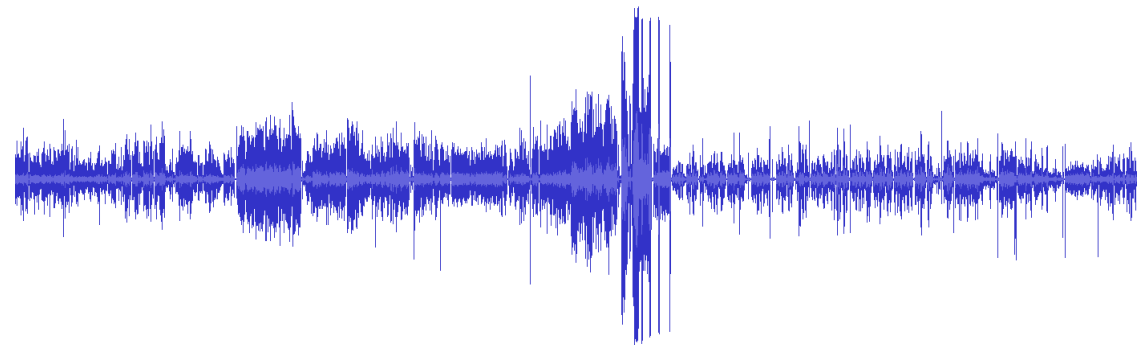
Smrt ljubice
2426.
(Jurna vas na Dolenjskem.)

Odbor za nabiranje
slovenskih narodnih pesmi.

Sensibile.

1. Snuč' sm dubu poštica, De m' je vmarla lubi-

ca. (n) Snuč' sm dubu poštica, De m' je vmarla ljubica,



262. NEZVESTA ZENA POBEGNE MOZU

K r a j: Kleče (Ljubljana), Gorenjsko
Z p s.: Matija Valjavec, pred 1856

Anzelj imel lepo ženó
pa se je vedno bal za njo,
da_b mu ne bila vgrabljena.
Enkrat se je primerilo,
5 da Anzelja doma ni blo,
da prideta popotna dva,
popotna dva, oba mladá,
za sladko vince dajata,
Alenčico napajata,

ti bom kosilo dobro dal,
da si ga bodeš sam jemal.«
40 Naprej, naprej mi spet zdrči,
da ogenj mu spod nog prši.
Kamor predaleč pridrči,
proti deveti deželi,
tja k oštariji zidani.
45 Na prag oštirka stala je,
šop ključev v roc držala je.

g g g / g f à / g(f) f / f..

38E8 hhh/ha_c/h(5)a/a..

(a)/c(5)c/chd/c(5)h/h..

/:h/d(5)d/d.e/c(5)c/c.

d/chh/ddd/cha/h...:/

T 118 / 254-272 1961 M. Šimen

Lajn-sku le - tu san se že - nu, e - no mla - do sem si vzev, lajn - sku

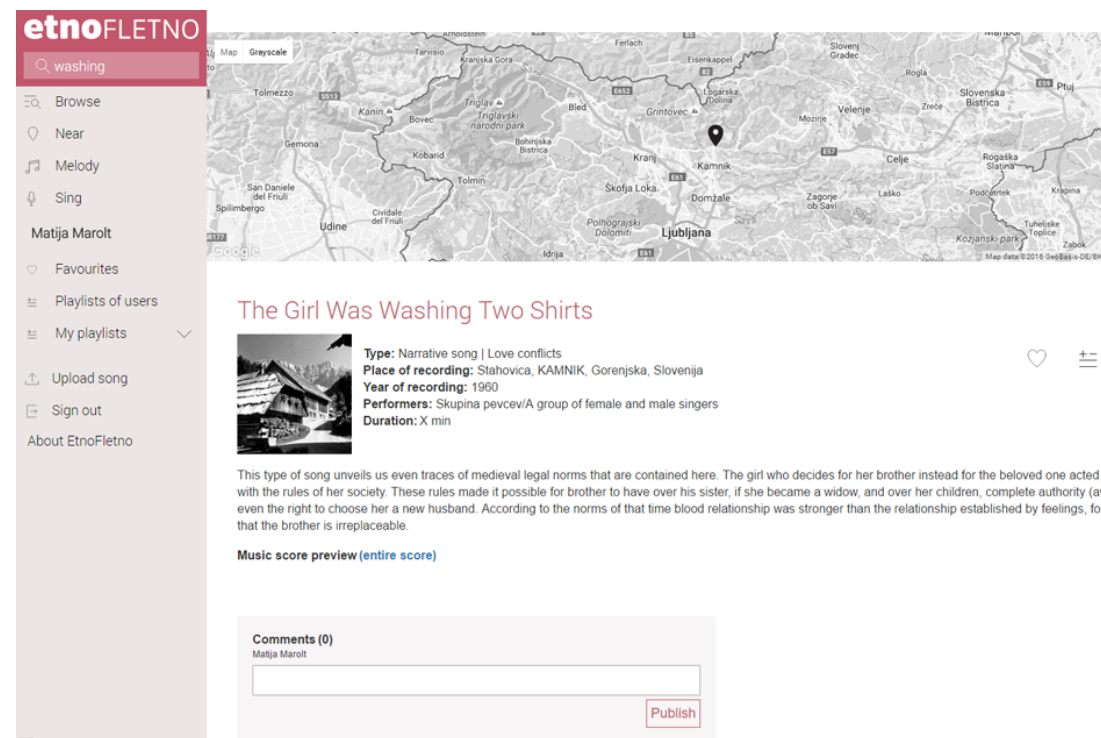
le - to san se u - že - nu, e - no mla - do sem si vzev.

Sodelovanje FRI in GNI

- GNI
 - velika količina raznolikih gradiv
 - zvočni posnetki
 - transkripcije
 - rokopisi
 - besedila
 - plesi
 - metapodatki
- Zamuden vnos in zbiranje metapodatkov
- Izziv: izluščiti čim več informacij iz gradiv
 - za lažjo analizo večje količine gradiv
 - za povečanje dostopnosti gradiv

Projekti

- [EtnoMuza](#), 2006
- [EtnoKatalog](#), TRP 2008-2011
- [Klik v domovino](#), CRP 2011-2012
- [EtnoFletno](#), eGradiva, 2014-2015



etnoFLETNO

washing

Browse

Near

Melody

Sing

Matija Marolt

Favourites

Playlists of users

My playlists

Upload song

Sign out

About EtnoFletno

The Girl Was Washing Two Shirts

Type: Narrative song | Love conflicts
Place of recording: Stahovica, KAMNIK, Gorenjska, Slovenija
Year of recording: 1960
Performers: Skupina pevcev/A group of female and male singers
Duration: X min

This type of song unveils us even traces of medieval legal norms that are contained here. The girl who decides for her brother instead for the beloved one acted in with the rules of her society. These rules made it possible for brother to have over his sister, if she became a widow, and over her children, complete authority (even the right to choose her a new husband). According to the norms of that time blood relationship was stronger than the relationship established by feelings, for that the brother is irreplaceable.

Music score preview ([entire score](#))

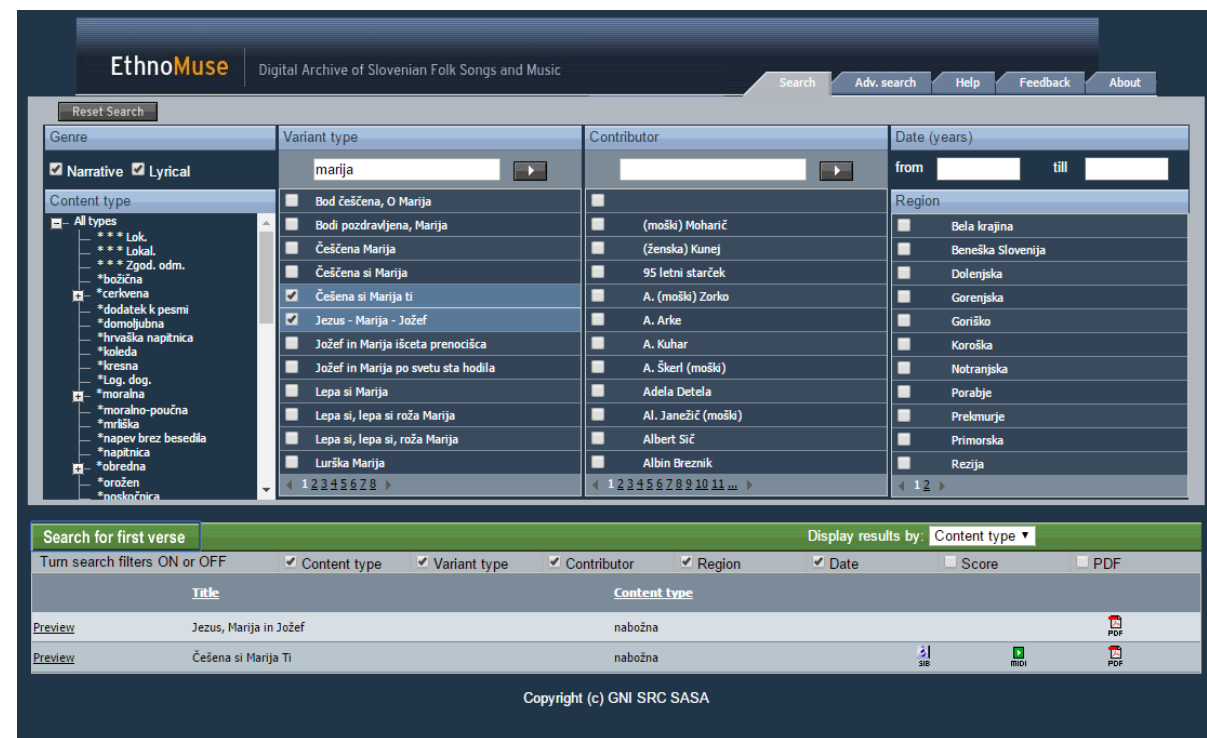
Comments (0)

Matija Marolt

Publish

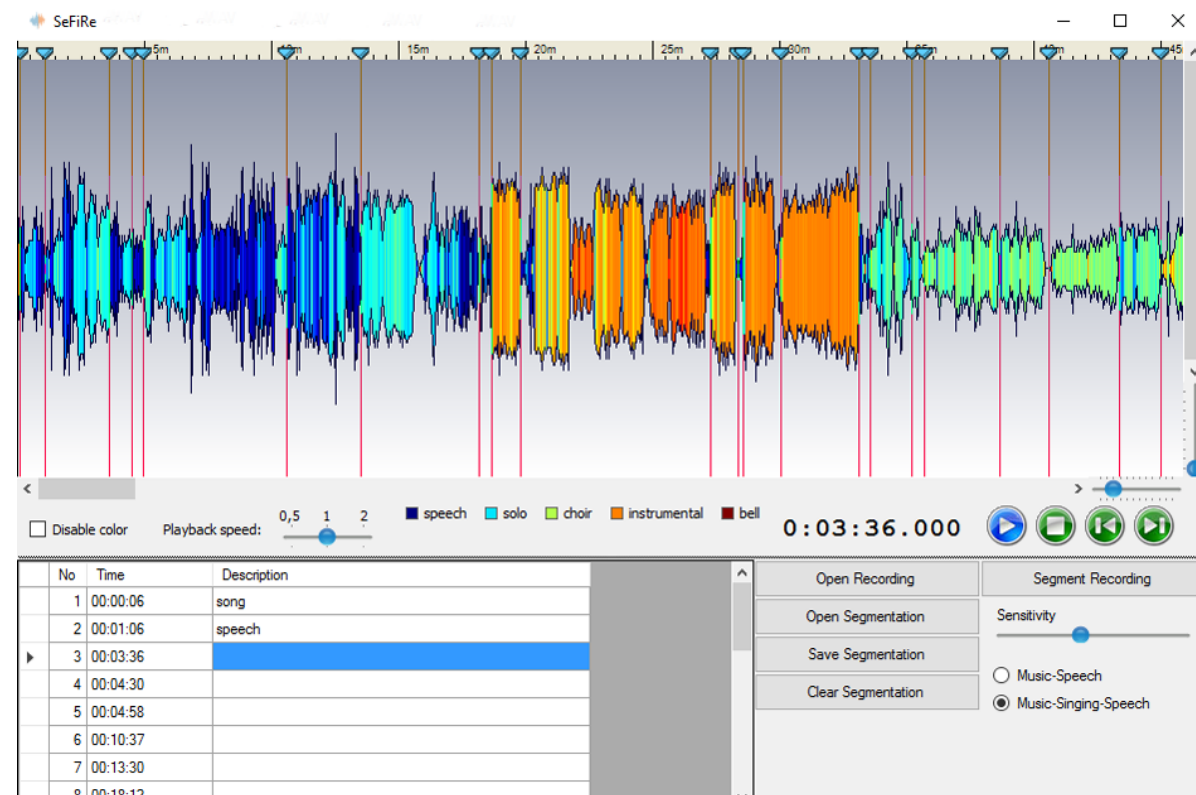
EtnoMuza – digitalni arhiv

- Začetek sodelovanja FRI in GNI (cca 2006)
 - Začetek vzpostavljanja digitalnega arhiva, temelj poznejših raziskav in projektov
- Orodja za zajem in shranjevanje metapodatkov v podatkovno bazo
 - rokopisi
 - terenski zapisniki, zvočni posnetki
 - slike
 - besedila
- [Spletni pregledovalnik](#) vsebin



Teme danes

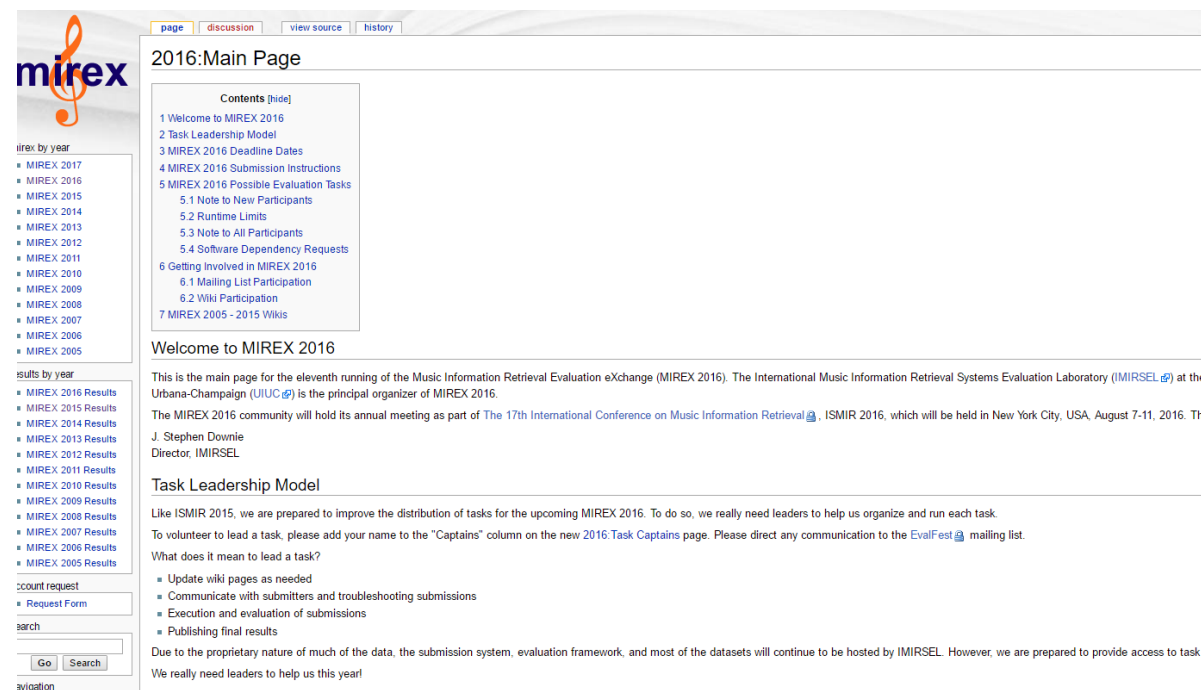
- Analiza zvočnih posnetkov
 - segmentacija
 - transkripcija
- Simbolični zapisi
 - iskanje
 - predstavitev
- Besedila
 - tematska analiza



Analiza zvočnih posnetkov

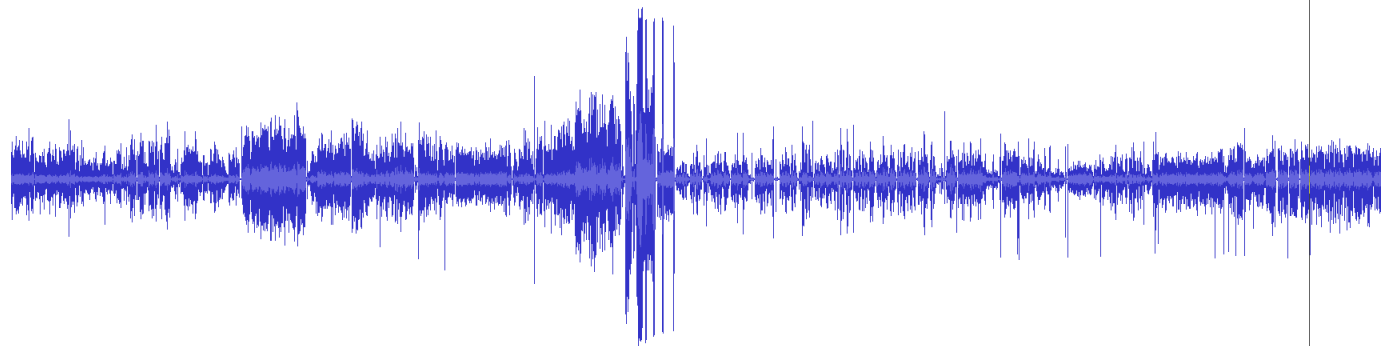
Analiza zvočnih posnetkov: MIREX

- MIREX: vsakoletna evalvacija algoritmov, v [2016](#):
 - Osnovni elementi
 - [Audio Onset Detection](#), [Offset Detection](#)
 - [Audio Chord Estimation](#), [Key Detection](#)
 - [Audio Melody Extraction](#)
 - [Multiple Fundamental Frequency Estimation & Tracking](#)
 - [Audio Beat Tracking](#), [Tempo Estimation](#), [Downbeat Estimation](#)
 - [Structural Segmentation](#)
 - [Audio Singing Voice Separation](#)
 - Značilnosti
 - [Audio Genre, Mood](#)
 - [Audio Tag Classification](#)
 - Iskanje
 - [Audio Fingerprinting](#)
 - [Audio Cover Song Identification](#)
 - [Audio Music Similarity and Retrieval](#)
 - [Set List Identification](#)
 - Izvedba
 - [Real-time Audio to Score Alignment \(a.k.a Score Following\)](#)
 - [Classical Composer Classification](#)



The screenshot shows the MIREX 2016 Main Page. At the top, there's a navigation bar with links for 'page', 'discussion', 'view source', and 'history'. The main heading is '2016:Main Page'. Below this, there's a 'Contents [hide]' section with a list of links: 1 Welcome to MIREX 2016, 2 Task Leadership Model, 3 MIREX 2016 Deadline Dates, 4 MIREX 2016 Submission Instructions, 5 MIREX 2016 Possible Evaluation Tasks (with sub-links for 5.1 Note to New Participants, 5.2 Runtime Limits, 5.3 Note to All Participants, and 5.4 Software Dependency Requests), 6 Getting Involved in MIREX 2016 (with sub-links for 6.1 Mailing List Participation and 6.2 Wiki Participation), and 7 MIREX 2005 - 2015 Wikis. To the left of the main content, there's a sidebar with 'mirex by year' and a list of links from MIREX 2017 down to MIREX 2005. Below that, there's a 'results by year' section with links for MIREX 2016 Results down to MIREX 2005 Results. Further down, there's a 'count request' section with a 'Request Form' link and a search bar. The main content area continues with a 'Welcome to MIREX 2016' section, followed by a 'Task Leadership Model' section. The 'Task Leadership Model' section includes text about improving task distribution for MIREX 2016, a call for volunteers to lead tasks, and a list of tasks to be completed: Update wiki pages as needed, Communicate with submitters and troubleshooting submissions, Execution and evaluation of submissions, and Publishing final results. It also mentions that due to the proprietary nature of much of the data, the submission system, evaluation framework, and most of the datasets will continue to be hosted by IMIRSEL, but access will be provided to tasks. The page ends with the statement 'We really need leaders to help us this year!'.

Analiza zvočnih posnetkov



- Specifike gradiv

- Izvajalci



- Posnetki



- Obstoječi pristopi večinoma niso primerni!

Računska (etno)muzikologija

- *Computational (Ethno)musicology*
- Področje se hitro razvija v zadnjem času
- Arhivi se digitalizirajo in so dostopni na spletu
 - [Alan Lomax](#), [Global Jukebox](#)
 - [CREM](#)
 - [British Library](#)
 - ...
- Potrebna je podpora za analizo, iskanje in odkrivanje znanj v velikih količinah gradiv



The screenshot shows the homepage of the Cultural Equity website. At the top, there is a header with the logo "cultural equity" and a navigation bar with links: HOME, RESEARCH CENTER, AVAILABLE PUBLICATIONS, ASSOCIATION FOR CULTURAL EQUITY, and ALAN LOMAX. Below the navigation bar, there is a sidebar with a list of categories: SOUND RECORDINGS (with a dropdown arrow), PHOTOGRAPHS, VIDEO RECORDINGS, RADIO PROGRAMS, DISCUSSIONS, INTERVIEWS & LECTURES, PERFORMANCE STYLE RESEARCH, GEO ARCHIVE, SONGS OF EARTH, TEACHING RESOURCES, FRIENDS & COLLEAGUES, and WEB RESOURCES. The main content area displays a table titled "Collection" with two columns: "Collection" and "Description". The table lists various recordings and interviews by Alan Lomax, including Bessie Jones, Big Bill Broonzy, Calypso Concert, Caribbean, Central Park Concert, Discussions, Lectures, and Interviews, Dominican Republic and Saint Eustatius, England, Haiti, Hally Wood, Ireland, Italy, Jean Ritchie, Kansas, Louisiana, Miscellaneous Recordings, Mississippi and Texas Church Recordings, Mississippi Prison Recordings, Morocco, Muhammad Glaus Waziri, New Orleans Jazz Interviews, New York Blues Interviews, Newport Folk Festival, Radio Programs and Ballad Operas, Romania, Scotland, Southern U.S., Soviet Union, Spain, Texas Gladden and Hobart Smith, Unknown or Unidentified, and Vera Hall.

Collection	Description
Bessie Jones 1961-1962	Alan Lomax, Zora Neale Hurston, and Mary Elizabeth...
Big Bill Broonzy 1952	Blues singer and guitarist Big Bill Broonzy (1893-...
Calypso Concert 1946	A live recording of Calypso At Midnight, a concert...
Caribbean 1962	In 1962 the British West Indies were on the verge ...
Central Park Concert 1965	A concert of Southern African American "antebellum...
Discussions, Lectures, and Interviews 1961-1994	Interviews and speeches given by Alan Lomax, and ...
Dominican Republic and Saint Eustatius 1967	In 1967, while on vacation in the Dominican Republ...
England 1951-1958	When Lomax started his British and Irish field wor...
Haiti 1936-1937	Alan Lomax's field trip to Haiti, undertaken at ag...
Hally Wood 1960	In the spring of 1960, on a brief vacation to Puer...
Ireland 1951 and 1953	This first anthology of Irish traditional music to...
Italy 1954-1955	From July 1954 through January 1955, Alan Lomax ma...
Jean Ritchie 1949 and 1950	Jean Ritchie was born and raised in Viper, Kentuck...
Kansas 1944	Images from a rodeo in Pittsburg, Kansas, in 1944...
Louisiana 1982-1985	
Miscellaneous Recordings 1950-1990	
Mississippi and Texas Church Recordings 1948	In early February 1948, a few days after the death...
Mississippi Prison Recordings 1947 and 1948	The Lomaxes, a...
Morocco 1967	In September 1967, Alan Lomax visited Morocco to m...
Muhammad Glaus Waziri 1966	A Pashto caravan song from Afghanistan, sung by Mu...
New Orleans Jazz Interviews 1949	In April 1949, as part of his research for a biogr...
New York Blues Interviews 1947	In 1947, using his own Presto disc recording machi...
Newport Folk Festival 1966	In 1966 Alan Lomax persuaded the Newport Folk Fest...
Radio Programs and Ballad Operas 1938-1957	Radio shows written, produced, and/or hosted by Al...
Romania 1964	These few recordings were made in the village of D...
Scotland 1951, 1953, and 1958	With guidance from Hamish Henderson, the MacLeans ...
Southern U.S. 1959 and 1960	In 1959 and 1960, Alan Lomax revisited the America...
Soviet Union 1964	In 1964 Alan Lomax visited the Soviet Union to att...
Spain 1952-1953	Alan Lomax's Spanish field recordings, made in 195...
Texas Gladden and Hobart Smith 1946	In 1946 Alan Lomax invited the prolific ballad sin...
Unknown or Unidentified	Recordings made in unidentified locations and by u...
Vera Hall 1948	These recordings of oral history, play songs, blue...

Segmentacija

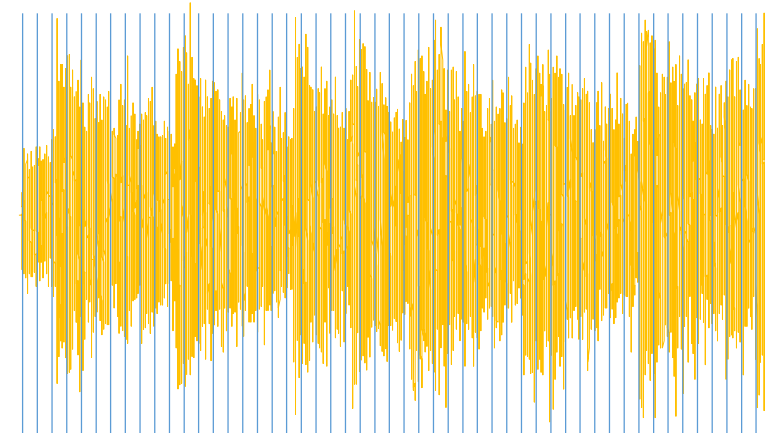
Segmentacija terenskih posnetkov

- Terenski posnetki
 - dokumenti s terena
- Posneto celotno dogajanje:
 - govor, petje, inštrumentali, pavze, šum ...
- Cilj:
 - avtomatski razrez posnetka na **smiselne enote**
 - klasifikacija enot (govor, petje, inštrumental)
- Standardni pristopi usmerjeni v *čiste* posnetke (npr. radijske oddaje ipd.)
 - težave pri *živih posnetkih z živim šumom*

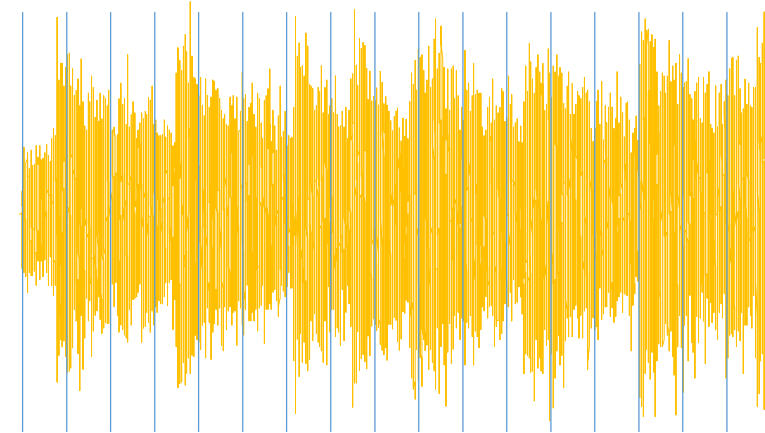


Segmentacija terenskih posnetkov - pristop

- Za vsak kratek odsek signala (50ms) izračunamo množico **značilk**, tipičnih za določanje **barve zvoka** in za razlikovanje med glasbo in govorom
 - MFCC, spektralna entropija itn.
- Izračunamo osnovne **statistike** teh značilk in njihovih **sprememb** preko daljših (3s) oken
 - npr. povprečje, varianco, varianco sprememb itn.



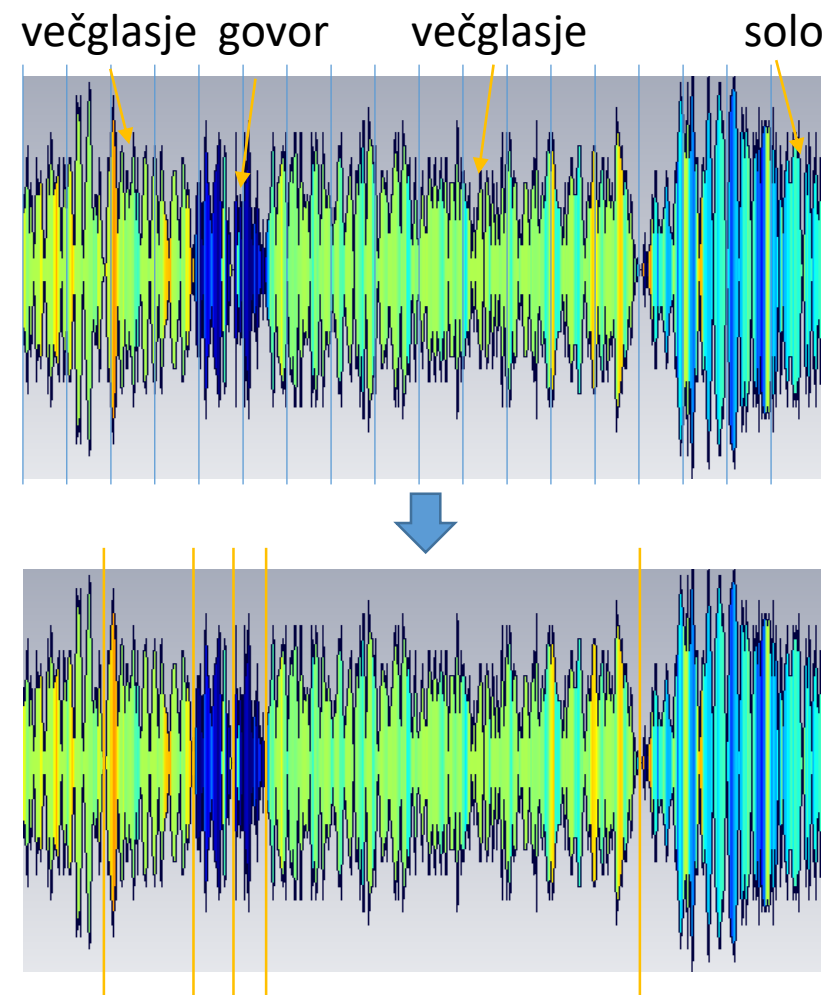
↓ značilke



↓ statistike značilk in *sprememb* značilk

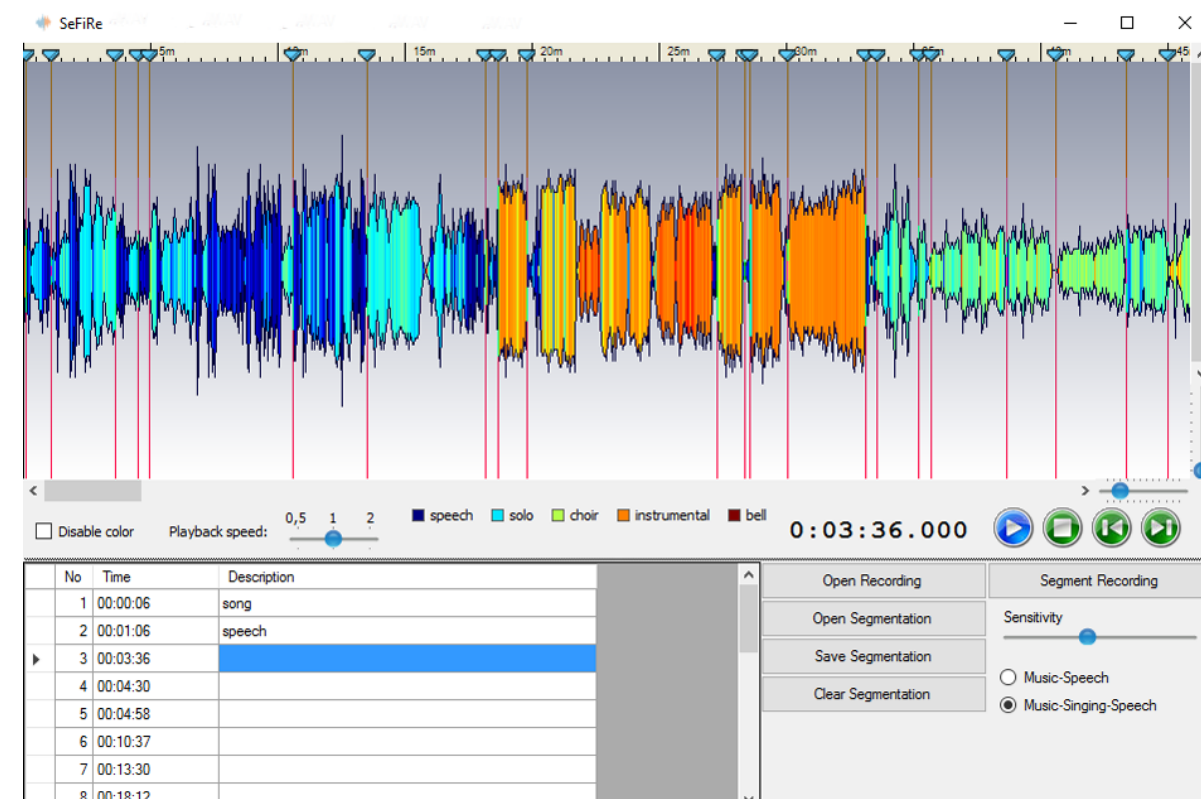
Segmentacija terenskih posnetkov - pristop

- Enostaven **klasifikator** (logistična regresija) klasificira statistike značilk v **5 razredov**:
 - govor
 - solo petje
 - večglasno petje (dva ali več glasov)
 - inštrumental (posnetek, ki vsebuje inštrumente)
 - pritrkavanje
- **Meje** med enotami postavimo glede na:
 - daljše obdobje tišine oz. šuma
 - spremembo porazdelitve razredov (npr. govor v petje ipd.)



SeFiRe

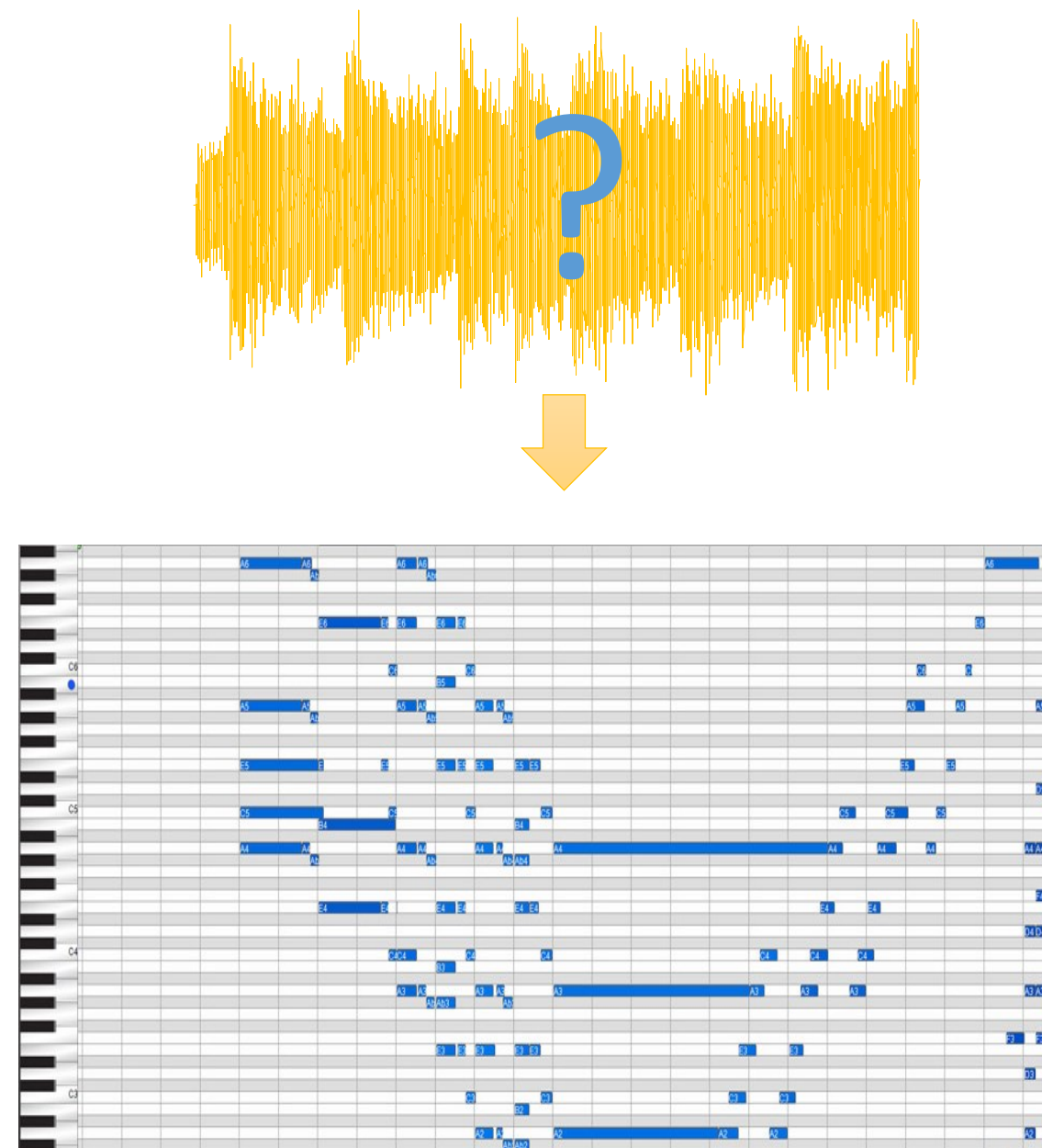
- Orodje za segmentacijo
 - [prosto dostopno](#)
 - **barvno** kodiranje razredov
 - avtomatsko postavljanje mej
 - spreminjanje mej, shranjevanje
- Dodaten algoritem prilagojen za razlikovanje govor-glasba
 - deluje tudi za druga (neslovenska) gradiva
 - MIREX 2015: Najboljši segmentacijski algoritem (terenski posnetki *British Library*)



Transkripcija

Transkripcija

- Iskanje not v zvočnem signalu
 - višina, začetek, *trajanje*
 - relevanten problem, ročna transkripcija je zelo zamudna (in v domeni ekspertov)
 - veliko raziskav
- Težek problem
 - v splošnem natančnost še vedno le okoli 50%
 - pričakovati je izboljšanje z uporabo globokih arhitektur
- Tudi nepopolna transkripcija je lahko uporabna
 - za iskanje
 - kot pomoč pri ročnem transkribiranju
- Boljši rezultati na omejenih domenah
 - pritrkavanje
 - večglasno petje

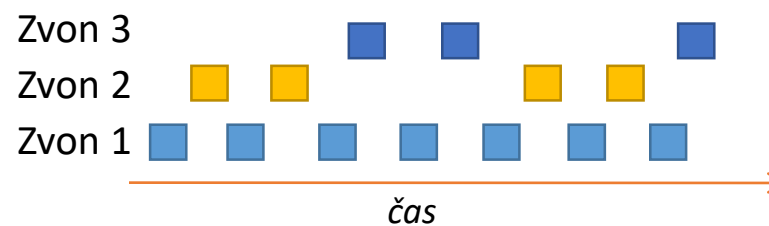
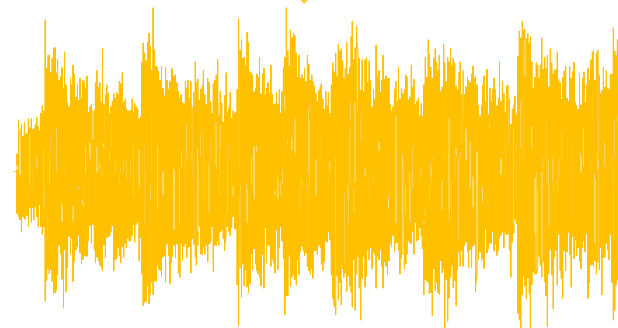


Pritrkavanje

- Tradicija igranja melodij/ritmičnih vzorcev na cerkvene zvonove

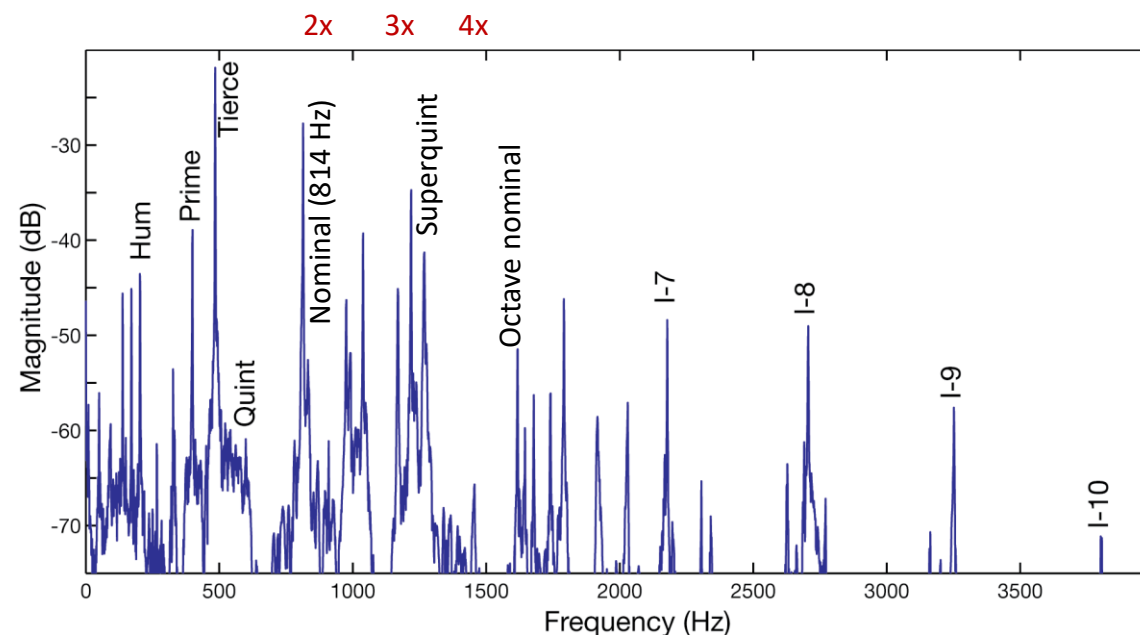
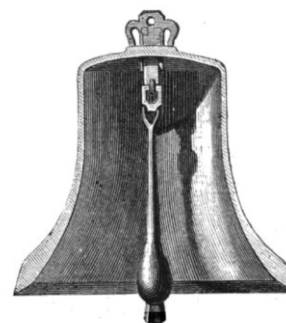


- **Transkripcija** pritrkavanj
 - zvočni posnetek -> ritmičen vzorec
- Omogoča
 - iskanje vzorcev
 - primerjavo vzorcev
 - analizo korpusa



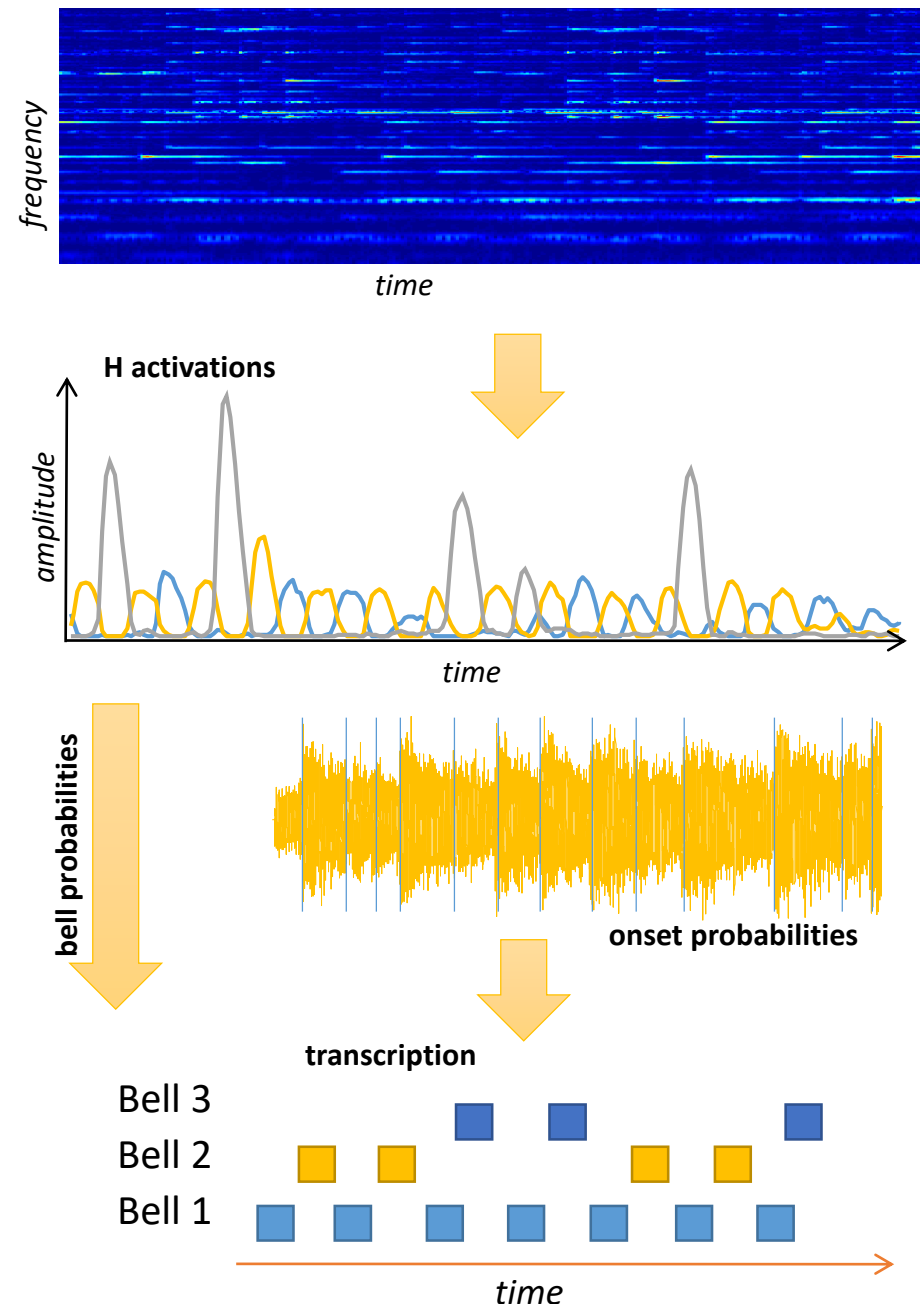
Transkripcija pritrkavanj - postopek

- Iskanje zvonov
- Vsak zvon je **drugačen**
 - drugače uglašen, ima drugačno razporeditev frekvenčnih komponent
- Toda...
 - **regularnosti** v spektru obstajajo
 - pri pritrkavanju zvoni niso zaigrani hkrati
- Algoritem za **iskanje zvonov**, ki upošteva korelacije amplitudnih ovojníc frekvenčnih komponent istega zvona
 - najde **število zvonov**
 - ugotovi približen **frekvenčni spekter** zvona
 - 95% natančnost (mera F1)



Transkripcija pritrkavanj - postopek

- **Nenegativna matrična faktorizacija** razstavi časovno-frekvenčno predstavitev signala na
 - spektre zvonov
 - aktivacije (amplitude) zvonov
- Inicializacija z rezultati algoritma za iskanje zvonov
 - Dodatne omejitve za redkost spektrov in nekoreliranost aktivacij
- Verjetnostni model, ki upošteva
 - **verjetnosti zvonov** glede na aktivacije
 - **verjetnosti začetkov** zvonov glede na analizo signala in detekcijo začetkov
 - **verjetnosti prehodov** med zvonovi



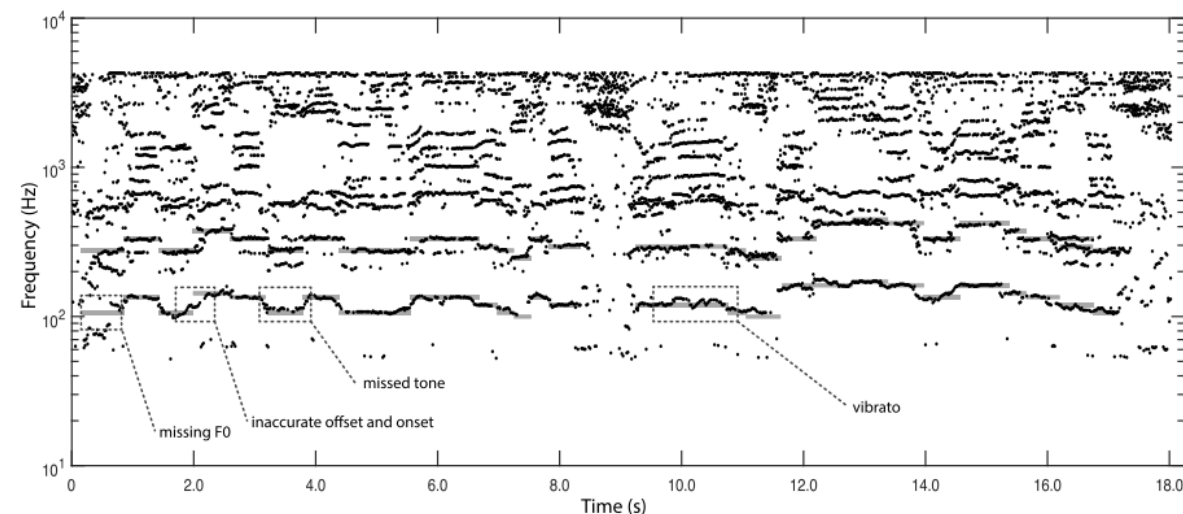
Transkripcija

- Rezultat je seznam zvonov in začetnih časov
 - natančnost cca 89% (mera F1)
 - [primeri transkripcij](#)
- Uporaba tudi drugje
 - iskanje melodij ur 18. stoletja ([The Meertens Institute](#))
 - primer:



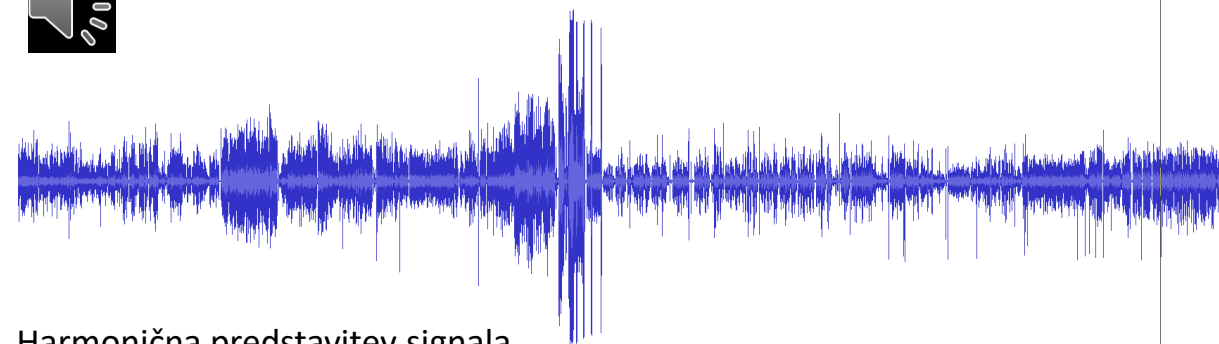
Transkripcija večglasnih pesmi

- Transkripcija solo (enoglasnih) pesmi je večinoma rešen problem
 - enostavno je najti višino tona
 - pretvorba v note ni nujno pravilna
- Transkripcija večglasnega petja je težavna
 - pristopi za enoglasje niso uporabni
 - barva glasov je spremenljiva
 - obstoječi pristopi večinoma ciljajo na inštrumentalno glasbo
- Uporaba strukture za izboljšanje transkripcije

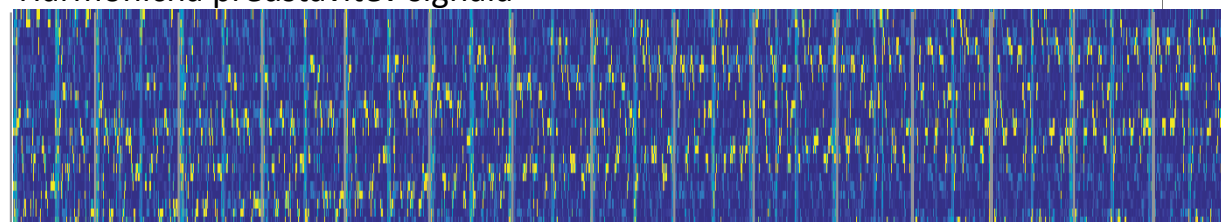


Struktura

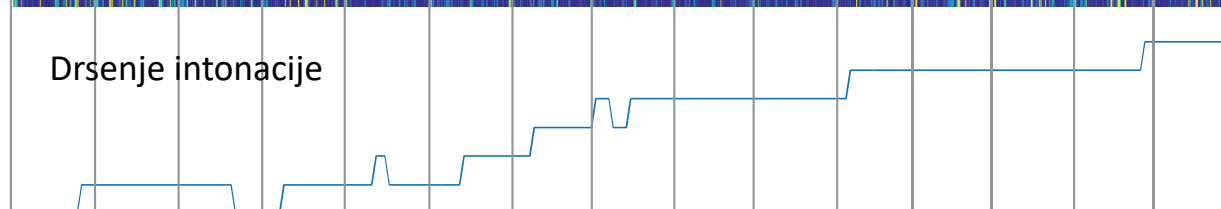
- Pesmi so sestavljene iz ponavljajočih delov - kitic
 - variacije zaradi netočnega petja, drsenja intonacije, spreminjanja tempa, napak, šuma
- Kako jih najdemo
 - harmonična predstavitev pesmi (kromatski vektorji)
 - dinamično ukrivljanje časa (*dynamic time warping*) za izračun razdalj med deli pesmi
 - ocena drsenja intonacije, kompenzacija
 - skriti markovski model za postavitve mej
- Natančnost
 - $F_1 = 0.76$ (pesmi GNI in iz zbirke inštituta Meertens)



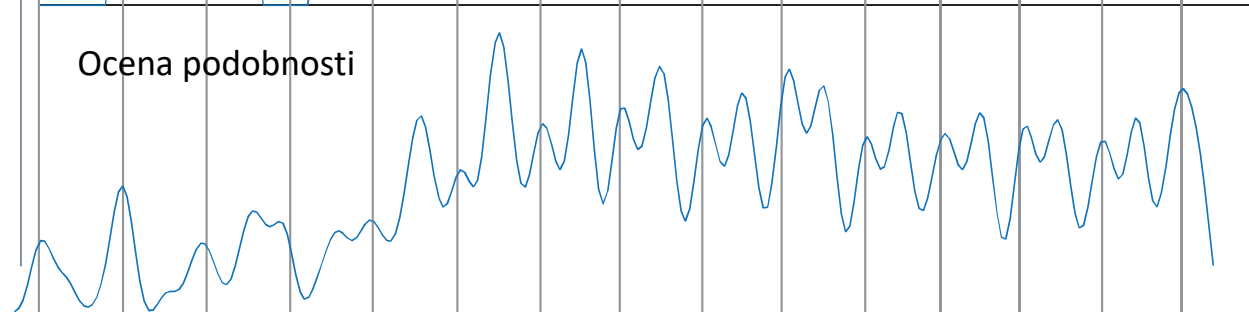
Harmonična predstavitev signala



Drsenje intonacije

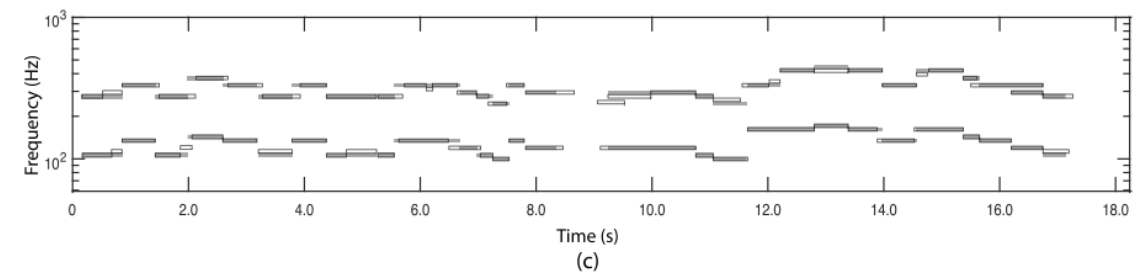
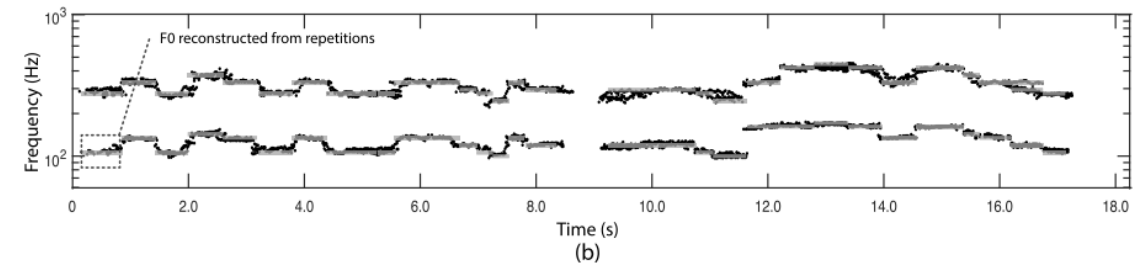
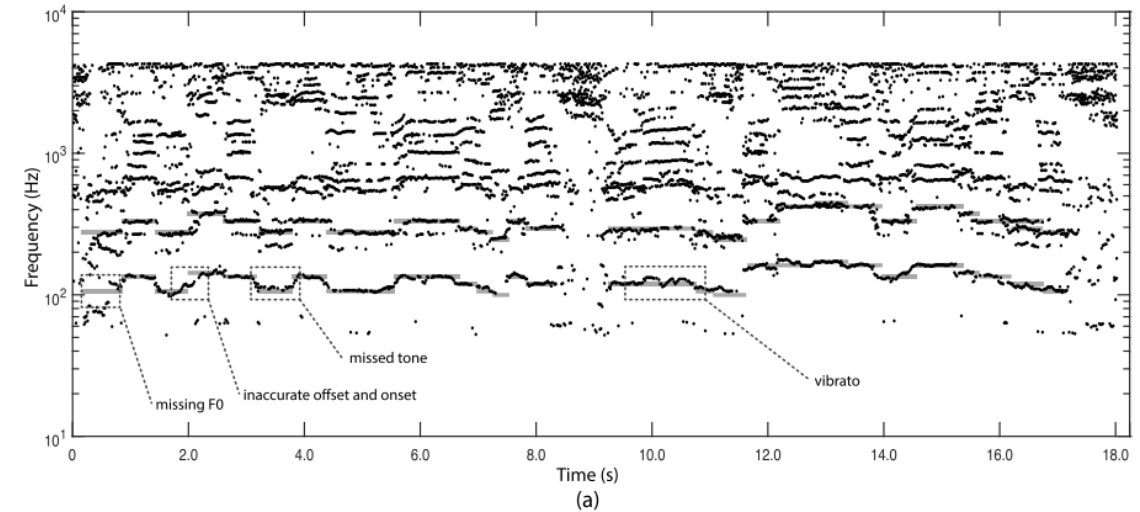


Ocena podobnosti

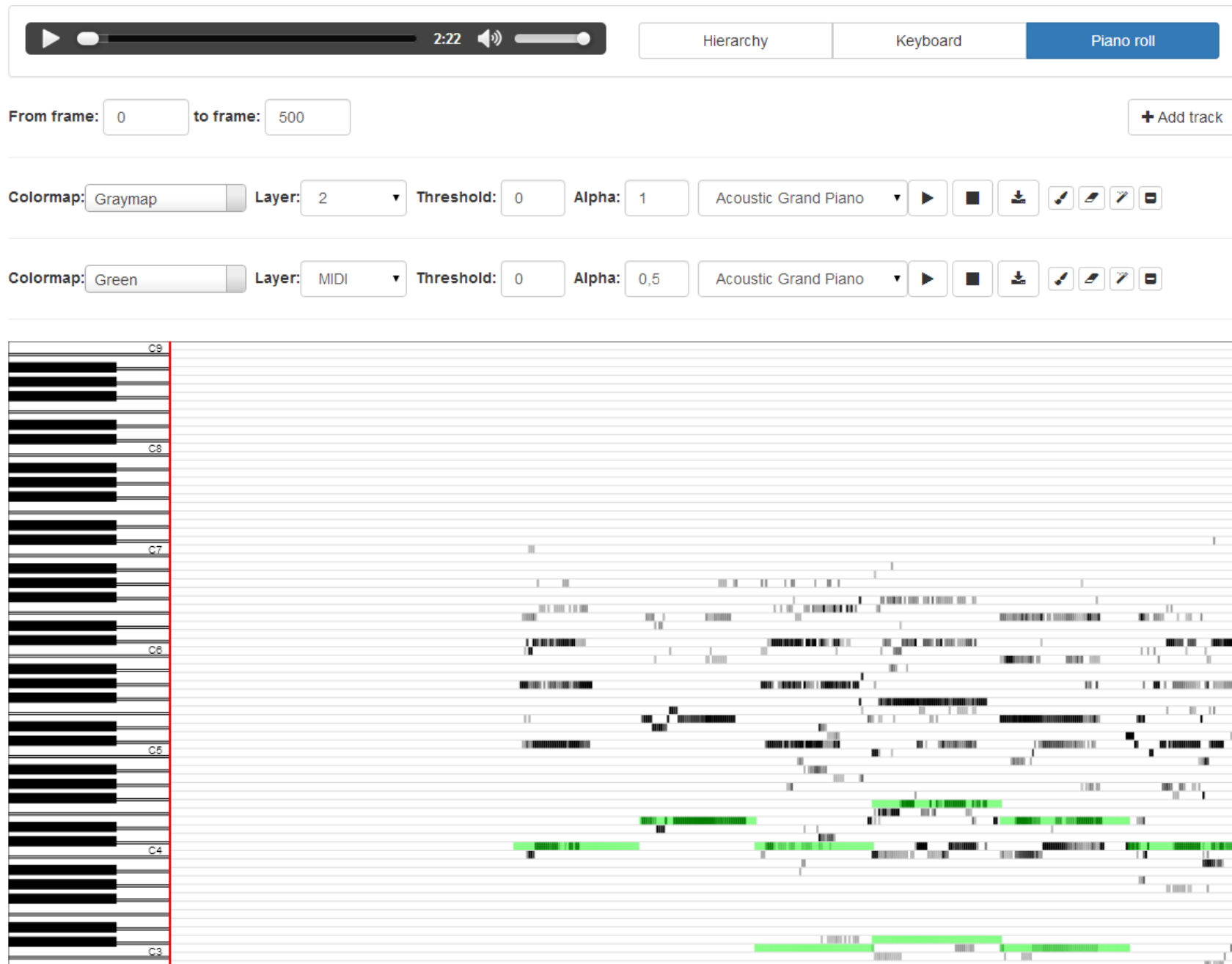


Transkripcija

- Naš pristop
 - vzamemo obstoječi pristop za transkripcijo
 - ga izboljšamo z uporabo strukture (ponavljanj)
- Izboljšanje SoA za cca 20% (F1)



Orodje



Reversing

Simbolični zapisi

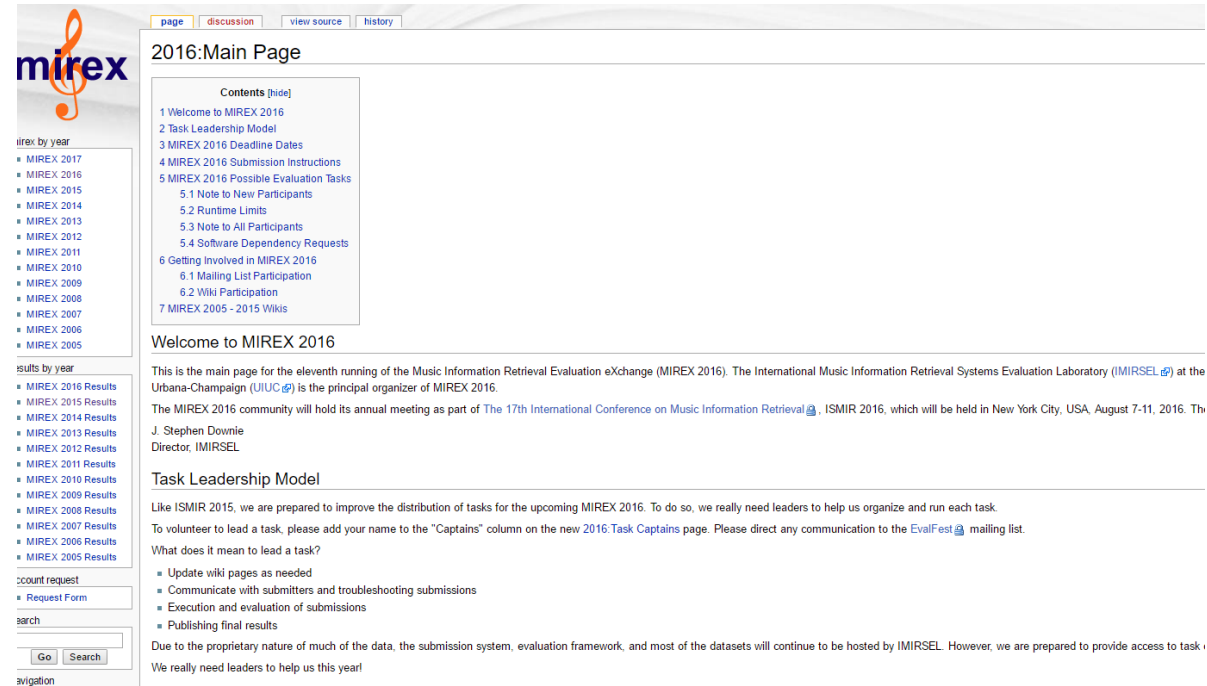
Simbolični zapisi glasbe

- Notni zapisi
 - rokopisi (prepis)
 - transkripcije zvočnih posnetkov
- Kaj bi počeli z njimi
 - iskanje
 - primerjava melodij (podobnost, variante)
 - vzorci



Simbolični zapisi: MIREX

- MIREX: vsakoletna evalvacija algoritmov, v [2016](#):
 - Iskanje
 - [Symbolic Melodic Similarity](#)
 - [Query by Singing/Humming](#)
 - [Query by Tapping](#)
 - Zbirke
 - [Discovery of Repeated Themes & Sections](#)
 - [Grand Challenge on User Experience](#)
 - Izvedba
 - [Real-time Audio to Score Alignment \(a.k.a Score Following\)](#)



The screenshot shows the MIREX 2016 Main Page. At the top, there's a navigation bar with links: page, discussion, view source, and history. The main heading is "2016:Main Page". Below it, there's a "Contents [hide]" section with a list of links: 1 Welcome to MIREX 2016, 2 Task Leadership Model, 3 MIREX 2016 Deadline Dates, 4 MIREX 2016 Submission Instructions, 5 MIREX 2016 Possible Evaluation Tasks (with sub-links 5.1 Note to New Participants, 5.2 Runtime Limits, 5.3 Note to All Participants, 5.4 Software Dependency Requests), 6 Getting Involved in MIREX 2016 (with sub-links 6.1 Mailing List Participation, 6.2 Wiki Participation), and 7 MIREX 2005 - 2015 Wikis. To the left of the main content, there's a sidebar with "mirex by year" (listing MIREX 2017 down to 2005) and "results by year" (listing MIREX 2016 Results down to 2005 Results). Below the sidebar, there's a "count request" section with a "Request Form" and a "search" section with a "Go" button and a "Search" button. The main content area starts with "Welcome to MIREX 2016" and a paragraph about the eleventh running of the Music Information Retrieval Evaluation eXchange (MIREX 2016). It mentions the International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) at the Urbana-Champaign (UIUC) as the principal organizer. It also mentions the MIREX 2016 community meeting as part of the 17th International Conference on Music Information Retrieval (ISMIR 2016) in New York City, USA, August 7-11, 2016. Below this, there's a "Task Leadership Model" section with a paragraph about improving the distribution of tasks for the upcoming MIREX 2016 and a list of tasks to lead: Update wiki pages as needed, Communicate with submitters and troubleshooting submissions, Execution and evaluation of submissions, and Publishing final results. It also mentions that due to the proprietary nature of much of the data, the submission system, evaluation framework, and most of the datasets will continue to be hosted by IMIRSEL, but access will be provided to task leaders.

Format zapisov na GNI

- Vnos v programu Sibelius
 - komercialen, lasten format
- Pretvorba v [MusicXML](#)
 - iskanje
 - analiza
 - prikaz (na spletu)
- Alternativa bi bila [MEI](#)
 - obstaja možnost pretvorbe
- Pretvorba v MIDI
 - predvajanje (na spletu)
- Posebnosti
 - vse v G
 - *tonus finalis*

267/62 Nezvesta gospa s tremi stražarji/B (GNI M 45873)

Tam dol na rav - nem po lju sto - ji en be - li, be - li grad, tam

dol na rav - nem po lju sto - ji en be - li grad. T.F.

```
<note>
  <pitch>
    <step>D</step>
    <octave>4</octave>
  </pitch>
  <duration>384</duration>
  <voice>1</voice>
  <type>eighth</type>
  <staff>1</staff>
  <beam number='1'>begin</beam>
  <lyric number='1'>
    <syllabic>end</syllabic>
    <text>Tam</text>
  </lyric>
</note>
```

Iskanje na podlagi not

- Vgrajeno v spletne iskalnike [EtnoMuza](#) in [EtnoFletno](#)
- Vprašanja
 - neodvisno od transpozicije?
 - kako delati z večglasji?
 - upoštevati dolžine not?
 - približno iskanje?

The screenshot shows the EtnoFletno web application. At the top is a red header with the text "etnoFLETNO" and a search bar labeled "Search string". Below the header is a sidebar menu with options: "Browse", "Near", "Melody", "Sing", "User", "Favourites", "Playlists of users", "My playlists", "Upload song", and "Sign out". At the bottom of the sidebar is a "PLAYER" section showing a play button, a progress bar at 0:26, and the song title "Anzel, Are You Alive or Dead?". The main content area displays a grid of search results, each featuring a thumbnail image, a title, and a location name.

Listen

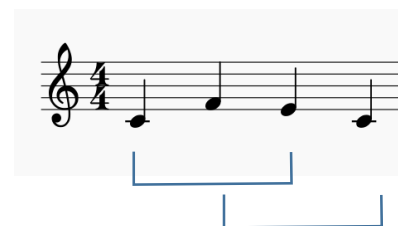
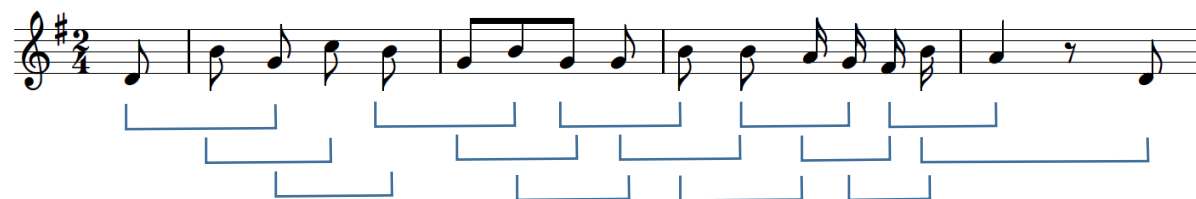
Results (208)

A grid of 15 search results from the EtnoFletno application. Each result consists of a thumbnail image, a title, and a location name. The results are as follows:

Thumbnail	Title	Location
	Anzel, Are You Alive or Dead?	Gorenjska
	I Got Married Last Year	Prekmurje
	Oh, Women, Come Here	
	A Woman's Head Is Too Weak	Prekmurje
	The Sun Has Already Set	
	We've Already Husked the Corn	
	The Sun is Already Shining Low	Stajerska
	The Wheat is Already Ripening in the Field	Stajerska
	Seven Years Have Passed	
	The Sad Mother Stands under the Cross	Dolenjska
	There Was a Forest Too Vast	
	Two Students Walk Along the Road	
	A Student Loved a Girl	Stajerska
	There are Forty of Us Prisoners	
	Four Musicians Are Playing	

Iskanje na podlagi not

- Uporabljen standarden pristop
 - iskanje z n-grami
- Indeksirajo se n-terice (3-5) not preko vseh skladb
 - indeksiramo intervale – **neodvisnost** od transpozicije, natančno do poltona
 - ne indeksiramo **trajanj** (bi jih lahko)
 - pri večglasjih
 - indeksiramo **vsak glas posebej** (če je informacija na voljo)
 - sicer samo zgornji glas (lahko vodi do napak)
- Pri iskanju
 - rangiramo odgovore glede na število n-gramov, ki se ujemajo – približno iskanje



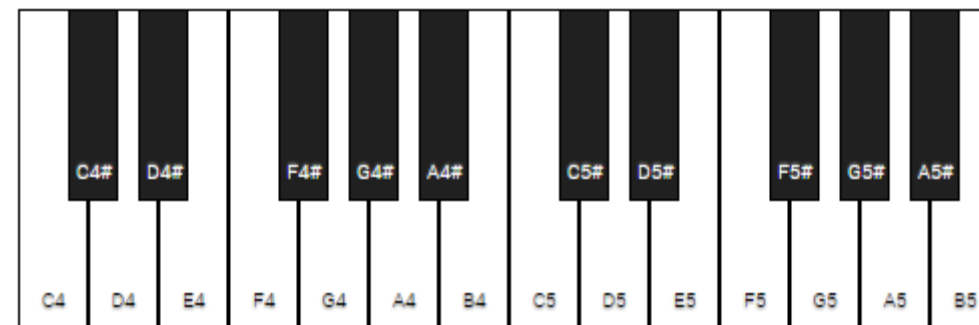
Iskanje na podlagi not

- Kako podati poizvedbo?
 - klaviatura
 - ročni zapis (ustrezna notacija)
- Prikaz not – v brskalniku?
 - Sibelius ima svojo rešitev – zahteva vtičnik
 - JavaScript okolja so še nekoliko okorna, se izboljšujejo
 - [vexflow](#)
 - [AbsWeb](#)
 - [Verovio](#)
 - Za EtnoFletno smo generirali slike
 - ne omogočajo npr. označevanja kateri del melodije smo našli

Melodija

Vnesite vsaj tri note za iskanje po melodiji.

G/4 A/4 B/4 B/4 C/5



< C4-B5 >

PREDVAJAJ



IZBRIŠI



Iskanje na podlagi zapete melodije

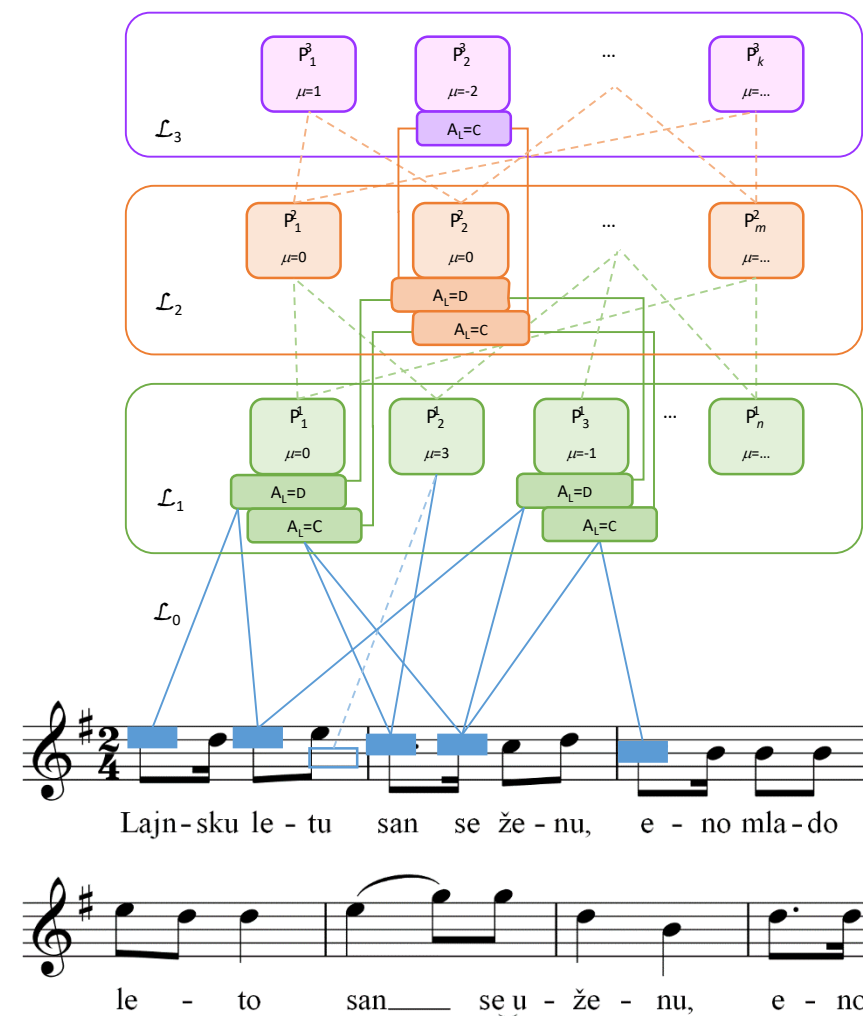
- *Query-by-humming*
- Za uporabnike, ki ne obvladajo notnega zapisa
 - je pa odvisno od pevskega talenta
- Iskanje bolj približno kot na podlagi not
 - upoštevamo, da je tudi poizvedba lahko napačna
 - napačno prepoznana



0:00

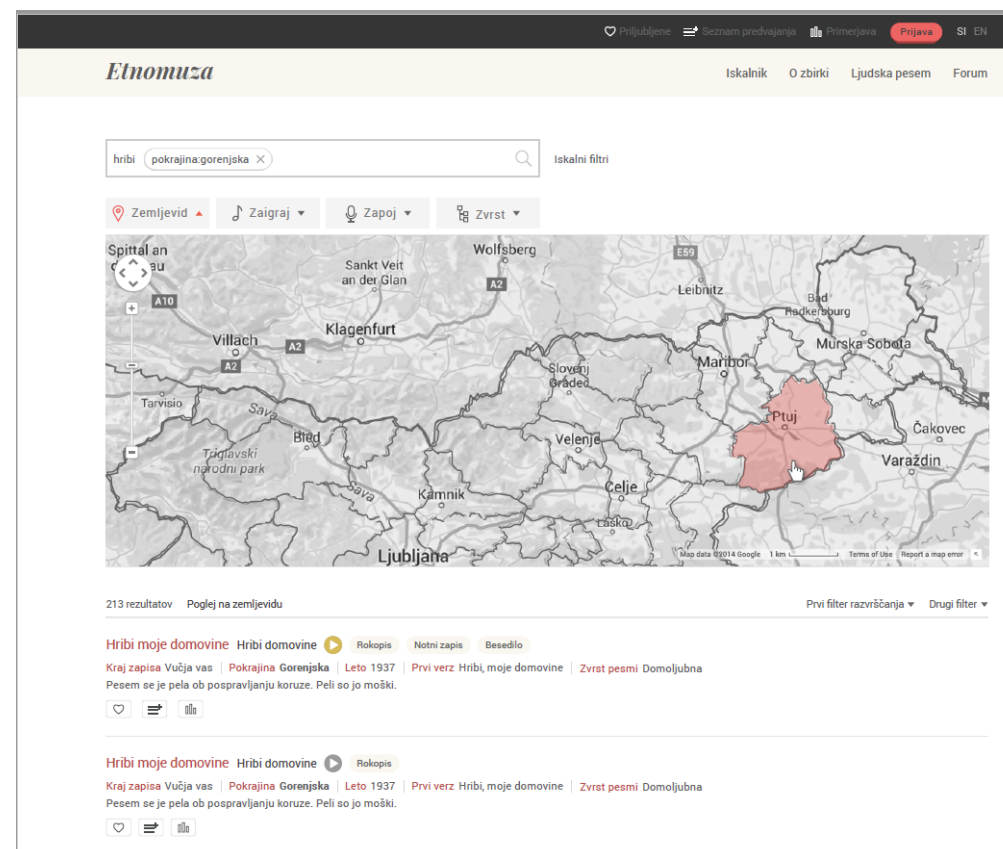
Iskanje vzorcev

- Kompozicionalni hierarhični model
 - globoka arhitektura
 - model se nauči hierarhijo vzorcev glede na statistične pojavitve zaporedij
- Uporaben za iskanje vzorcev in primerjavo pesmi
 - melodičnih
 - ritmičnih
 - vzorci se lahko spreminjajo
- Potencialna uporaba
 - računanje podobnosti, gručenje pesmi
 - uvrščanje melodij v variantne tipe
 - primerjave med tipi glede na vzorce
- Na evalvaciji *MIREX 2015*: dobra uvrstitev pri iskanju ponovljenih tem in sekcij

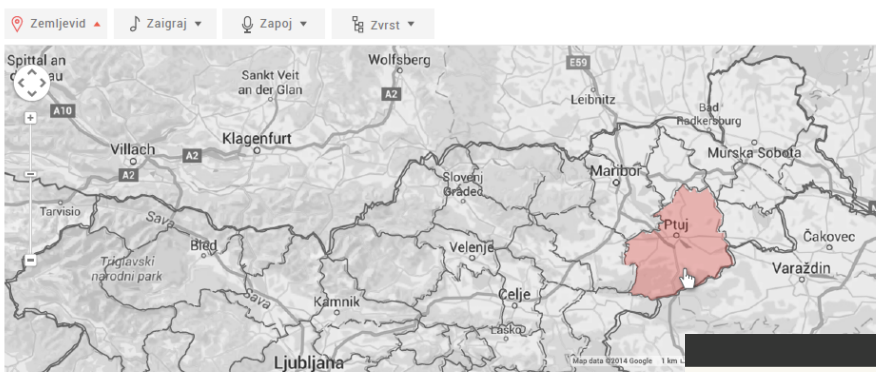
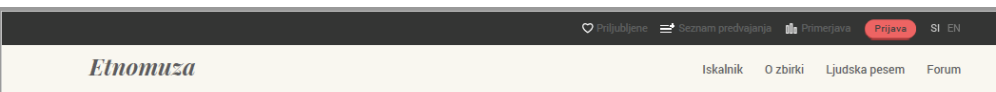


Uporabniški vmesniki za iskanje

- Kdo so **uporabniki**?
 - raziskovalci
 - izvajalci, folklorne skupine ...
 - javnost
- Kakšni so njihovi cilji?
- Načrt prototipov vmesnik



Uporabniški vmesniki za iskanje



213 rezultatov Poglej na zemljevidu

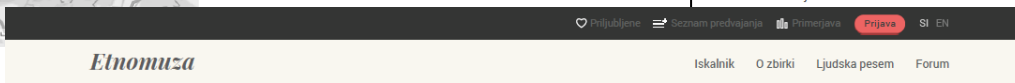
Hribi moje domovine Hribi domovine Rokopis Notni zapis Besedilo

Kraj zapisa Vučja vas Pokrajina Gorenjska Leto 1937 Prvi verz Hribi, moje domovine Zvrst pesmi Domoljubna
Pesem se je pela ob pospravljanju koruze. Peli so jo moški.



Hribi moje domovine Hribi domovine Rokopis

Kraj zapisa Vučja vas Pokrajina Gorenjska Leto 1937 Prvi verz Hribi, moje domovine Zvrst pesmi Domoljubna
Pesem se je pela ob pospravljanju koruze. Peli so jo moški.



hribi Iskalni filtri

Zemljevid Zaigraj Zapoj Zvrst

213 rezultatov Poglej na zemljevidu

Hribi moje domovine Hribi domovine Rokopis Notni zapis Besedilo

Kraj zapisa Vučja vas Pokrajina Gorenjska Leto 1937 Prvi verz Hribi, moje domovine Zvrst pesmi Domoljubna
Pesem se je pela ob pospravljanju koruze. Peli so jo moški.



Hribi moje domovine Hribi domovine Rokopis

Kraj zapisa Vučja vas Pokrajina Gorenjska Leto 1937 Prvi verz Hribi, moje domovine Zvrst pesmi Domoljubna

00:05 3:45 Hribi moje domovine Heart icon List icon Map icon

Hribi moje domovine Hribi domovine Rokopis

Kraj zapisa Vučja vas Pokrajina Gorenjska Leto 1937 Prvi verz Hribi, moje domovine Zvrst pesmi Domoljubna
Pesem se je pela ob pospravljanju koruze. Peli so jo moški.

Heart icon List icon Map icon 3 različice



hribi Iskalni filtri

Hribi moje domovine

Originalni posnetek Sintetiziran zvok Rokopis Notni zapis Besedilo

Osnovni podatki

Naslov prvega verza: Lore m ipsum dolor amet
Zvrst: Amet

Izvor in zapis

Kraj zapisa: Cven
Bližnji kraj zapisa: Ljutomer
Pokrajina zapisa: Prlekija
Država zapisa: Slovenija
Leto zapisa: 1958
Izvajalec: Tereza Horvat
Kraj izvora: Cven

Priljubljene Seznam predvajanja Primerjava

Notni zapis



Besedilo

Marička sedi na kamenu.
Marička češe zlate lase.
Marička začne jokati.
Pa k njej pristopi Karol brat.

»Zakaj pa jokaš ti, dekle?«
»Zato, k bom mogla jutri umret.«
Marička je zdej angelček.

Marička sedi na kamenu.
Marička češe zlate lase.
Marička začne jokati.
Pa k njej pristopi Karol brat.

»Zakaj pa jokaš ti, dekle?«
»Zato, k bom mogla jutri umret.«
Marička je zdej angelček.

Besedila

Besedila pesmi

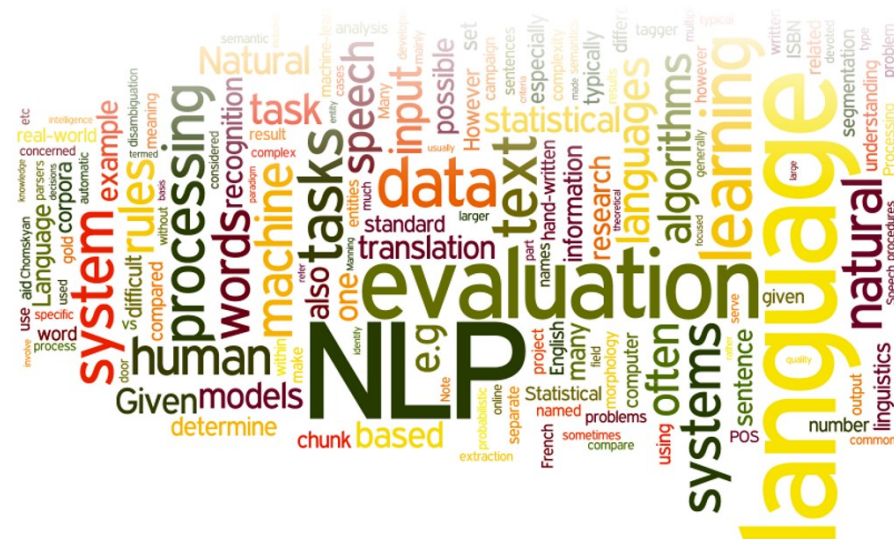
- Analiza **besedil** – ali lahko ugotovimo:
 - vsebinski **tip** (npr. ljubezenska, moralna, pivska ...)
 - **povezave/podobnosti** med pesmimi
- Melodije so v ljudski pesmi pogosto izposojene, se prenašajo med pesmimi



ljubezenska?
moralna?
legendarna?
govori o smrti?
pivska?
družinska?

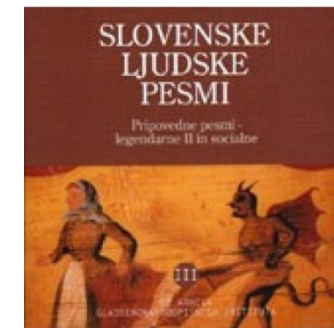
Cilj

- Eksperimenti
 - lahko v besedilih ljudskih pesmi odkrijemo tematike?
 - lahko besedila uvrstimo v skupine glede na tematike?



Korpus

- Na novo ustvarjen iz knjig *Slovenske ljudske pesmi I-V*, ZRC SAZU (1970-2007)
 - novejše v elektronski obliki (Word)
 - polavtomatska pretvorba, izločanje besedil in metapodatkov
 - starejše
 - skenirane in prepoznane z OCR
- 4095 Slovenskih ljudskih pripovednih pesmi
 - od 18. stoletja naprej
 - 349 variantnih tipov
 - od 1 do 180 pesmi na variantni tip



Pretvorba

- Ločevanje besedila od metapodatkov in drugih
- Pri OCR sprotno učenje posebnih simbolov
- Izločanje drugih metapodatkov (kraj, informatorji, zapisovalci itn.)
 - težave z nekonsistentnostjo...
- Preslikava in zapis vseh posebnih simbolov v Unicode

mikrofon vključen. Sprva raztrgana, skromna obnova je mimogrede prerasla v samostojno verzijo naše pesmi — B. V nji najdemo nekaj novih prvin, ki jih tu natisnjena verzija (A) nima:

Matjaža so dali v ječo. Linčica se je vanj zaljubila, menda je bil lep človek. Linčica je dala »dormido« (uspavalni napoj ali prašek) očetu in materi, kasneje pa je uspavala tudi stražo. O mostu, ki je slonel na lažnih opornikih in ki sta ga ubežnika

zadnji hip pred prihodom nasledovavcev zažgala, slišimo, da je bil na meji (turške dežele).

Za pesem, ki je na samem začetku okrnjena, pomenijo ti dodatki dragoceno dopolnilo. Ob navedenosti, da je Linčica dala očetu in materi napoj za spanje (dopolnilo pevkine sestre Marije), se je Ana celo glasno začudila, kot bi hotela reči: »Glej glej, tega pa nisem vedela!« ali »Na to pa sem pozabila!«

6.

Kraj: Bila v Reziji

Pela in pripovedovala: Ana Buttolo, vd. Zanetti, Pécawa (1894)

B

Z p.s.: Matičetov-Vodušek, 20. junija 1963



1 Od 4. kitice dalje na tem mestu »c«.

- | | |
|--|---|
| 1. Linčica Turkinčica
na wze-la noga Wóngarja. | 9. Dwa dni na jě mu púlela
nú na jě mu fys rakla: |
| 2. To bila noga kraja šči
nu won to bil de Wóngar krej. | 10. »Da ti me vliku jes plažáš
nu ti maš víde' me womože» |
| 3. ke an je bil prožonir
nu za tri dni je bil zažán. | 11. ke ja čon te wiğá' wozdž,
ja bon te fis wiğál' wozdž.» |
| 4. Nu Linčici je zaplašél,
na je prosila grač(jico). ¹ | 12. »Káko bėj maš mé wóženě,
ke já si wžě wóženjen, |
| 5. Nu won je jo koncčđinal,
na drě delá pokléknula | 25 13. a) ma já man šće dwa bratra tapar
iši,
bój lipča nekuj mle.» |
| 6. nu nōge na mu bášnula
nu drě ga bášnula pa njagá. | 13. b) man dwa lipča bratra někoj mle, ²
ke téđwa tó bo te wzéto.» |
| 7. »Da Linčeca Turkinčeca,
wse tō ke boš me bārala, | Nu dwa dni na je mu púlela jěst
anu kjūče na je ískala
wod te vlíkih dur tow Fránciji,
ke grejo skuz,
na vlíka galeríja skuz, |
| 15 8. já ja bon te kontantəl,
ke ti si moja sama ščil'» | 30 |

6. B 1. Linčica Turkinčica / je vzela Ogra, / 2. je bila kraljeva hči / in on je bil kralj Oger, / (5) 3. ki je bil jetnik, / in za tri dni je bil zaprt / 4. in Linčica se je zaljubila vanj, / Ona je prosila [očeta] dobroto / [da bi nosila 3 dni jetniku jest v ječo]. / 5. In on je privolil. / (10) Ona je dol pokleknila / 6. in noge mu poljubila / in brž poljubila še njega. / 7. »Oj Linčica Turkinčica, / vse, kar me boš prosila, / (15) 8. jaz te bom zadovoljil, / ker ti si moja edina hči!« / 9. Dwa dni mu [jetniku] je nosila [jest] / in ona mu je rekla: / 10. »Ti si mi prav močno všeč / (20) in ti glej, da se oženiš z mano, / 11. ker jaz te bom rešila od tod, / jaz te bom resnično rešila od tod!« / 12. »Kako se boš omožila z mano, / saj jaz sem že oženjen! / (25) 13. a) Ampak jaz imam še dva brata doma, / lepša od mene.« / b) Imam dva lepša brata kot sem jaz, / ta dva te bosta vzela!« / In dva dni mu je nosila jest, / in ključje je iskala / (30) od velikih duri v Franciji, / ki gredo skozi, / skozi velik

Pretvorba v nevtrarno obliko

- Za strojno analizo potrebujemo **nevtrarno obliko**
- **Pravila** za zamenjavo
 - simbole, ki so karakteristični za narečne skupine (polglasniki, naglasi itn.) so zamenjani s slovničnimi ekvivalenti
- Ustvarili smo narečni **slovar** za preslikavo besed v slovnični jezik
 - vse, česar nismo našli v SSKJ
 - >18000 besed in besednih oblik
- **Oblikoslovni označevalnik** za slovenski jezik Obeliks za lematizacijo
 - besedam pripiše dodatne informacije, npr. besedno vrsto, v katero in kakšne so njene lastnosti
 - dobimo osnovno obliko - lemo

3. Nəč predowga, nəč prekratka,
sej ne bom plesala u nji.



Nəč predowga, nəč prekratka,
sej ne bom plesala u nji.



Slovar

...	
nəč	nič
...	
predowga	predolga
...	
rigelci	zapahi
...	

Nič predolga, nič prekratka,
saj ne bom plesala v nji



nič predolg nič prekratek
saj ne biti plesati v n

Eksperimenti

- Ožji izbor pesmi
 - pesmi o ljubezenskih usodah in konflikatih (SLP IV)
 - družinske pripovedne pesmi (SLP V)
- Teme povezane s smrtjo, umorom, samomorom, nezvestobo itn.
 - Smrt neveste pred poroko
 - Samomor nune zaradi ljubezeni
 - Nezvesti študent – novomašnik
 - Tašči se izjalovi umor snahe
 - ...



Obdelava korpusa

- Uporabljene metode dokumente (pesmi) obravnavajo kot vreče besed
 - vrstni red ni pomemben, le število pojavitev besed
- Vse pesmi postavimo v matriko besede-dokumenti
 - besede ovrednotimo z mero *tf-idf*
 - *term frequency-inverse document frequency*
 - poudari besede, ki se pogosto pojavijo v pesmi, vendar ne pogosto v celem korpusu

1

Kata, Katalena
v senčini sedela,
v senčini sedela,
drogen šiv šivala.



kata katalena v senca sedeti
v senca sedeti
droben šiv šivati
...

2.

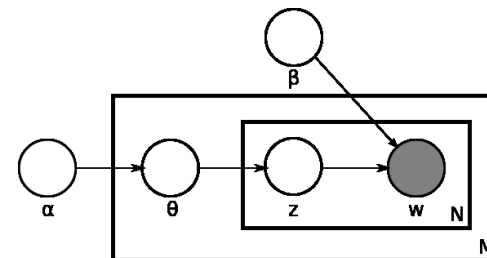
K njoj sta priletela
siva dva goloba.
V klunekaj prinesla
sveto Rešnje telo.



	p1	p2	p3	p4	p5	p6
kata	2	0	0	0	4	0
katalena	2	0	0	0	2	0
v	3	1	0	0	0	2
senca	2	1	0	0	0	3
sedeti	2	2	3	1	2	1
droben	1	0	1	0	0	4
šiv	1	0	1	0	3	1
šivati	2	0	1	0	0	0
...						

Tematska analiza

- Ali lahko z osnovnimi tehniki obdelave naravnega jezika najdemo **smiselne teme** v besedilih
 - se teme povezujejo z **variantnimi tipi**?
- LDA - *Latent Dirichlet Allocation*
 - verjetnostni model
 - dokument (pesem) je mešanica **tem**
 - verjetnosti **besed** v dokumentu izhajajo tem
- Ugotovitve
 - teme lahko povežemo z različnimi variantnimi tipi



LDA variant types and topics

DEATH AT A REUNION

t1: heart boy Breda head sad hunter Danube

MURDER OUT OF JEALOUSY

t2: love sword kneel sharp neighbor boyfriend blame

BRIDE INFANTICIDE

t3: home shepherd Mary uncle birth shred rockcradle

UNFAITHFUL STUDENT/NEW PRIEST

t4: undertaker love priest parish love promise letter

NUN'S SUICIDE FOR LOVE

t5: love Urška convent boy Jesus farewell sword

REJECTED LOVER

t6: seduce blood house Vida linden Ljubljani's death

WIDOWER ON BRIDE'S GRAVE

t7: tender abandon blood bread jesus rockcradle married

ABANDONED ORPHANS

t8: bury window chamber wound grow crying dead

PUNISHMENT FOR THE WICKED SONS AND DAUGHTERS-IN-LAW

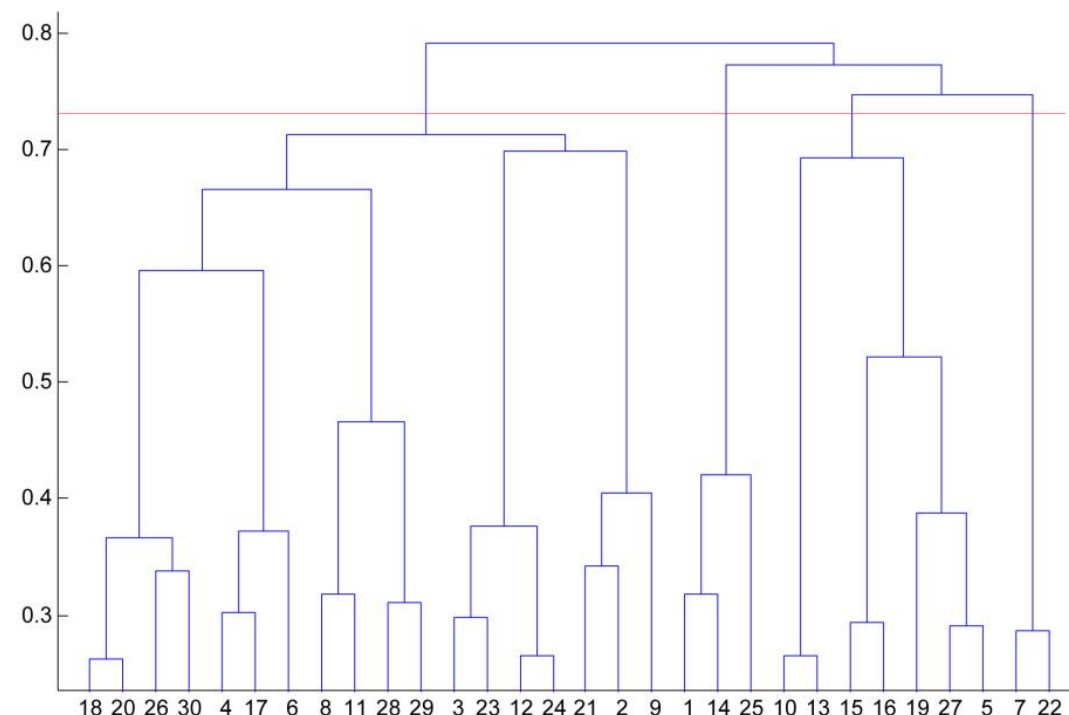
t9: gold sea mountain rooster fear crying darling son

MISTRESS' LOYALTY REPAID

t10: boy fenced heart nosegay dead grieve loyal

Gručenje pesmi

- Ali lahko **razvrstimo** pesmi v ljubezenske in družinske na podlagi tematske analize
 - težavnost – intertekstualnost, teme obeh skupin so podobne
- **Hiearhično gručenje** za razdelitev variantnih tipov glede na
 - podobnost povprečne tematske porazdelitve po temi
- Rezultat
 - prostor je moč razdeliti na obe skupini
 - tehnika omogoča, da uvrstimo pesem (tudi novo ali neznano) v prostor pesmi in analiziramo razmerja do ostalih



family clusters 1 (2:6) and 4 (13:31)

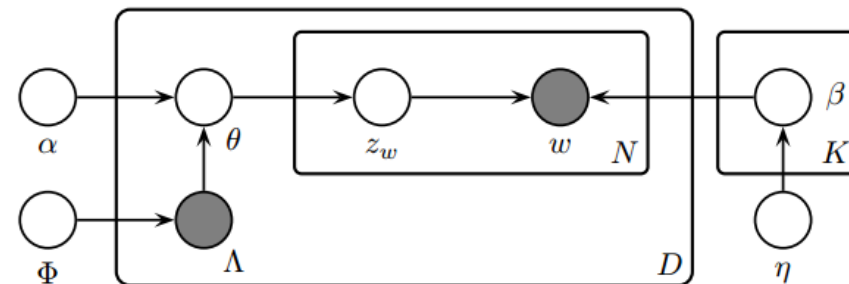
- hunter earth unfortunately rifle **son** mother remember
- noble castle **son** stand cry dress letter dress give
- **mother wife children** find gold adultery measure colorful stick boy
- mountain will water **mother** hero angry dam girlfriend mother-in-law
- **brother father** house dear ours sister see

love clusters 2 (17:11) and 3 (6:4)

- field three maid sun golden like ark sea **lover**
- things **husband** voice eat say young white know sin school
- **mistress** unlock boy saint window pot die lie
- stepmother run home getup graveyard rough get out go home

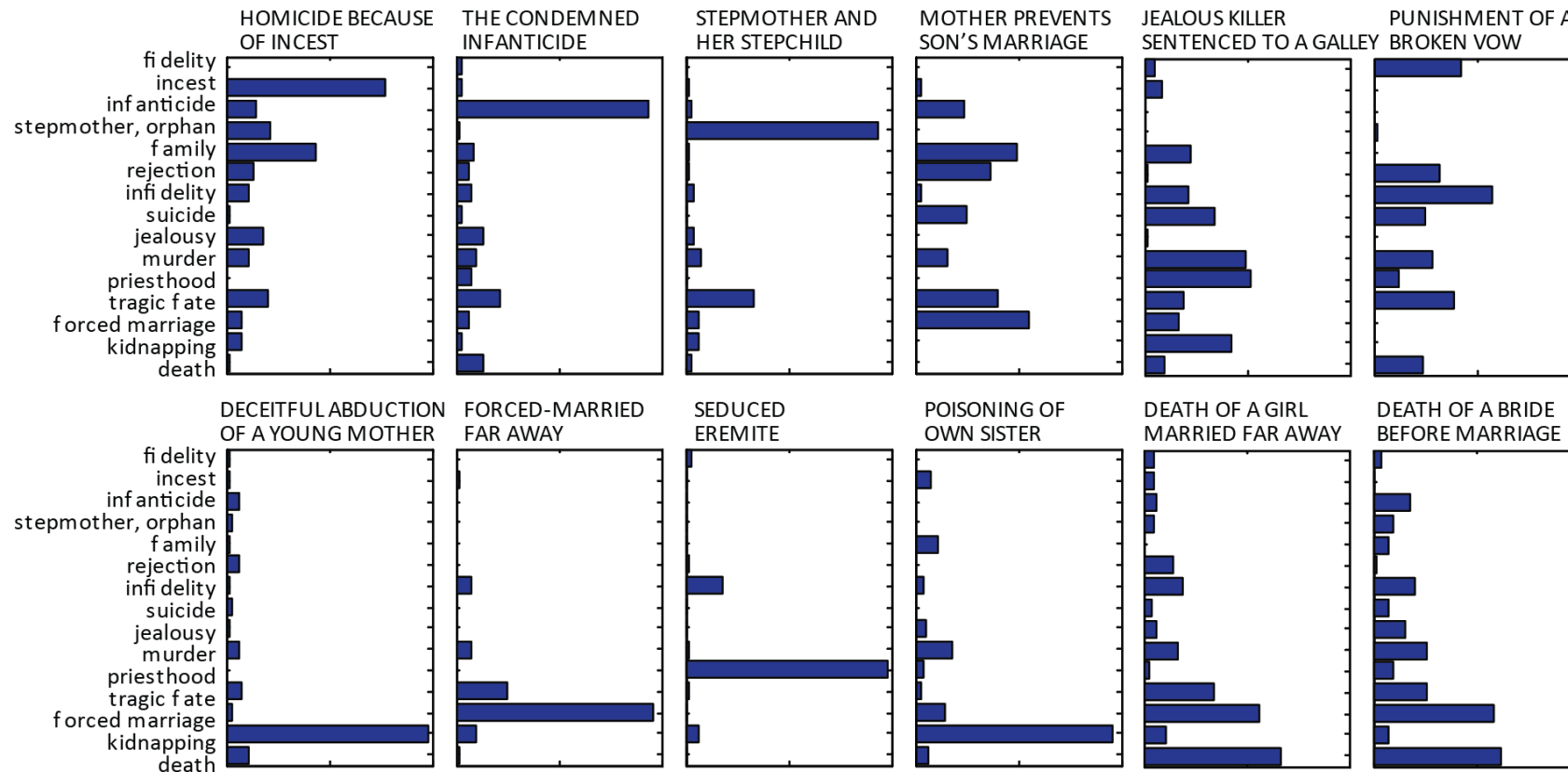
Napovedovanje tem

- Lahko z LDA ugotovimo karakteristične teme za posamezne pesmi oz. variantne tipe
- Nadzorovano učenje: *Labeled LDA*
 - **sami določimo teme** za nekatere pesmi
 - 18% korpusa
 - LLDA se nauči porazdelitve za te teme
- Naučene porazdelitve lahko posplošimo na cel korpus
 - dobimo porazdelitve po temah za vse pesmi



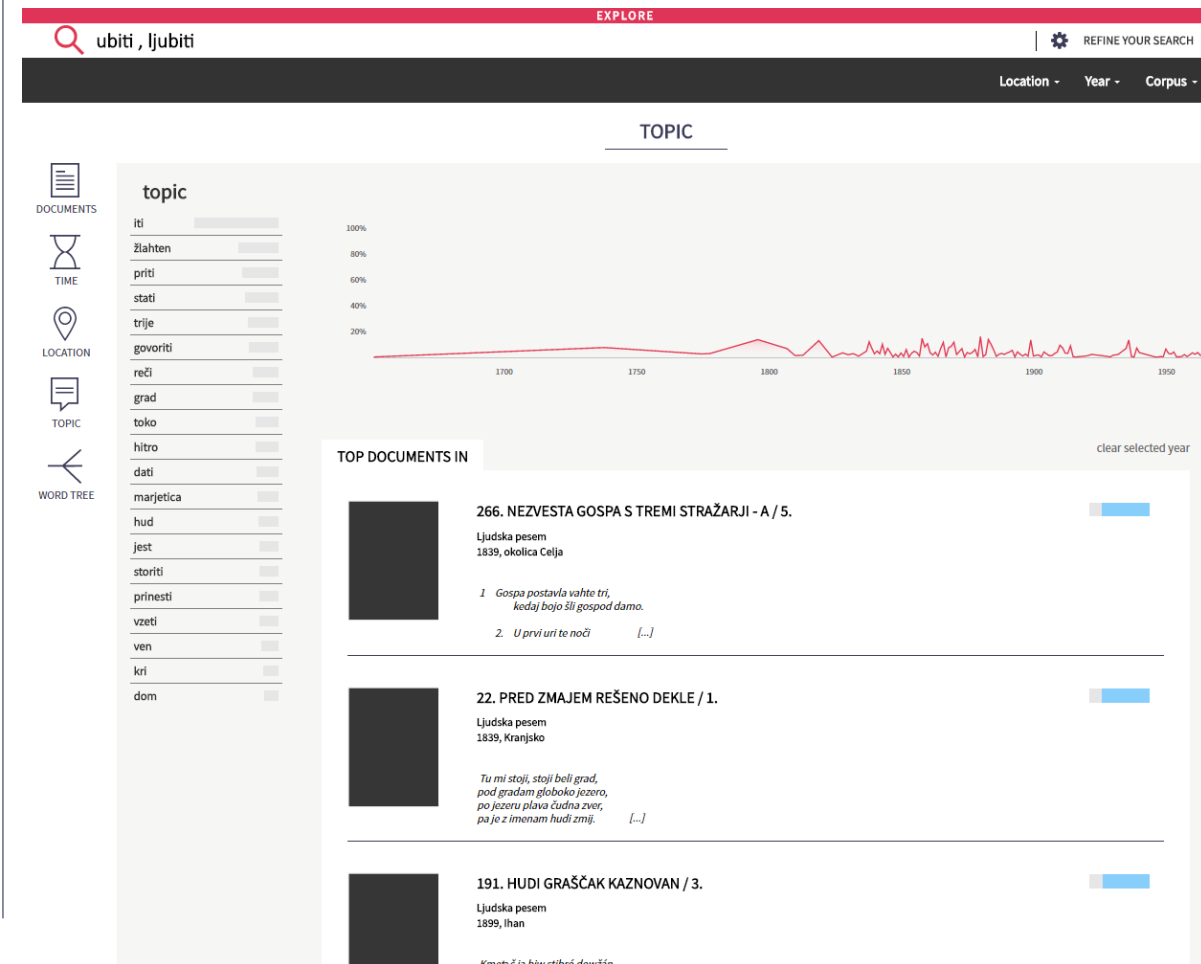
Napovedovanje tem

- Veliko pesmi ima več tem, glavna tema je večinoma najbolj prisotna
- Lahko razločimo med podobnimi temami (npr. nesrečna ljubezen)



Stranski projekt - TextExplore

- Omogočiti enostavno izdelavo in vpogled v tematske analize poljubnih korpusov
 - uvoz podatkov
 - izdelava tematskih modelov (Mallet)
 - [vizualizacija](#) dokumentov, tem, časovnega razvoja, lokacij, [besednih dreves](#)



Hvala