Creation and Analysis of the Yugoslav Rock Song Lyrics Corpus from 1967 to 2003[1]

UDC 811.163.41'322
DOI 10.18485/infotheca.2019.19.1.1

Ljudmila Petković
ljudmila.petkovic@gmail.com
University of Belgrade
Belgrade, Serbia

ABSTRACT: The paper analyses the process of creation and processing of the Yugoslav rock song lyrics corpus from 1967 to 2003, from the theoretical and practical perspective. The data have been obtained and XML-annotated using the Python programming language and the libraries lyricsmaster/yattag. The corpus has been preprocessed and basic statistical data have been generated by the XSL transformation. The diacritic restoration has been carried out in the Slovo Majstor and LeXimir tools (the latter application has also been used for generating the frequency analysis). The extraction of sociocultural topics has been performed using the Unitex software, whereas the prevailing topics have been visualised with the TreeCloud software.

KEYWORDS: corpus linguistics, Yugoslav rock and roll, web scraping, natural language processing, text mining.

PAPER SUBMITTED: 15 April 2019
PAPER ACCEPTED: 19 June 2019

[1] This paper originates from the author's Master's thesis "Creation and Analysis of the Yugoslav Rock Song Lyrics Corpus from 1945 to 2003", which was defended at the University of Belgrade on March 18, 2019. The thesis was conducted under the supervision of the Prof. Dr Ranka Stanković, who contributed to the topic's formulation, with the remark that the year of 1945 was replaced by the year of 1967 in this paper.

1 Introduction

1.1 The Phenomenon of the Yugoslav rock and roll

Music analysts, sociologists and anthropologists unanimously agree that the Yugoslav rock and roll (also known as the *Yu rock*) had left a deep impact on the territory of the former Yugoslavia. It is a musical style that was conceived in 1961, along with the appearance of the bands Uragani, Bijele Strijele and Siluete, while during the same decade the groups Crveni Koralji (1962), Zlatni Dečaci (1962) and Korni Grupa (1968) were formed (Janjatović, 1998). The Yu rock had been developing in parallel with the then-flourishing British *beat* scene (Раковић, 2011), made famous by the bands The Beatles and The Rolling Stones (Cooper and Cooper, 1993). In the late fifties and early sixties, the rock and roll in Yugoslavia had been equated with the rock and roll/twist dance, and not with the music genre per se (Раковић, 2011).

In 1963, the rock and roll in the Socialist Federal Republic of Yugoslavia (abbr. SFRY) had acquired the status of genre, which since then had been primarily performed with electric guitars, bass and drums (Раковић, 2011). Classic rock and rockabilly had represented very popular musical forms among the Yugoslavs, as evidenced by the covers of Elvis Presley, Buddy Holly, Chuck Berry and the like. (Арсенијевић et al., 2016). The 1970s had brought the influence of the hippie movement, with additional genre layering and the emergence of hard rock, progressive rock, art rock, and similar subgenres. The end of the 1970s and the early 1980s had been marked by the advent of punk rock and new wave, based on the equivalent seminal forms originating from the USA and the UK (Арсенијевић et al., 2016). In fact, the "Yugo-rock" has found its place in a society that had enthusiastically embraced the products of Western culture, both via domestic journals (e.g. *Rock* and *Džuboks*) and foreign radio programmes broadcast by Radio Luxembourg, military and pirate stations, as well as in form of films from the West. Also, Western fashion trends has also been followed, which had dictated, among other things, the men's wearing of long hair or the girls' wearing of mini skirts (Раковић, 2018).

Nevertheless, what made Yu-rock particularly distinctive were the libertarian lyrics, characteristic of punk and new wave music. Namely, the lyrics content was often politically engaged, ironic or vulgar, and therefore treated as unfit for the then-Yugoslav social circumstances (Гајић, 2018). The systematisation of the (self-)censorship cases in Yu-rock represents evidence that this was not an isolated phenomenon.[2] On the other hand, numerous songs from that era extolled the homeland, while in the others there were allusions to the partisans and the brigade (Божиловић, 2016). Božilovic adds that the core of Yu-rock was constituted by rock musicians' rebellion and fight for democracy, which had been encapsulated into unconventional, simple, direct and improvised musical forms and lyrics. However, Yu-rock, like its Western predecessor, also gave birth to the artists who sublimed their personal life attitudes into lyrics with lyrical, philosophical and introspective tone (e.g. the bands Azra, Idoli or Ekatarina Velika).

1.2 State of the Art

To the best of the author's knowledge, the studies of Yugoslav rock and roll have until now been primarily based on theoretical considerations, without application of computer technologies (some of such works are listed in the subsection 1.1). Zörnig et al. (2016) have conducted the only research that dealt with the quantitative analysis of song lyrics from the Yugoslav rock and roll era. Part of the corpus in the aforementioned work contains lyrics of the band Riblja Čorba and his frontman Bora Đorđević. The same paper presents the method of calculating the word frequencies and lexical variability using the *relative repeat rate* and *h-point* measures in the first place, which allowed the classification of corpus texts with multivariate analysis.

On the other hand, automatic collecting of song lyrics from the web, their electronic processing and analysis are obtaining more and more attention, as evidenced by the vast number of projects and scientific studies. There is a large number of international song lyrics corpora, e.g. The Million Song Dataset[3] (Bertin-Mahieux et al., 2011), which can be used in the computational text analysis researches. So far, the lists of the most frequent words in the annotated corpora of pop (Kreyer and Mukherjee, 2007) and rock song lyrics (Falk, 2013) have been generated. Topic modeling techniques were applied to the librettos of the Beijing opera (Zhang et al., 2017), as well as to the corpus of song lyrics harvested from the website SongMeanings (Lukic, s.d.).

When it comes to song lyrics from the domain of rock or related genres, the computational analysis of their content represents a promising and highly developed scientific practice. Some interesting results that these techniques

[2] See the website Balkanrock.
[3] The Million Song Dataset (on-line).

have produced are the structure extraction of The Beatles' song lyrics (Mahedero et al., 2005) or plotting the song lyrics written by Paul McCartney and John Lennon on the *emotion clock* circumplex (Whissell, 1996). The research carried out by Petrie et al. (2008) is another example of interest in computational processing of The Beatles' song lyrics. The given research is concerned with the change of the prevailing mood in their lyrics, while the use of stylometric analysis reveals stylistic similarities and differences between Lennon, McCartney, and George Harrison, as songwriters.

Falk (2013) has brought the results of a diachronic analysis of linguistic peculiarities in the corpus of international rock song lyrics from 1950 to 1999, based on the most frequent words used in each decade. Haslam (2017) has traced the evolution of thematic motives in the songs of the singer-songwriter Leonard Cohen, which has been visualised in the form of *word clouds*. From the standpoint of corpus linguistics, Taina et al. (2014) has explored the linguistic features in the song lyrics that distinguish heavy metal subgenres from each other. The tools used in psychometric research in order to calculate the textual cohesion (Coh-Metrix)[4] and to discover emotional, cognitive, and structural components in texts (Linguistic Inquiry and Word Count)[5] have been implemented in the comparative analysis of rock, folk, country, punk and grunge song lyrics (Lightman et al., 2007). The aim of the mentioned research was to identify differences in the writing style between artists who committed suicide and those who were not suicidal. Thematically more remote, but methodologically close to the present research is the paper dealing with the techniques for the automatic extraction of multi-word expressions in the lyrics of one ancient religious Hindu poem using local grammars in Unitex[6] software (Stein, 2012).

In all likelihood, the analysis of Yugoslav rock song lyrics from the aspect of computational linguistics seems to be an underdeveloped field of research. For that reason, this paper seeks to examine Yu-rock songs through the prism of interdisciplinarity, in order to reinforce the existing interpretations of the genre with the results produced by computational corpus processing tools.

1.3 Theoretical-Methodological Framework

One of the focal points of the current research is the application of corpus linguistics techniques. Corpus linguistics is a developed scientific methodol-

[4] Coh-Metrix (on-line).

[5] Linguistic Inquiry and Word Count (abbr. LIWC) (on-line).

[6] Unitex/GramLab (on-line).

ogy, while many scholars also consider it a discipline, theory, paradigm and a tool (Taylor, 2008). This methodological approach deals with the handling of structured, machine-readable and purposely chosen texts, which represent the basis for analysing various aspects of language and its usage. In the general context of corpus linguistics, the "purposely chosen texts" mean the collection of certain textual units sampled for the purpose of achieving representativeness in corpus construction. Frequency of a particular word or phrase usage, their retrieval as keywords in context by generating a concordance, or metadata extraction (e.g. name of the author of the text, date of publication, language in which the text is written, etc.) from an annotated corpus are just some of the possible tasks of corpus linguistics (McEnery and Hardie, 2012).

The data for the Yu-rock song lyrics corpus have been collected by using the web scraping[7] method. Algorithms for automatic collection, preprocessing and annotating the corpus in accordance with the XML syntax have been implemented in the Python programming language using the `lyricsmaster`, `xml.sax.utils` and `yattag` libraries. The XSLT transformation of the XML document into XHTML format allows the overview of the corpus statistics.[8] Automatic diacritic restoration has been conducted using the Slovo Majstor[9] and LeXimir[10] applications. Corpus linguistics and natural language processing methods have been applied with the purpose of frequency analysis of tokens and collocations using LeXimir, while the finite-state automation has been constructed in the Unitex software for the extraction of socio-political and culturological topics. The TreeCloud tool (Gambette and Véronis, 2009) has been used for visualising dominant topics in the corpus, within the text mining framework (serb. *kopanje po tekstu*, according

[7] The above English term is standardised in foreign literature, unlike in the articles and studies published in this area, in which numerous attempts of unique terminological determination are made (e.g. *nalaženje podataka na vebu* ("finding web data"), *struganje podataka* ("data scraping") or *grebanje veba* (web scraping), to name just a few). The aforementioned terms originate from the web articles "Nalaženje podataka na internetu" and "SEO optimizacija kroz rečnik najbitnijih termina".

[8] Codes are available at the author's GitHub repository (on-line).

[9] Slovo Majstor (on-line).

[10] LeXimir (on-line).

to Кешељ and Шипка (2008) or *iskopavanje iz teksta*[11]), which is based on discovering text patterns in unstructured data.

The paper consists of five sections. Basic features of Yu-rock, some relevant works in the field of computational processing of domestic and international song lyrics, as well as the research methods used in this paper were considered in the introductory part. The section 2 describes the process of automatic collection of material for the Yu-rock song lyrics corpus. The semi-automatic[12] preprocessing and corpus annotation methods were presented in the section 3. The experimental results of the application of the corpus are shown in the section 4, while the section 5 gives concluding remarks and guidelines for future work.

2 Data Collection via Web Scraping Methods

The narrower field of research proposed in this paper is concerned with the computational creation and analysis of the corpus of Yugoslav rock song lyrics during the era of two Yugoslav states – the Socialist Federal Republic of Yugoslavia (1945-1992) and the Federal Republic of Yugoslavia (1992-2003). As for the criteria for data collection, the corpus is composed of lyrics written in the former Serbo-Croatian language, which had been originally standardised by the Vienna Literary Agreement in 1850, but eventually split into Serbian, Croatian and Bosnian language, with the disintegration of Yugoslavia in 1991 (Hentschel, 2003). Accordingly, the lyrics in the mentioned three languages have been sought, whereas the lyrics written in other languages that had been in use in Yugoslavia – Macedonian, Slovene and minority languages – were excluded.

Regarding the web scraping method, it refers to the direct downloading of digitalised content with the aid of a specific computational tool. The core of this practice can be described as the automatisation of the comprehensive and relatively fast extraction and storage of web data. Web scraping is imposed as a far more efficient solution in comparison with the manual methods of copying and collecting each information unit individually. Additionally, the technique in question can represent the only possible solution for data extraction in case the "copy/paste" options are disabled in web browser.

[11] The latter term (lit. "excavation from the text") is taken from the presentation of the Prof. Dr. Cvetana Krstev (on-line).

[12] The semi-automatic method involves a combination of techniques for automatic and manual preprocessing and corpus annotations.

2.1 Description of the Data Source: the LyricWiki Website

LyricWiki[13] is a commercial music website which hosts lyrics of domestic and international artists. The website is searchable by name of artist, album, song, genre, and record label. Various lists are also available, such as the lists of the most popular albums from a certain year, according to the opinion of the editors of prominent music magazines, and lists of albums of film soundtracks, as well as plenty of other textual content related to the music field. As stated in the description of this website, LyricWiki is publicly available and is licensed to publish reliable lyrics.

The website's database stores more than 2,054,289 texts (data from June 15, 2019).[14] Since LyricWiki contains a relatively large number of songs in Serbian and former Serbo-Croatian language, at the initial stage of the research (while the website was still open for editing), the lyrics of the songs performed by selected musicians from the former Yugoslavia have been collected.[15] The additional reason for choosing this site for the scraping of textual data was the fact that lyrics could be extracted from it via the specific API, i.e. the `LyricWiki` module from the `lyricsmaster` library using the Python language, which will be discussed in detail in the next section.

2.2 Functionality of the lyricsmaster Library in the Python Language

Automatic collection of lyrics from the Internet represents an efficient and relatively commonly leveraged method that precedes the electronic corpus analysis. One of the representative examples of such a practice is the use of the `lyricsmaster` programming library, which is available on the PyPI's repository website[16] containing Python programming packages. Using this API, the lyrics stored in the databases of some of the most popular music websites (hereinafter referred to as *providers*), such as AZLyrics, Genius, etc., can be collected. On the same webpage, the usage of the aforementioned tool was demonstrated with the purpose of directly downloading and saving the lyrics of the late American rapper Tupac Shakur, which are available on LyricWiki. For the anonymisation of the user's IP address, there is an option for scraping the website via the Tor Proxy Server[17].

[13] LyricWiki (on-line).

[14] Statistics (on-line).

[15] The list of artists starts from the webpage Language/Serbian.

[16] lyricsmaster 2.8.1 (on-line).

[17] Tor Project (on-line).

The implementation of the aforementioned library has resulted in automatic extraction of textual content from various LyricWiki webpages, each of which was dedicated to some of the popular Yugoslav performers whose lyrics have been collected. Specifically, the web scraping function has been defined, and the `get_lyrics()` method of the class `lyricsmaster.providers.LyricWiki(tor_controller = None)`[18] has been used in order to retrieve the desired pages, from which only the lyrics were collected, and not the other content visible on the same webpages (e.g. header or sidebar information). An algorithm had been constructed which has collected the lyrics content by artist name in the following manner:

1. The LyricWiki website was firstly selected as the lyrics provider;
2. A list of thirty artists, whose lyrics were to be collected, was defined with the variable `izvodjaci` ("performers"). The artists' names were referenced exactly as they had been listed on the website (e.g. instead of `'Yu-Grupa'`, `'yu grupa'`, etc., entering the string value `'YU Grupa'` was only allowed). This also applied to cases in which two artists shared the same name (e.g. the Serbian and Finnish group Negative). The LyricWiki editors have made a difference between the mentioned groups by assigning the appropriate "RS" tag of the country of origin to the Serbian band, in order to explicitly refer to the group from the Republic of Serbia. With that in mind, the modified label `Negative (RS)`[19] was entered in the web scraping code;
3. The `corpus()` function was created whose arguments were the elements of the aforementioned list of artists, and for each of them it was attempted to collect the lyrics with the method `get_lyrics()`;
4. The `discography`, `album` and `song` objects were then created, while the introduction of the parameters `title` and `lyrics` resulted in accessing the album titles (`album.title`), song titles (`song.title`) and lyrics content (`song.lyrics`);
5. Afterwards, the collected corpus material was stored in the local computer memory using the `save()` method, which was applied to the `discography`, `album` i `song` objects.[20] The default absolute path was `/{user}/Documents/LyricsMaster/`, and the lyrics content were saved in the `/artist/album/` directory in the `song.txt` format.

[18] `get_lyrics` is the main method of the specified class (from the documentation on-line).

[19] The Finnish group did not receive any label, but only the name "Negative".

[20] `save()` is the method of three classes: `lyricsmaster.models.Discography`, `lyricsmaster.models.Album` and `lyricsmaster.models.Song` .

On the LyricWiki webpages with the lyrics of certain artists (e.g. YU Grupa), most of the song titles had been formatted as hyperlinks that pointed to the webpages with available lyrics. However, for several song titles the lyrics content had not been created at all (e.g. for the song „Čovek i Marsovac", i.e."A Man and a Martian"), despite the formal existence of the song titles, which was causing the algorithm to stop prematurely. In order to handle the exceptions, the `try` statement was introduced for the error detection in the place(s) where the problem(s) had occured. The `except` and `continue` statements, which eliminate the errors and allow the program to continue the started procedure, were also added. The mentioned errors are summarized in the Table 1.

2.3 Limitations of the lyricsmaster Library

It was not possible to collect lyrics with the `lyricsmaster` library for certain bands and solo artists (Riblja Čorba, Ekatarina Velika, Azra, Film, Gibonni, Đorđe Balašević, etc.), either due to the restricted access to data or the lack of lyrics by a specific artist on LyricWiki. Furthermore, at the stage of collecting the lyrics it was concluded that the corpus should contain female artists' lyrics, but for certain female artists whose musical style can be characterized as "rock" (Kaliopi, Slađana Milošević, etc.) their lyrics could not be automatically extracted from the given website. Likewise, some artists did not even exist in the database, and so as their lyrics (as was the case with Maja Odžaklijevska).

Song/Artist	**Error**	**Cause**	**Solution**
"Čovek i Marsovac"	`AttributeError`	Link unavailable	`try-except-continue`
`'Yu-Grupa'`	`TypeError`	Incorrect name	`'YU Grupa'`

Table 1: Problematic cases during the collection of lyrics.

In order to compensate for this deficiency, the alternative criteria have been applied: they refer to the subsequent selection of artists by expanding the genre's scope that will be covered by the corpus material. More precisely, the pop artists (such as Nina Badrić and Zana), whose songs have been influenced by rock music, were also included in the corpus. Other added artists were Madame Piano[21] and Goran Karan[22]. The corpus proposed in this paper consists of songs of the artists listed in the Table 2:

Bajaga	Električni Orgazam	Neverne Bebe
Bajaga i Instruktori	Zabranjeno Pušenje	Negative
Bebi Dol	Zana	Nina Badrić
Bijelo Dugme	Idoli	Oktobar 1864
Van Gogh	Indexi	Partibrejkers
Galija	Josipa Lisac	Prljavo Kazalište
Goran Bregović	YU Grupa	Rani Mraz
Goran Karan	Kerber	Smak
Divlje Jagode	Madame Piano	Hari Mata Hari
Dino Merlin	Mirzino Jato	Haustor

Table 2: List of 30 artists in the corpus.

The oldest album in the corpus is *Naše doba* (1967) by the band Indexi; in contrast, the most recently published albums which are included in the corpus are the Divlje Jagode's album *Od neba do neba*, and *Collection* by Nina Badrić (both published in 2003).

2.4 Creating a Directory Tree for Storing the Corpus

By executing the web scraping code, a hierarchically organised data structure was created, that is, a directory tree with its root and branches. The resulting tree structure of the corpus can be represented in form of an output from the command line of an operating system, after running the code in the programming language Bash:[23]

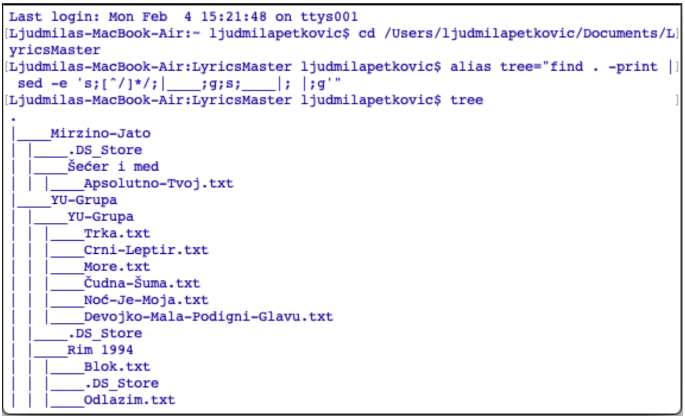
[21] The Yu-rock expert, Petar Janjatović, included Madame Piano in his YU rock encyclopedia (Janjatović, 1998), owing to the impact that the aforementioned representative of jazz and world music sound had on the development of the Yugoslav music scene.

[22] Goran Karan is also not a representative example of a rock artist, but does not deviate too much from the target frame. Although Karan gained popularity by performing pop songs "with a Dalmatian tone", he had began his career as a rock singer (the information obtained from his biography on-line).

[23] Based on the Bash code provided on-line.

```
alias tree="find . -print | sed -e ’s;[^/]*/;|____;g;s;____|; |;g’"
```

The Figure 1 depicts a partial view of the structure in question that was taken from the Terminal application included with the macOS operating system, after the initial navigation to the directory `LyricsMaster`, where the lyrics corpus was located.



```
Last login: Mon Feb  4 15:21:48 on ttys001
[Ljudmilas-MacBook-Air:~ ljudmilapetkovic$ cd /Users/ljudmilapetkovic/Documents/L]
yricsMaster
[Ljudmilas-MacBook-Air:LyricsMaster ljudmilapetkovic$ alias tree="find . -print |]
 sed -e 's;[^/]*/;|____;g;s;____|; |;g'"
[Ljudmilas-MacBook-Air:LyricsMaster ljudmilapetkovic$ tree                        ]
.
|____Mirzino-Jato
| |____.DS_Store
| |____Šećer i med
| | |____Apsolutno-Tvoj.txt
|____YU-Grupa
| |____YU-Grupa
| | |____Trka.txt
| | |____Crni-Leptir.txt
| | |____More.txt
| | |____Čudna-Šuma.txt
| | |____Noć-Je-Moja.txt
| | |____Devojko-Mala-Podigni-Glavu.txt
| |____.DS_Store
| |____Rim 1994
| | |____Blok.txt
| | |____.DS_Store
| | |____Odlazim.txt
```

Figure 1: Directory tree “LyricsMaster” from the command line.

As can be concluded from the Figure 1, the root directory carries the default name “LyricsMaster”, within which the directories with the artists’ names (e.g. Mirzino Jato or YU Grupa) were located. Then, for each artist, the album titles (*Šećer i med, YU Grupa, Rim 1994*, etc.) were listed. The albums contain lyrics in the .txt format which represent the terminal nodes in the tree structure (e.g. “Apsolutno tvoj”, “Crni leptir”, etc.). Among the mentioned data, the .DS_Store file also appeared, which was not of any significance for further corpus processing and analysis, and which was therefore eliminated at a later phase. After the web scraping procedure, the collected lyrics content in the .txt format, stored in separate directories, in the next stage of the corpus preparation was unified into an unique XML file and annotated.

3 Corpus Preprocessing

3.1 DTD Specification

Prior to the automatic generation of the corpus in the XML format, the *document type declaration* (abbr. DTD) had been created, which specified the logical structure of the XML document to which the document refers, and in relation to which its validity was checked. This step was also important because it represented a prerequisite for the further processing of the XML document, such as generating basic statistics from it in a clearly presented XHTML format using the XSLT processor, which will be discussed in the subsection 4.1. In the valid XML corpus document, for example, the DTD declares that the root element `<exYuPesme>` can have multiple authors, but that its attribute value is a fixed empty string.

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="tabele-css-classes.xsl"?>
<!DOCTYPE exYuPesme [
<!ELEMENT exYuPesme (autor)+>
<!ATTLIST exYuPesme
xmlns CDATA #FIXED ''> ... ]
```

Figure 1: Document Type Declaration – excerpt.

3.2 Locating the Lyrics Using the os Module

The first step in the process of generating the XML document was the implementation of the `os` module using the `listdir()` method, from The Python Standard Library. When the method is called from the root "LyricsMaster" directory, it returns the directory list, i.e. the artists' names. In order to collect the names of the subdirectories (album titles of the given artist) and files (song titles on those albums), it was necessary to perform a string formatting in order to create the absolute paths using `format()` function. The final parts of the path, that is, a string of everything that comes after the last slash in the path argument (*base name*) represents the values of the target categories. In this process, two functions of the `os.path` module were used: with the `join()` function, the elements of the list of directories/files were concatenated to the ends of the current directory paths in order to access these elements. The second function, `isfile()`, was combined with the flow control statements `if` and `if not` in order to correctly determine whether a list of directories or files was generated. For illustration purposes, below are listed the paths based on which it was possible to extract the name of the band Smak, their album title *Crna Dama* and the song title from the same album – "Daire":

```
../../LyricsMaster/Smak
../../LyricsMaster/Smak/Crna-Dama
../../LyricsMaster/Smak/Crna-Dama/Daire.txt
```

Then the lyrics content was loaded using the `open()` and `read()` functions, whereas the song structure of lyrics containing newlines was retained using the `split('\n')[:-1]` function. The purpose of the described procedure was to correctly allocate the data when annotating the file in the XML language.

3.3 Elimination of the Redundant Content

Certain lyrics from the corpus had been written exclusively in foreign languages. Since the subject of this paper was the processing of the corpus of lyrics in Serbian or Serbo-Croatian language, the manual removal of the mentioned songs or even the whole albums from the corpus directories was performed. In the present subsection, some representative examples of corpus preprocessing were listed. In particular, the lyrics written in Greek, Macedonian, Romanian, English, Portuguese and Polish language were excluded from the analysis. A specific case was the presence of multilingualism in the lyrics of Goran Bregović as a solo artist, who was also the author of the songs "Κέρνα μας", from the album *Alkohol: šljivovica & champagne*; "7/8 & 11/8", "Ederlezi", "TV screen", "Ausência" (album *Ederlezi*) and "To nie ptak" (*Kayah & Bregović*). On the album soundtrack from the film *Arizona Dream* all the songs were in English, so that album was also removed.

The lyrics which were attributed to an author by mistake were also excluded from the corpus. For example, among the lyrics on the Nina Badrić's LyricWiki webpage there was also the song "Ubila si del mene", which belongs to the Slovene boy band Game Over. Two songs ("Muistoja" and "Myrsky") by the homonymous band from Finland had also been incorrectly assigned to the Serbian group Smak from Kragujevac. Besides, on the album *Zašto ne volim sneg* by the band Smak there were several instrumental songs. The songs for which the lyrics content on LyricWiki had been missing (instead of the lyrics there were only the suspension points), as in the case of the song "Ne mogu da kapiram", by the group Partibrejkers, were also not taken into account. Moreover, there were the cases whereby the duplicates of lyrics appeared under different song titles. For example, the lyrics of the song

"Počasna salva" by the band Zabranjeno Pušenje were retained under that song title, while the duplicates with the false titles "Manijak", "Vuk" and "Ujka Sam" were deleted.

Particular attention was paid to selecting the songs from concert or compilation albums. Namely, it had been initially intended to immediately remove such directories, as it was expected that they would have contained the songs that had already existed in the corpus. However, the presence of certain number of unpublished songs on the concert albums was noticed (e.g. "Na vrhovima prstiju" from the album *Neka svemir čuje nemir* by the band Bajaga i Instruktori). From the greatest hits album *Collection* by Nina Badrić few unique songs were kept. Another type of duplicate was also found in form of some song covers, such as for the song "Tako ti je mala moja kad ljubi Bosanac", originally published by Bijelo Dugme in 1975 and covered by the group Zabranjeno Pušenje in 1998.

3.4 Annotating and Preprocessing the XML Corpus

`Yattag`[24] represents the Python programming library that can automatically insert HTML and XML tags while structuring the document. This API automatically inserts open and closed angle brackets, and each start-tag is followed by a end-tag. In this manner, the annotator can tag texts more quickly and easily, because the program reduces the possibility of reporting syntax errors. Using the module `indent`, the `Yattag` library also supports automatic indentation according to the general hierarchical structure of an XML document, and the size of the indented space can be modified. Since the annotation procedure was to be carried out on a large corpus, the objective was to exploit the functionality of the present library in order to annotate the document more efficiently, in accordance with the defined rules for marking up the lyrics content. the content of the texts. These rules concerned the inclusion of the XML elements' and attributes' tags for describing the corresponding parts of the corpus content using the following procedure:

- The root element was defined by the tag `<exYuPesme>`;
- The authors, i.e. performers, were defined by the element `<autor>`, whose attributes are `ime`, `brojAlbuma`[25], `pol`, `zanr`, `rodnoMesto` (songwriters were not included because they had not either been listed on the website);

[24] Yattag (on-line).

[25] That is, the number of albums included by the lyrics that form part of the corpus.

- The albums were defined by the `<album>` element, which has the attributes `naziv`, `godina` and `izdavac`;
- The songs were defined by the `<pesma>` element which has the `naslovPesme` attribute, while the tag of the `<li>` element was reserved for the verse lines.

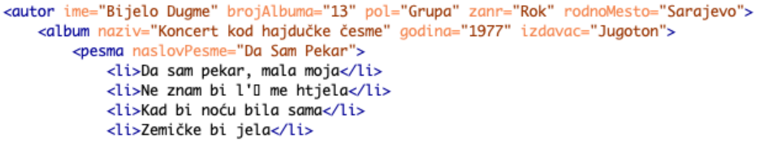
Since the method `tag()` creates XML tags, only the desired tag names for marking up the elements and their attributes were forwarded as the method's arguments. For example, the album as an element has the attributes for the title, publication year and the record label, which was defined as follows: `with tag('album', naziv=album, godina="", izdavac=""):`. On the other hand, the method `text()` generated the text content that was not the name of the tag. For brevity of code, the `Doc` instance of the class `yattag.Doc` and the joint method `tagtext()`, which adds the content produced by the specified method to this instance (names of elements, attributes, and the plain text), were used. After defining all the necessary parameters, the `getvalue()` method was applied, in order to convert the whole content into a long character string. This way, the lyrics were transferred into a new XML file.

After the XML corpus had been experimentally generated, the special ampersand character (`&`) was noticed, which, as a rule, was replaced by the escape sequence character (`&amp;`). This was performed in order not to interpret the ampersand as the beginning of the entity reference and to fully comply with the principles of proper XML structuring. For that reason, the `escape()` function of the module `xml.sax.saxutils` was inserted into the code for annotating the corpus. Although the apostrophe symbol (`'`) often appears in the corpus, it was not replaced by its XML equivalent `&apos;` in the process of automatic character substitution, which did not hinder the process of generating a well-formed XML document. For the sake of greater transparency, the hyphens present in the titles of the songs and albums that had remained during the web scraping procedure, were automatically removed (e.g. the initial song title "Da-Sam-Pekar" was changed to "Da Sam Pekar").

As for the manual annotation of the corpus, the values "pop", "rock" and "world music"[26] of the attribute `zanr` were appended to the element `<autor>`. Besides the genre, the values of the attribute `rodnoMesto` which refer to the artist's birthplace and `izdacac` for the record label that published the album, were also added. An example of the semi-automatically annotated lyrics from

[26] According to the genre classification available on the website Discogs.

the the album *Koncert kod Hajdučke česme* by Bijelo Dugme, can be seen in the Figure 2.



```
<autor ime="Bijelo Dugme" brojAlbuma="13" pol="Grupa" zanr="Rok" rodnoMesto="Sarajevo">
    <album naziv="Koncert kod hajdučke česme" godina="1977" izdavac="Jugoton">
        <pesma naslovPesme="Da Sam Pekar">
            <li>Da sam pekar, mala moja</li>
            <li>Ne znam bi l'▯ me htjela</li>
            <li>Kad bi noću bila sama</li>
            <li>Zemičke bi jela</li>
```

Figure 2: Excerpt from the annotated corpus in the oXygen XML Editor software.

3.5 Automatic diacritic restoration – LeXimir software

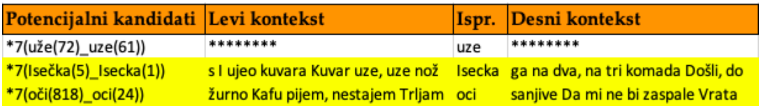
The corpus normalisation and its preparation for the computational analysis does not represent a trivial task whatsoever, and, as a rule of thumb, it contributes to generating more informative results in comparison with the computational analysis of the non-preprocessed text. The collected lyrics in the Yu-corpus were rather uneven in terms of the use of the writing systems: most of the lyrics were written in Latin script, and there were quite a few lyrics in which the special Latin letters (č, ć, ž, đ, š) did not contain the necessary diacritical marks. Simultaneously, fewer songs were originally written in Cyrillic. The initial idea had been that the whole corpus be translated into Cyrillic; this idea was later rejected because the lyrics in the Serbian language also contained some excerpts in foreign languages (e.g. "la musique c'est fantastiqu / prepare la revolution / et la femme est tres jolie / tre jolie comme un bonbon").[27] It was therefore decided that during the transliteration process the lyrics written in the degraded Latin and Cyrillic alphabet be automatically converted into the standardised Serbian Latin alphabet containing diacritical marks. The transliterated corpus lyrics in the .txt format were used for extracting lexical units (see the subsection 4.2).

In order to select the most suitable tool, the diacritic restoration was performed in the Slovo Majstor and LeXimir applications. The second software uses electronic morphological dictionaries (with valid word forms), local grammars and the Corpus of Contemporary Serbian Language

[27] According to the original lyrics, with typographical and spelling errors.

(Krstev et al., 2018). After evaluating both methods, it was concluded that Slovo Majstor solidly solved the problem of converting the degraded Latin alphabet into the standardised one, with the remark that the application incorrectly added diacritical marks to some words, which violated the word semantics. Thus, from the verb form “isecka” the word “isečka” was incorrectly created. The favourable aspect of this application was the ability to subsequently correct the incorrectly transliterated words.

Conversely, the LeXimir tool, which exploits the functionalities of Unitex corpus processing software, left the previously mentioned word in the correct form. In the Figure 3, the potential candidates for selecting an adequate form were the word “isecka”, which is an inflected form of the noun “isečak” in the genitive singular case, and “isecka”, that is, third-person singular of the aorist of the verb “iseckati”. The diacritic restoration algorithm based on the left and right textual context of the potential word for the lexical correction performed the *disambiguation* (ambiguity resolution). The second option was selected, where the pronoun “ga” can co-occur with the verb “isecka”, and not with the inflected form of the noun “isečka”, since the other construction would be grammatically incorrect. Nevertheless, manual evaluation is to be performed after the implementation of the described tool as well, because in some words (e.g. in the words “oci” in the context “Trljam oci sanjive”), no diacritical marks were added where they had been supposed to exist.



Potencijalni kandidati	Levi kontekst	Ispr.	Desni kontekst
*7(uže(72)_uze(61))	********	uze	********
*7(Isečka(5)_Isecka(1))	s I ujeo kuvara Kuvar uze, uze nož	Isecka	ga na dva, na tri komada Došli, do
*7(oči(818)_oci(24))	žurno Kafu pijem, nestajem Trljam	oci	sanjive Da mi ne bi zaspale Vrata

Figure 3: Diacritic restoration in the LeXimir tool.

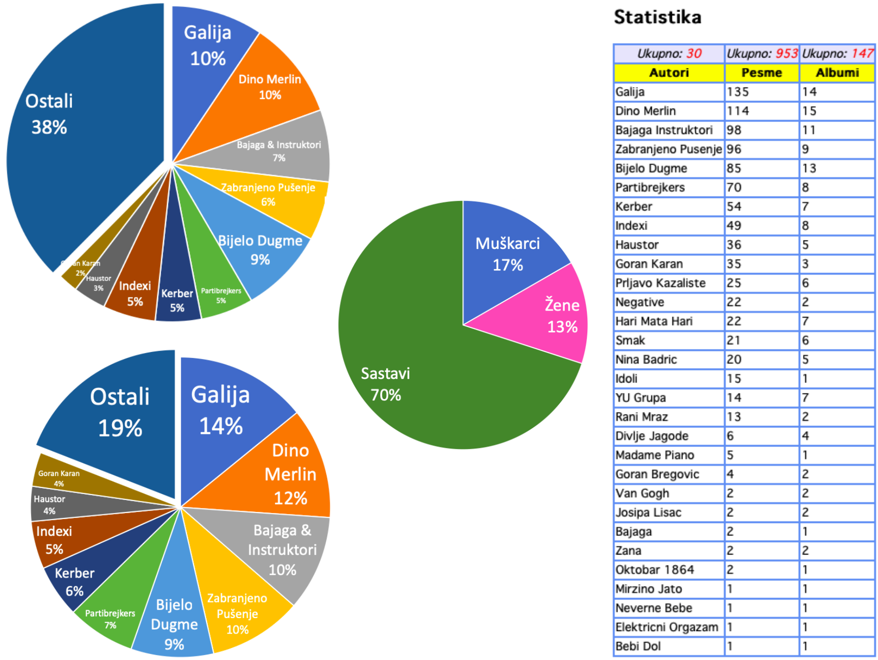
4 Computational Corpus Analysis

4.1 Corpus Statistics

The DTD specification enabled submitting queries by navigating over the given document using the XSLT language for transforming an XML document into other formats. For instance, the following XSLT expression selects the value of the attribute `ime` for each element `<autor>` in the document, which generates a list of all artists in the corpus:

```
<xsl:for-each select="//autor">
        <xsl:value-of select="@ime" />
</xsl:for-each>
```

Data from the XML document were transformed into the XHTML table from the Figure 4, where it can be observed the total number of artists, songs and albums in the corpus. The pie charts from the same figure illustrate (clockwise, top-down): the artists' percentage in the corpus by the number of albums, the percentage of male, female and group artists, as well as the percentage by the number of songs ("ostali" means "others").

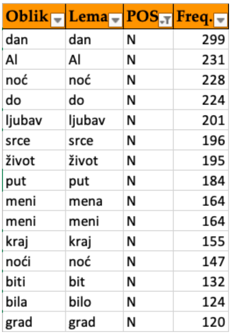


Statistika

Ukupno: 30	*Ukupno: 953*	*Ukupno: 147*
Autori	**Pesme**	**Albumi**
Galija	135	14
Dino Merlin	114	15
Bajaga Instruktori	98	11
Zabranjeno Pusenje	96	9
Bijelo Dugme	85	13
Partibrejkers	70	8
Kerber	54	7
Indexi	49	8
Haustor	36	5
Goran Karan	35	3
Prljavo Kazaliste	25	6
Negative	22	2
Hari Mata Hari	22	7
Smak	21	6
Nina Badric	20	5
Idoli	15	1
YU Grupa	14	7
Rani Mraz	13	2
Divlje Jagode	6	4
Madame Piano	5	1
Goran Bregovic	4	2
Van Gogh	2	2
Josipa Lisac	2	2
Bajaga	2	1
Zana	2	2
Oktobar 1864	2	1
Mirzino Jato	1	1
Neverne Bebe	1	1
Elektricni Orgazam	1	1
Bebi Dol	1	1

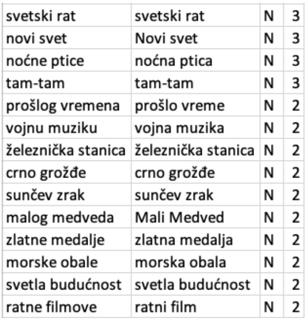
Figure 4: Yu-corpus statistics.

LeXimir also produced basic statistics of the Yu-corpus. After the diacritic restoration, the corpus contains 248,807 tokens, 16,964 unique lexical units, 116,972 simple forms (16,909 different) and 268 numbers (10 different). LeXimir also provides support for displaying inflected forms of tokens, lemmatised forms, parts of speech, words' and collocations' frequencies. The results of the frequency analysis can be filtered by the part of speech. Based on the exported .xlsx file it can be determined, for instance, what are the most frequent nouns or collocations whose headword is a noun. The Figure 5 below shows the partial display of the results which reveal the most frequent noun tokens, among which the words *dan, noć, ljubav, srce*, etc., appear. The Figure 6 provides an overview of certain collocations related to war (*svetski rat, vojnu muziku, ratne filmove*).



Oblik	Lema	POS	Freq.
dan	dan	N	299
Al	Al	N	231
noć	noć	N	228
do	do	N	224
ljubav	ljubav	N	201
srce	srce	N	196
život	život	N	195
put	put	N	184
meni	mena	N	164
meni	meni	N	164
kraj	kraj	N	155
noći	noć	N	147
biti	bit	N	132
bila	bilo	N	124
grad	grad	N	120

Figure 5: Tokens.



svetski rat	svetski rat	N	3
novi svet	Novi svet	N	3
noćne ptice	noćna ptica	N	3
tam-tam	tam-tam	N	3
prošlog vremena	prošlo vreme	N	2
vojnu muziku	vojna muzika	N	2
železnička stanica	železnička stanica	N	2
crno grožđe	crno grožđe	N	2
sunčev zrak	sunčev zrak	N	2
malog medveda	Mali Medved	N	2
zlatne medalje	zlatna medalja	N	2
morske obale	morska obala	N	2
svetla budućnost	svetla budućnost	N	2
ratne filmove	ratni film	N	2

Figure 6: Collocations.

4.2 Extracting topics from the corpus

It is known that the socio-political and culturological topics were present in the Yugo-rock lyrics, thus the Unitex software was used for constructing the graph of finite-state automation for recognising the indicators of the above topics. Regular expressions and lexical masks served for the retrieval of lexical forms in the form of abbreviations[28] composed of:

- capital letters, hyphens and a sequence of numbers (e.g. *B-52*);
- sequence of capital letters + full point repeating at least two times (*K.P.*);

[28] See the webpage Skraćenice i objašnjenja.

- sequence of capital letters + space repeating at least two times (*ES EF ER JOT*);
- sequence ranging from two to five capital letters (*CZ, CIA*).

Among them the names of state authorities (e.g. *AVNOJ*), sports clubs (*PFC*), the name of broadcasting companies (*BBC*), etc. came to the fore. The graph from the Figure 7, in addition to the abbreviated names presented in the concordance result in the Figure 8, also recognised the tokens: *CZ, ES EF ER JOT, FK, TV, JRT, K.P., KGB , KK, MUP, O.K., PC, PFC, SUP* and *TAS*.

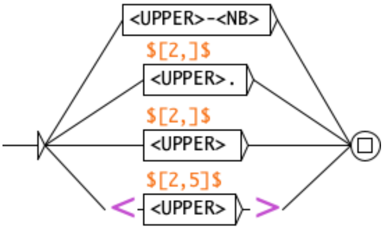
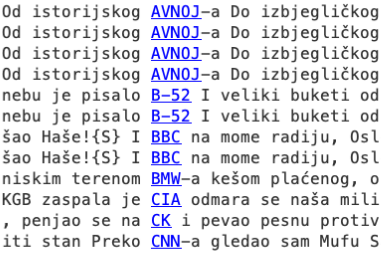


Figure 7: Graph that recognises the abbreviations.



```
Od istorijskog AVNOJ-a Do izbjegličkog
Od istorijskog AVNOJ-a Do izbjegličkog
Od istorijskog AVNOJ-a Do izbjegličkog
Od istorijskog AVNOJ-a Do izbjegličkog
nebu je pisalo B-52 I veliki buketi od
nebu je pisalo B-52 I veliki buketi od
šao Haše!{S} I BBC na mome radiju, Osl
šao Haše!{S} I BBC na mome radiju, Osl
niskim terenom BMW-a kešom plaćenog, o
KGB zaspala je CIA odmara se naša mili
, penjao se na CK i pevao pesnu protiv
iti stan Preko CNN-a gledao sam Mufu S
```

Figure 8: Concordance excerpt.

In order to determine whether the socially engaged topics were statistically significant in the corpus, a visual overview of predominant topics in the form of a *tree cloud* had been generated using the TreeCloud tool, with the remark that the analysis had been performed on non-lemmatised forms. This program combines the word cloud graphical representation of data with a tree structure, so that, apart from the word frequencies, it is also possible to observe the way of clustering the tokens based on their distance in this structure (see the Figure 9). The application also includes a built-in stop words list for Serbian language, which was supplemented with some words for the sake of better visualisation of the given corpus. A visual representation was created, in which one can distinguish the clusters indicating feelings, expressed by the words *tuga, ljubav/-i, bol, srce/-a*. Urban topics were represented by the words *grad/-u, ulice*, body parts were associated with the words *ruke, lice, oči, srce/-a/-u*. The motive of time is also frequent in the corpus (*život, vr(ij)eme, godina/-e*). According to this visualisation, socially engaged topics are not sufficiently present in the corpus.

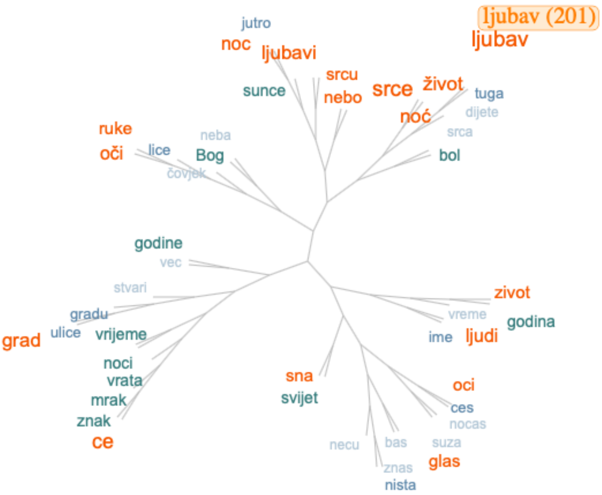


Figure 9: Tree cloud of the whole corpus.

5 Concluding Remarks

The paper discussed the project of electronic creation and processing of the Yugoslav rock song lyrics from 1967 to 2003. Automatic data collection from the LyricWiki website was performed using the `lyricsmaster` library in the Python programming language, and the preprocessed corpus was automatically annotated in compliance with the XML syntax rules, using the `yattag` tool, along with the manual adding of some attribute values. Automatic diacritic restoration was also carried out. XSL transformation of the corpus into XHTML format was also shown, as well as the extraction of socio-political and culturological topics in the Unitex software and the visualisation of the prevailing topics with the TreeCloud tool.

Further work would include the corpus evaluation with the aim of correcting the spelling and typographical errors (e.g. *Dunav* instead of *dunav*, *Avale* instead of *Aavale* etc.). Also, the implementation of electronic morphological dictionaries of named entities is planned, in order to extract the names of musicians, politicians, athletes and other celebrities who had exerted a significant impact on the Yugoslav society.

References

Bertin-Mahieux, Thierry, Daniel PW Ellis, Brian Whitman and Paul Lamere. "The million song dataset". In *Proceedings of the 12th International Conference on Music Information Retrieval*, 591–596. 2011. Accessed: 22/10/2019, http://ismir2011.ismir.net/papers/OS6-1.pdf.

Cooper, Laura E. and B. L. Cooper. "The pendulum of cultural imperialism: Popular music interchanges between the United States and Britain, 1943-1967". *Journal of Popular Culture* Vol. 27, no. 3 (1993): 61–78. Accessed: 22/10/2019, https://sci-hub.tw/10.1111/j.0022-3840.1993.00061.x#.

Falk, Johanna. "We will rock you : A diachronic corpus-based analysis of linguistic features in rock lyrics". Bachelor's thesis, Linnaeus University, Department of Languages, 2013. Accessed: 22/10/2019, http://www.diva-portal.org/smash/get/diva2:605003/FULLTEXT02.pdf.

Gambette, Philippe and Jean Véronis. "Visualising a Text with a Tree Cloud", *IFCS'09: International Federation of Classification Societies Conference* (2009): 561–569. Accessed: 22/10/2019, https://hal-lirmm.ccsd.cnrs.fr/lirmm-00373643/file/2009GambetteVeronis.pdf

Haslam, Thomas J. "Mapping the Great Divide in the Lyrics of Leonard Cohen". *Rupkatha Journal on Interdisciplinary Studies in Humanities* Vol. 9, no. 1 (2017): 1–10, Accessed: 22/10/2019, http://rupkatha.com/V9/n1/v9n1s01.pdf

Hentschel, Elke. "The expression of gender in Serbian". In *Gender across languages: The linguistic representation of women and men*, Vol. 3, John Benjamins Publishing Company, 287–309, 2003. Accessed: 22/10/2019, https://epdf.tips/gender-across-languages-volume-iii-the-linguistic-representation-of-women-and-me.html

Janjatović, Petar. *Ilustrovana YU rock enciklopedija: 1960-1997*, Geopoetika, 1998. Accessed: 22/10/2019, https://monoskop.org/images/c/ca/Janjatovic_Petar_Ilustrovana_YU_Rock_Enciklopedija_1960-1997.pdf

Kreyer, Rolf and Joybrato Mukherjee. "The style of pop song lyrics: A corpus-linguistic pilot study". *Anglia-Zeitschrift für englische Philologie* Vol. 125, no. 1 (2007): 31–58, Accessed: 22/10/2019, https://doi.org/10.1515/ANGL.2007.31

Krstev, Cvetana, Ranka Stanković and Duško Vitas. "Knowledge and Rule-Based Diacritic Restoration in Serbian". In *Proceedings of the Third International Conference Computational Linguistics in Bulgaria* (2018): 41–51, Accessed: 22/10/2019, https:

//www.researchgate.net/publication/328416358_Knowledge_and_Rule-Based_Diacritic_Restoration_in_Serbian

Lightman, Erin J., Philip M. McCarthy, David F. Dufty and Danielle S. McNamara. "Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal songwriters". In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 29, 2007. Accessed: 22/10/2019, https://cloudfront.escholarship.org/dist/prd/content/qt0dh4553j/qt0dh4553j.pdf.

Lukic, Alen. "A Comparison of Topic Modeling Approaches for a Comprehensive Corpus of Song Lyrics", Tech report, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, s.d. Accessed: 22/10/2019, http://alenlukic.com/assets/docs/lyric_topic_modeling.pdf

Mahedero, Jose P.G. Álvaro Martínez, Pedro Cano, Markus Koppenberger and Fabien Gouyon. "Natural language processing of lyrics". In *Proceedings of the 13th annual ACM international conference on Multimedia*, 475–478. 2005. Accessed: 22/10/2019, https://www.researchgate.net/profile/Pedro_Cano5/publication/221573745_Natural_language_processing_of_lyrics/links/00b7d52826f623edfb000000.pdf

McEnery, Tony and Andrew Hardie. *Corpus Linguistics: Method, Theory and Practice.* Cambridge University Press, 2012. Accessed: 22/10/2019, http://gen.lib.rus.ec/book/index.php?md5=33c1c5b6d73ea816dfb2a034f73bb176

Petrie, Keith J., James W. Pennebaker and Borge Sivertsen. "Things We Said Today: A Linguistic Analysis of the Beatles". *Psychology of Aesthetics, Creativity, and the Arts* Vol. 2, no. 4 (2008): 197. Accessed: 22/10/2019, https://www.uvm.edu/pdodds/files/papers/others/2008/petrie2008a.pdf.

Stein, Daniel. "Multi-Word Expressions in the Spanish Bhagavad Gita, Extracted with Local Grammars Based on Semantic Classes". In *LREC '2012 Workshop: LRE-Rel, Language Resources and Evaluation for Religious Texts*, 88–93. 2012. Accessed: 22/10/2019, https://www.academia.edu/26035335/Linguistic_and_Semantic_Annotation_in_Religious_Memento_mori_Literature

Taina, Jesse et al.. "Keywords in heavy metal lyrics: A Data-Driven Corpus Study into the Lyrics of Five Heavy Metal Subgenres", 2014. Accessed: 22/10/2019, https://helda.helsinki.fi/bitstream/handle/10138/136524/keywords.pdf?sequence=1.

Taylor, Charlotte. "What is corpus linguistics? What the data says". *ICAME journal* Vol. 32 (2008): 179–200. Accessed: 22/10/2019, http://sro.sussex.ac.uk/id/eprint/53389/1/what_is_corpus_linguistics.pdf

Whissell, Cynthia. "Traditional and Emotional Stylometric Analysis of the Songs of Beatles Paul McCartney and John Lennon". *Computers and the Humanities* Vol. 30, no. 3 (1996): 257–265. Accessed: 22/10/2019, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.7171&rep=rep1&type=pdf.

Zhang, Shuo, Rafael Caro Repetto and Xavier Serra. "Understanding the expressive functions of jingju metrical patterns through lyrics text mining". In *18th International Society for Music Information Retrieval Conference*, 397–403. 2017. Accessed: 22/10/2019, https://repositori.upf.edu/bitstream/handle/10230/32652/Zhang_ISMIR2017_unde.pdf?sequence=1&isAllowed=y

Zörnig, Peter, Emmerich Kelih and Ladislav Fuks. "Classification of Serbian texts based on lexical characteristics and multivariate statistical analysis". *Glottotheory* Vol. 7, no. 1 (2016): 41–66. Accessed: 22/10/2019, http://homepage.univie.ac.at/emmerich.kelih/wp-content/uploads/2016_Zoernig_Kelih_Fuks_www.pdf.

Арсенијевић, Александра, Милена Обрадовић and Михаило Шкорић. "Израда мултимедијалног документа „YU рок сцена"". *INFOtheca – Journal for Digital Humanities* Vol. 16, no. 1-2a (2016): 113-129. Accessed: 22/10/2019, https://infoteka.bg.ac.rs/ojs/index.php/Infoteka/article/view/2016.16.1_2.6_sr

Божиловић, Никола. "Култура сећања и југословенски рокенрол". *Култура* no. 152 (2016): 257–280. Accessed: 22/10/2019, https://scindeks-clanci.ceon.rs/data/pdf/0023-5164/2016/0023-51641652257B.pdf

Гајић, Златомир. "Рок поезија Бранимира Штулића и њени медијски одјеци". PhD thesis, Faculty of Philosophy, University of Novi Sad, 2018. Accessed: 22/10/2019, http://nardus.mpn.gov.rs/bitstream/handle/123456789/9926/Disertacija17602.pdf?sequence=1&isAllowed=y

Кешељ, Владо and Данко Шипка. "Приступ изградњи стемера и лематизатора за језике с богатом флексијом и оскудним ресурсима заснован на обухватању суфикса". *INFOtheca – Journal for Digital Humanities* Vol. 9, no. 1-2 (2008): 21–31. Accessed: 22/10/2019, http://infoteka.bg.ac.rs/pdf/Srp/2008/04%20Vlado-Danko_Stemeri.pdf

Раковић, Александар. "Бит мода, рокенрол и генерацијски сукоб у Југославији 1965-1967". *Етноантрополошки проблеми* Vol. 6, no. 3 (2011): 745–762. Accessed: 22/10/2019, https://www.eap-iea.org/index.php/eap/article/view/604/594

Раковић, Александар. "Рокенрол у Социјалистичкој Југославији: од забаве градске омладине до националне културе". In *Сан о граду : зборник радова*, 427–439. The Andrić Institute, 2018. Accessed: 22/10/2019, http://doi.fil.bg.ac.rs/pdf/eb_book/2018/ai_san_o_gradu/ai_san_o_gradu-2018-ch18.pdf