



# Садржај презентације

- Ресурси: једнојезични и вишејезични корпуси, електронски речници, термилошке базе.
- Апликације: паралелни текстови, лексичка база, терминологија, веб сервиси за проширење упита, за анотације.
- Скупови података за обучавање модела.
- Модели за анотацију: врсте речи, лематизацију, именованих ентитета
- Векторске репрезентације речи, ГПТ модел за српски
- Ресурси семантичког веба.

# Састав ЈеРТеха

- Друштво за језичке ресурсе и технологије ЈеРТех окупља истраживаче различитих профила:
  - математичаре,
  - информатичаре, <https://jerteh.rs/>
  - лингвисте,
  - лексикографе,
  - терминологе и
  - друге који су заинтересовани за развој и примену језичких технологија на научном, стручном и практичном нивоу.
- Председник: проф.др Душко Витас
- Отворено друштво за све заинтересоване сличних интересовања  
<https://jerteh.rs/index.php/pristupnica/>

# Ресурси

Међу ресурсима које Друштво развија и одржава се издвајају:

- једнојезични и вишејезични корпуси,
  - платформа No Sketch Engine <https://noske.jerteh.rs/>
  - ТХМ
- електронски речници,
  - Unitex, Leximirka
- термиолошки речници
  - Терми
- повезани лингвистички отворени подаци.

# Корпуси

- Корпуси омогућују брзо и лако добијање велике количине емпиријских података о језику
- Практична примена у лексикографији, настави језика, превођењу, обучавању векторских репрезентација речи, реченица,...
- Корпуси општег језика: СрпКор2013, СрпКор2021, ВикиКорпус, SrpELTeC...
- Доменски корпуси: рударски, геолошки, електроенергија, библиотекарство, математички, фудбалски, ...
- Двојезични: енглеско-српски, немачко-српски, италијанско-српски, француско-српски,...

Статистика изабраних корпуса са <https://noske.jerteh.rs/>

Корпуси	Токени	Речи	Документи	Реченице	Параграфи
BiKes_sr	1.848.802	1.547.103	336	82.459	
GeoSrpKor	1.316.646	1.067.583	69		
Matematika	1.932.271	1.341.242	89		
RudKorp	3.542.016	2.713.594	172		
SkolKor	4.736.884	3.650.849	82		
SrpELTeC	5.923.024	4.766.056	108	144.387	
SrpKor2013	145.275.324	121.142.927	5.038	5.178.311	
SrpKor2021	716.878.652	606.683.682	1.916	29.718.407	2.934.786
WikiKorpus	81.267.249	64.510.812	609.380	609.380	
ZlatniKorpus	1.333.804	1.104.864	8		
<b>Укупно</b>	<b>964.054.672</b>	<b>808.528.712</b>	<b>617.198</b>	<b>35.732.944</b>	<b>2.934.786</b>

## SrpKor2021

Регистар	Речи	Документи	%
административни	4.659.111	149	0,7
књижевност	47.561.951	541	7,1
наука	18.319.938	572	2,7
новине	510.701.630	336	75,7
разговорни	92.102.304	301	13,7
друго	1.141.607	17	0,2
<b>Укупно</b>	<b>674.486.541</b>	<b>1.916</b>	<b>100,0</b>

### Семинар Друштва Јертех

- 29.12.2022. „Приказ структуре и функција новог СрпКор-а (верзија 4.0)“ – Ранка Станковић – Рударско-геолошки факултет Универзитета у Београду [\[pdf\]](#)
- 29.12.2022. „О српским корпусима и корпусима српског језика, а посебно оним који су на вебу“ – Душко Витас – Друштво за језичке ресурсе и технологије [\[pdf\]](#)

## SrpKor2013 + SrpKor2021 + wiki + domenski+...

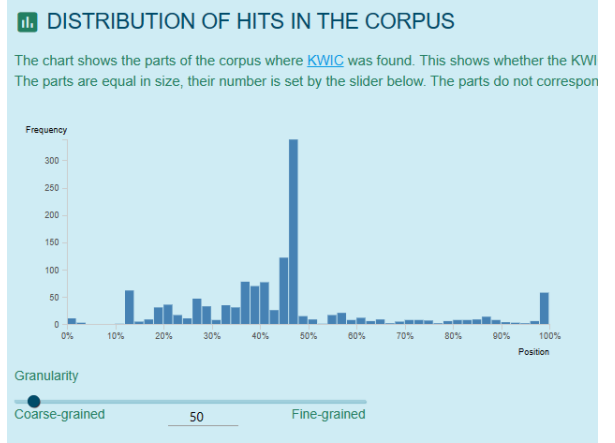
Регистар	Речи	Документи	%
административни	11.505.283	1.070	1,3
књижевност	66.785.237	1.000	7,6
наука	28.843.058	744	3,3
новине	600.892.578	3.581	68,4
разговорни	92.102.304	301	10,5
уџбеници	4.147.771	97	0,5
вики	64.510.812	609.380	7,3
друго	9.601.134	362	1,1
<b>Укупно</b>	<b>878.388.177</b>	<b>616.535</b>	<b>100</b>

# CONCORDANCE

SrpKor2021

CQL [lemma="veštački"][lemma="inteligencija"] • 1,307  
1.82 per million tokens • 0.00018%

	Details	Left context	KWIC	Right context
1	doc#1	itacija , tehnička podrška , tehničke mogućnosti , upravljачka struktura ,	veštačka inteligencija	, zakonska regulativa , zaštita spomenika kulture i dr. Celokupan popis i
2	doc#1	, računarska tehnologija ( 17 ) , učenje na daljinu ( 32 ) , web sajt ( 5 ) ,	veštačka inteligencija	( 4 ) itd. .</s></s>Kod termina koji su sačinjeni od imenice kojoj prethod
3	doc#1	1 sin vektorski model 4 1 N2X grf03 sin veličina slova 4 1 AXN grf01 sin	veštačka inteligencija	4 1 AXN grf01 sin visokoškolska biblioteka 95 1 AXN grf01 sin visokošk
4	doc#1	grf03 sin upravljanje znanjem 6 NXN grf06 sin web sajt 5 AXN grf01 sin	veštačka inteligencija	4 AXN grf01 sin visokoškolska ustanova 9 AXN grf01 sin visokoškolsko
5	doc#62	kovinama digitalizacije , otvoren je čitav spektar novih tema , kao što su	veštačka inteligencija	, teleprisustvo i telerobotika , baze podataka , internet aktivizam , video-
6	doc#62	e i obuhvatala teoriju informacija , opštu teoriju sistema , strukturalizam ,	veštačku inteligenciju	, kibernetiku i slično , mada u svojim radovima nisu koristili neposredno
7	doc#82	korišćenom u Srbiji .</s></s>Ekspertni sistemi ( ES ) su oblast primene	veštačke inteligencije	, a zasnivaju se na ideji da se sposobnost kopiranja , koja je rezultat sf
8	doc#82	EMI U ANALITIČKOJ HEMIJI Ekspertni sistemi ( ES ) su oblast primene	veštačke inteligencije	koja je počela da se razvija sredinom 60-tih godina XX veka ( Liao , 200
9	doc#82	, obrazovanje , poljoprivreda i dr. Osnova ES baziranih na saznanju je	veštačka inteligencija	i oni predstavljaju pokušaj da se razume i podstakne znanje ljudi u komj
10	doc#82	:S zasnovanog na ovom principu .</s></s>ES su glavna oblast primene	veštačke inteligencije	u hemiji ( Ramos et al. , 1986 ) .</s></s>Razvoj ES u oblasti hemije poč
11	doc#82	JMSTUNE ( Dessy , 1984 b ) i ESCA ( Buydens et al. , 1990 ) .</s></s>	veštačka inteligencija	i posebno ES , imaju sve veću ulogu u obezbeđivanju , ugrađene — inte
12	doc#173	od elitnih američkih univerziteta , gde sam proučavao primenu i teoriju	veštačke inteligencije	u projektovanju proizvoda i tehnologije .</s></s>Sedam godina sam radi
13	doc#186	ošlost , pa i kod nas , a ne novom informatičkom društvu koje je , slično	veštačkoj inteligenciji	i informatičkim tehnologijama , okrenuto uvek ka dograđivanju , povratn



## Упитни језик CQL (Corpus query language)

1	doc#1	tehničke mogućnosti , upravljачka struktura ,	veštačka inteligencija	zakonska regulativa , zaštita spomenika kulture i
2	doc#1	32 ) , web sajt ( 5 )	veštačka inteligencija	( 4 ) itd. .</s></s> Kod termina
3	doc#1	N2X grf03 sin veličina slova 4 1 AXN grf01 sin	veštačka inteligencija	4 1 AXN grf01 sin visokoškolska biblioteka 95
4	doc#1	em 6 NXN grf06 sin web sajt 5 AXN grf01 sin	veštačka inteligencija	4 AXN grf01 sin visokoškolska ustanova 9 AXN g
5	doc#62	je čitav spektar novih tema , kao što su	veštačka inteligencija	, teleprisustvo i telerobotika , baze podataka
6	doc#62	nacija , opštu teoriju sistema , strukturalizam	veštačku inteligenciju	, kibernetiku i slično , mada u svoj
7	doc#82	Ekspertni sistemi ( ES ) su oblast primene	veštačke inteligencije	, a zasnivaju se na ideji da se spos
8	doc#82	Ekspertni sistemi ( ES ) su oblast primene	veštačke inteligencije	koja je počela da se razvija sredinom 60-tih godi
9	doc#82	vreda i dr. Osnova ES baziranih na saznanju je	veštačka inteligencija	i oni predstavljaju pokušaj da se razume i
10	doc#82	ovom principu .</s></s>ES su glavna oblast primene	veštačke inteligencije	u hemiji ( Ramos et al. 1986 )
11	doc#82	( Buydens et al. 1990 ) .</s></s>	veštačka inteligencija	i posebno ES , imaju sve veću ulogu u
12	doc#173	verziteta , gde sam proučavao primenu i teoriju	veštačke inteligencije	u projektovanju proizvoda i tehnologije .</s></s> Se
13	doc#186	ne novom informatičkom društvu koje je , slično	veštačkoj inteligenciji	i informatičkim tehnologijama , okrenuto uvek ka
14	doc#195	, zasnovana je na elektronicima , robotici	veštačkoj inteligenciji	i da snaga tog preobražaja jeste u nauci



(16 items, 1,307 total frequency)

	Word	Frequency	Relative ?
1	<input type="checkbox"/> veštačke inteligencije	581	0.81
2	<input type="checkbox"/> veštačka inteligencija	220	0.31
3	<input type="checkbox"/> veštačku inteligenciju	196	0.27
4	<input type="checkbox"/> veštačkoj inteligenciji	106	0.15
5	<input type="checkbox"/> Veštačka inteligencija	102	0.14
6	<input type="checkbox"/> veštačkom inteligencijom	75	0.10
7	<input type="checkbox"/> Veštačkoj inteligenciji	7	< 0.01
8	<input type="checkbox"/> Veštačke inteligencije	5	< 0.01
9	<input type="checkbox"/> Veštačku inteligenciju	5	< 0.01
10	<input type="checkbox"/> VEŠTAČKE INTELIGENCIJE	4	< 0.01
11	<input type="checkbox"/> veštačkih inteligencija		
12	<input type="checkbox"/> veštačku inteligencije		
13	<input type="checkbox"/> veštačke inteligencija		
14	<input type="checkbox"/> VEŠTAČKA INTELIGENCIJA		
15	<input type="checkbox"/> veštačkim inteligencijama		
16	<input type="checkbox"/> veštačka inteligenciju		

Фреквенције: облика, лема, врста речи

Колокације и линкови ка изабраним конкорданцама

## Collocations

CHANGE CRITERIA

BACK TO CONCORDANCE

	Word	Cooccurrences ?	Candidates ?	T-score	MI	LogDice ↓
1	<input type="checkbox"/> AI	34	1,005	5.83	14.18	8.91 ...
2	<input type="checkbox"/> robotike	20	307	4.47	15.12	8.67 ...
3	<input type="checkbox"/> Ai	11	288	3.32	14.35	7.82 ...
4	<input type="checkbox"/> mašinskog	12	726	3.46	13.15	7.60 ...
5	<input type="checkbox"/> robotiku	8	105	2.83	15.35	7.54 ...
6	<input type="checkbox"/> mašinsko	8	292	2.83	13.88	7.36 ...
7	<input type="checkbox"/> Veštačka	7	218	2.65	14.10	7.23 ...
8	<input type="checkbox"/> robotika	6	94	2.45	15.10	7.13 ...
9	<input type="checkbox"/> ARTifical	5	5	2.24	19.07	6.96 ...
10	<input type="checkbox"/> roboti	8	934	2.83	12.20	6.87 ...

	Word	Cooccurrences ?	Candidates ?	T-score	MI	LogDice ↓
11	<input type="checkbox"/> robotizacije	4	33	2.00	16.02	6.61 ...
12	<input type="checkbox"/> big	6	726	2.45	12.15	6.60 ...
13	<input type="checkbox"/> generativnoj	4	60	2.00	15.16	6.58 ...
14	<input type="checkbox"/> robota	9	1,770	3.00	11.45	6.58 ...
15	<input type="checkbox"/> IoT	4	104	2.00	14.36	6.54 ...
16	<input type="checkbox"/> gramatici	4	548	2.00	11.97	6.14 ...
17	<input type="checkbox"/> dronovi	3	103	1.73	13.96	6.12 ...
18	<input type="checkbox"/> učenja	16	6,429	4.00	10.41	6.08 ...
19	<input type="checkbox"/> algoritme	3	149	1.73	13.43	6.08 ...
20	<input type="checkbox"/> mašinskom	3	201	1.73	13.00	6.03 ...

Rows per page: 20

1-20 of 377



1

/ 19



[tag="A"]+[lemma="inteligencija"]

## SrpKor2021

(1,203 items, 5,856 total frequency)

Lemma	Frequency	Relative ?
<input type="checkbox"/> veštački inteligencija	1,307	1.82
<input type="checkbox"/> pošten inteligencija	269	0.38
<input type="checkbox"/> srpski inteligencija	256	0.36
<input type="checkbox"/> emocionalan inteligencija	200	0.28
<input type="checkbox"/> misionarski inteligencija	120	0.17
<input type="checkbox"/> poslovan inteligencija	110	0.15
<input type="checkbox"/> kritički inteligencija	99	0.14
<input type="checkbox"/> tehnički inteligencija	92	0.13
<input type="checkbox"/> socijalan inteligencija	86	0.12
<input type="checkbox"/> ljudski inteligencija	82	0.11
<input type="checkbox"/> nacionalan inteligencija	79	0.11
<input type="checkbox"/> liberalan inteligencija	62	0.09
<input type="checkbox"/> visok inteligencija	59	0.08
<input type="checkbox"/> natprosečan inteligencija	57	0.08
<input type="checkbox"/> politički inteligencija	55	0.08
<input type="checkbox"/> koeficijent inteligencija	53	0.07
<input type="checkbox"/> ruski inteligencija	50	0.07
<input type="checkbox"/> prosečan inteligencija	44	0.06
<input type="checkbox"/> elementaran inteligencija	43	0.06
<input type="checkbox"/> kolektivan inteligencija	41	0.06

## SrpKor2013

(463 items, 1,149 total frequency)

Lemma	Frequency	Relative ?
<input type="checkbox"/> veštački inteligencija	94	0.65
<input type="checkbox"/> srpski inteligencija	58	0.40
<input type="checkbox"/> emocionalan inteligencija	32	0.22
<input type="checkbox"/> socijalan inteligencija	29	0.20
<input type="checkbox"/> pošten inteligencija	27	0.19
<input type="checkbox"/> stvaralački inteligencija	23	0.16
<input type="checkbox"/> ljudski inteligencija	21	0.14
<input type="checkbox"/> koeficijent inteligencija	21	0.14
<input type="checkbox"/> visok inteligencija	19	0.13
<input type="checkbox"/> test inteligencija	17	0.12
<input type="checkbox"/> velik inteligencija	15	0.10
<input type="checkbox"/> koeficijentom inteligencija	13	0.09
<input type="checkbox"/> francuski inteligencija	13	0.09
<input type="checkbox"/> kritički inteligencija	13	0.09
<input type="checkbox"/> evropski inteligencija	13	0.09
<input type="checkbox"/> ruski inteligencija	12	0.08
<input type="checkbox"/> srpski stvaralački inteligenc	12	0.08
<input type="checkbox"/> tehnički inteligencija	10	0.07
<input type="checkbox"/> natprosečan inteligencija	10	0.07
<input type="checkbox"/> građanski inteligencija	10	0.07

## SrpELTeC

(26 items, 33 total frequency)

Lemma	Frequency	Relative ?
<input type="checkbox"/> slab inteligencija	3	0.51
<input type="checkbox"/> viša inteligencija	3	0.51
<input type="checkbox"/> seoski inteligencija	2	0.34
<input type="checkbox"/> visok inteligencija	2	0.34
<input type="checkbox"/> srpski inteligencija	2	0.34
<input type="checkbox"/> narodni inteligencija	1	0.17
<input type="checkbox"/> prirodan inteligencija	1	0.17
<input type="checkbox"/> svet inteligencija	1	0.17
<input type="checkbox"/> energiju brajkovački inteligencija	1	0.17
<input type="checkbox"/> borila inteligencija	1	0.17
<input type="checkbox"/> svedok inteligencija	1	0.17
<input type="checkbox"/> velik inteligencija	1	0.17
<input type="checkbox"/> varoški inteligencija	1	0.17
<input type="checkbox"/> jak muzički inteligencija	1	0.17
<input type="checkbox"/> oskudevala prirodan inteligencija	1	0.17
<input type="checkbox"/> brajkovački inteligencija	1	0.17
<input type="checkbox"/> nov inteligencija	1	0.17
<input type="checkbox"/> sakrovište inteligencija	1	0.17
<input type="checkbox"/> korisnu inteligencija	1	0.17
<input type="checkbox"/> ceo sakrovište inteligencija	1	0.17

# SrpELTeC-NERfull

<role>[]+ </role>[]+ within <pers/>

```
<pers><role>Kapetan kraljeve garde</role>  
<persName_full>Žarko Belimarković</persName_full></pers>  
stupa u „<org_pol>Ujedinjenje ili smrt</org_pol>“.
```

CONCORDANCE

DOWNLOAD

Download first \_\_\_\_\_ rows

Or save the current view as

Download first	rows	TXT	CSV	XLS	XML
		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

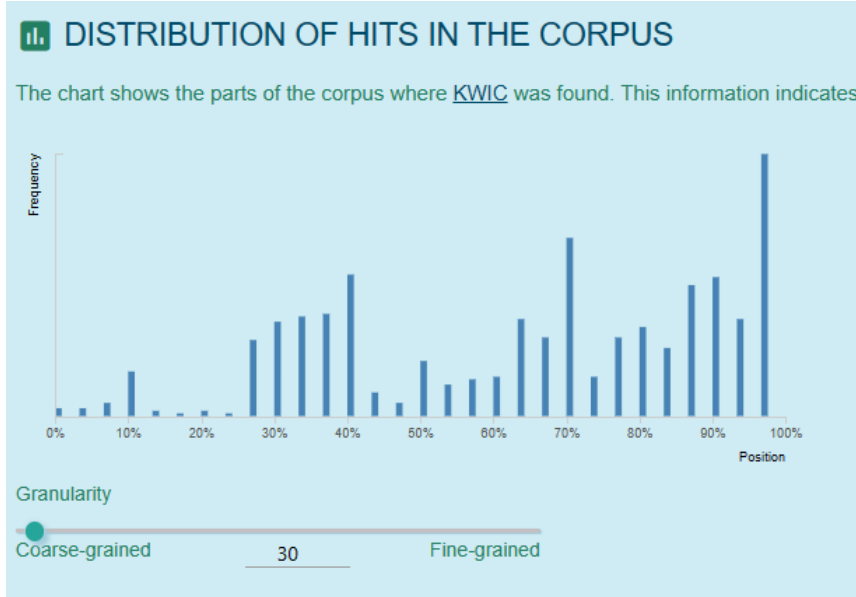
PDF

You may need to allow pop-ups in your browser when downloading.

Frequency

Show frequency per million

Lemma	↓ Frequency	Frequency per million
81 komandant Ugrinović	2	2.12
82 kapetan kraljev garda Žarko Belimarković	2	2.12
83 kapetan kraljev garda Žarko	2	2.12
84 kapetan Pribičević	2	2.12
85 harmonikašica Mara	2	2.12
86 harambaša Marinka	2	2.12
87 gospođica Amalija Bertoldina	2	2.12



# Претрага паралелног корпуса

The screenshot shows the PARALLEL CONCORDANCE web application. The browser address bar displays the URL: [https://noske.jerteh.rs/#parconcordance?corpname=BiKes\\_sr&tab=advanced&formValue=%7B%22queryselector%3A%22%3A%22%22%7D](https://noske.jerteh.rs/#parconcordance?corpname=BiKes_sr&tab=advanced&formValue=%7B%22queryselector%3A%22%3A%22%22%7D). The page title is "PARALLEL CONCORDANCE" and the current corpus is "BiKes\_sr".

The main search area shows a query: `simple [tag="V.*"][](,2)[lemma="zakon"] within <s/>` with 166 results (89.88 per million) and a note that "translations not found".

The interface is divided into two panes. The left pane is titled "CHANGE CRITERIA" and has tabs for "BASIC", "ADVANCED", and "ABOUT". It includes a search for "Srpski" and a "Query type" dropdown menu with options: simple, lemma, phrase, word, character, and CQL (which is selected). Below the dropdown is a CQL query editor with the text: `[tag="V.*"][](,2)[lemma="zakon"] within <s/>`. There is an "Insert" toolbar with symbols for brackets, arrows, quotes, ampersand, backslash, pipe, tilde, and hash. Below the editor are buttons for "TAGS" and "CQL BUILDER". A "Default attribute" dropdown is set to "word", and a "Subcorpus" dropdown is set to "none (the whole corpus)".

The right pane is titled "where" and has a search for "English". It includes a "does" dropdown menu with the option "contain" selected. It also has a "Query type" dropdown menu with options: simple, lemma, phrase, word, character, and CQL (which is selected). Below the dropdown is a CQL query editor with the text: `[tag="V.*"][](,3)[lemma="law"] within <s/>`. There is an "Insert" toolbar with symbols for brackets, arrows, quotes, ampersand, backslash, pipe, tilde, and hash, followed by a "TAGS" button. Below the editor is a "CQL BUILDER" button. A "Default attribute" dropdown is set to "word". At the bottom of the right pane, there is a checked checkbox for "Filter out untranslated segments?".

BiKes\_sr  
BiKes\_en


<http://biblisha.jerteh.rs/>



# Апликације

- Од развијених апликација издвајамо:
  - за претраживање колекције паралелних текстова,
  - за управљање лексичком базом,
  - за развој термилошких ресурса,
  - за екстракцију терминологије,
  - скуп веб сервиса за проширење упита за различите намене,
  - сервиси за анотацију текстова именованим ентитетима,
  - апликација за анотацију врста речи и лематизацију,
  - ...

# Урављање паралелним текстовима <http://biblisha.jerteh.rs/>


[ Log In ]  
[ Register ]

Home Metadata browse Metadata search Mongo search Manage data Help Tutorial About

## WELCOME TO ALIGNED TEXT COLLECTION SEARCH TOOL!

Keyword   Text collection

Synonyms	en	sr
<input checked="" type="checkbox"/> WordNet <a href="#">...</a>	AI,artificial intelligence	AI,veštačka inteligencija
<input checked="" type="checkbox"/> Dictionary of Librarianship <a href="#">...</a>	\$.: artificial intelligence - AI,AI - artificial	veštačka inteligencija
<input checked="" type="checkbox"/> Biblimir <a href="#">...</a>		
<input checked="" type="checkbox"/> GeolISSTerm <a href="#">...</a>		veštačka inteligencija
<input checked="" type="checkbox"/> RudOnto <a href="#">...</a>		veštačka inteligencija
<input checked="" type="checkbox"/> Termini <a href="#">...</a>		veštačka inteligencija

Include Hypernyms  Include Hyponyms  

Match query: both sentences  , only one language sentences  , '-' no filtering.

Morphological query expansion

Classification of Terms on a Positive-Negative

n257 – **Artificial intelligence**: automated conversation in natural language and work with clients.

n257 – **Veštačka inteligencija**: automatizovanje konverzacije na prirodnom jeziku, održavanje razgovora i rad sa strankama.

Miladinović et al., 2011, No. 19, ID: 2.2011.19.4 [metadata](#)

The new IT are based on the latest achievements in the development of computer equipment, software, communications and satellite navigation technology, robotics, **artificial intelligence**, measurement and control techniques, fuzzy logic and systems analysis.

Nove IT su zasnovane na savremenim dostignućima u razvoju računarske opreme, računarskih programa, komunikacionih i satelitskih navigacionih tehnologija, robotike, **veštačke inteligencije**, merno-regulacione tehnike, fazi logike i sistemske analize.

Rehm et al., 2013, vol. XIV:1, ID: 1.2013.1.1 [metadata](#)

These barriers harm the EU's long-term goal of establishing a single digital market because they hinder the free flow of knowledge, information, goods, products, and services," says Prof. Hans Uszkoreit, coordinator of META-NET and scientific director at DFKI (German Research Center for **Artificial Intelligence**).

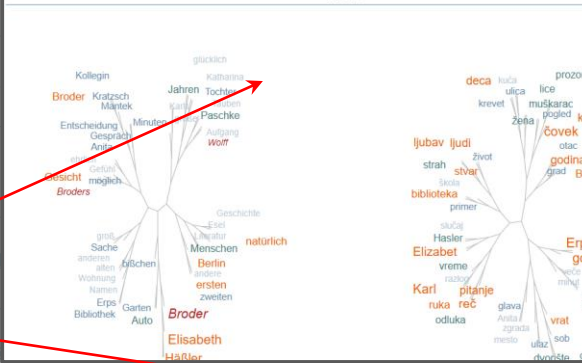
Ove barijere nanose štetu dugoročnom cilju EU da uspostavi jedinstveno digitalno tržište, jer ometaju slobodan protok znanja, informacija, roba, proizvoda i usluga, " kaže dr Hans Uskorajt, koordinator META-NET-a i naučni direktor u DFKI (Nemački istraživački centar za **veštačku inteligenciju**).

# SRPNEMKOR

Number of documents: 14 Number of sentences: 48004

Journal Document	About Document
11.1.001 vol. 1 11.1	<b>Moje nagrade</b> [VIAF] [GND] [WikiData] [LCN] authors: <b>Thomas Bernhard</b> [VIAF] [GND] [WikiData] [detaljnije] [tmx] [pdf]
11.1.002 vol. 1 11.1	<b>Pijanistkinja</b> [VIAF] [GND] [WikiData] [LCN] authors: <b>Elfride Jelinek</b> [VIAF] [GND] [WikiData] [detaljnije] [tmx] [pdf]
11.1.003 vol. 1 11.1	<b>Beč, juli 1999</b> [VIAF] [GND] authors: <b>Milo Dor</b> [VIAF] [GND] [WikiData] [LCN] [detaljnije] [tmx] [pdf]
11.1.004 vol. 1 11.1	<b>Hodov raka</b> [VIAF] [GND] [WikiData] [LCN] authors: <b>Günter Grass</b> [VIAF] [GND] [WikiData] [LCN] [detaljnije] [tmx] [pdf]
11.1.005 vol. 1 11.1	<b>Buridanov magarac</b> [VIAF] [GND] [WikiData] authors: <b>Günter de Bruyn</b> [VIAF] [GND] [WikiData] [detaljnije] [tmx] [pdf]
11.1.006 vol. 1 11.1	<b>Poslednji svet</b> [VIAF] [GND] [WikiData] authors: <b>Christoph Ransmayr</b> [VIAF] [GND] [WikiData] [LCN] [detaljnije] [tmx] [pdf]
11.1.007 vol. 1 11.1	<b>Parfem: hronologija jednog zločina</b> [VIAF] [GND] [WikiData] authors: <b>Patrik Süskind</b> [VIAF] [GND] [WikiData] [LCN] [detaljnije] [tmx] [pdf]
11.2.001 vol. 1 11.1	<b>Mamac</b> [VIAF] [GND] [LCN] authors: <b>David Albahari</b> [VIAF] [GND] [WikiData]

Buridans Esel	Buridanov magarac
SrpNemKor, Roman [pdf]	SrpNemKor, roman [pdf] [VIAF] [GND] [WikiData]
ID: 11.1.005 Number: 1 Volume: 1 Month: 1 Year: 1 UDC: 821.112.2-31	[bow] [tmx]
<b>Günter de Bruyn</b>	<b>Günter de Bruyn</b> [VIAF] [GND] [WikiData] [LCN]
<b>Abstract</b> Dieser Roman von Günter de Bruyn ist eine Liebesgeschichte. Der Titel des Buches, der sich auf die Fabel des französischen Philosophen Buridan vom Esel, der sich zwischen zwei gleichgroßen Heubündeln nicht entscheiden kann und verhungert, bezieht, dient als eine im weiten Sinne gefälschte Metapher für die Erzählung eines Dreieckskonflikts.	<b>Apstrakt</b> De Brojnovli junaci doživljavaju svoje radosti muku u Istočnom Berlinu sezdesetih godina pozadini realocijalizma kao da se još jače boje autorove bogate palete. De Brojn nas rukom provodi kroz meandre ljudske palhe ogledalo i izmamljujući nam osmehe koji ni Ovo je prvi roman Gintera de Brojna koji se prevodu na srpski.
Pages: 199	Strane: 272
Publishing place: Frankfurt am Main Publisher: Fischer-Taschenbuch Publishing year: 1992	Mesto izdanja: Beograd Izdavač: Fabrika knjiga Godina izdanja: 2007
Translator:	Prevodilac: s nemačkog prevela Aleksandra Vučen



BURDANS ESEL / GUNTER DE BRUYN = BURIDANOV MAGARAC / GINTER DE BROJN, ID: 11.1.005	
En/De/Fr/It- (first 9 out of 2890 sentences) [pdf]	Srpski - (prvih 9 od 2890 rečenica) [pdf]
n1 Günter de Bruyn	n1 Ginter de Brojn
n2 Buridans Esel: Roman	n2 BURIDANOV MAGARAC
n3 Anfechtungen hat es so:	n3 Evo kako je sve počelo:
n4 Karl Erp lächelte beim Erwachen und wußte nicht, warum.	n4 Karl Erp se probudio nasmejan a nije znao otkud to.
n5 An einen Traum entsann er sich nicht.	n5 Nije mogao da se seti da je lišta sanjao.
n6 Erst später, nicht viel später, aber doch erst danach, fiel ihm Fräulein Broder ein.	n6 Tek kasnije, ne baš mnogo kasnije, ali ipak danach, fiel ihm Fräulein Broder ein.
n7 Sagen wir lieber:	n7 Ili bolje da kažemo ovako:
n8 Nachträglich schien ihm, daß er an jenem Morgen lächelnd erwacht sei.	n8 naknadno mu se učinilo da se tog jutra probudio nasmejan.
n9 (Und so erzählte er Fräulein Broder auch:	n9 (Pa je to tako I ispriča gospodici Broder:

<https://www.wikidata.org/wiki/Q110637016>

Q110637016	Erp	gospodica	Broder	sal
Буриданов магарац : роман : СрпНемКор издање	srpski jezik			
jezik dela ili imena	srpski jezik			
datum izdavanja	2019			
mesto izdavanja	Beograd			
- 0 referencе				
+ dodaj referencу				
Буриданов магарац : роман (издање из 1992. године)	немачки jezik			
jezik dela ili imena	немачки jezik			
datum izdavanja	1992			
mesto izdavanja	Франкфурт на Мајни			
- 0 referencе				
+ dodaj referencу				
Буриданов магарац : роман (издање из 2007. године)	srpski jezik			
jezik dela ili imena	srpski jezik			
datum izdavanja	2007			
mesto izdavanja	Beograd			
- 0 referencе				
+ dodaj referencу				
+ dodaj vrednost				



# veštačka inteligencija



Check in dictionaries:

- [show WordNet](#)

ENG30-06133203-n, BCS=, +0.00000, -0.00000

grana računarstva koja se bavi pisanjem računarskih programa koji mogu da rešavaju probleme  
veštačka inteligencija 1

AI 2

DOM: computer\_science, SUMO: FieldOfStudy @

---> *hypernym*: ENG30-06128570-n računarstvo; računarska nauka

... <--- *hyponym* <- *hypernym*: ENG30-06133654-n robotika

... <--- *hyponym* <- *hypernym*: ENG30-06133503-n mašinsko prevodenje

- [show Terminološki](#)

- [show Bi-lista](#)

Check in external dictionaries: [Wiktionary](#) [Babelnet](#) [Termi](#) [Glosbi](#)

Senses (1):

## 1. +Comp

Domains:

Properties: polileksemna jedinica

Is composed of:

Form	Lemma	FST Code	Gram Cat	Separator
<a href="#">veštačka</a>	veštački	A2	aefs1g	
<a href="#">inteligencija</a>	inteligencija	N600	fs1q	

URL	Text
https://leximirka.jerteh.rs/LexicalEntry/Dela?eid=216789 - Profil 1 - Microsoft Edge	
https://leximirka.jerteh.rs/LexicalEntry/Dela?eid=216789	
	veštačka inteligencija,veštačka(veštački.A2:aefs1g) inteligencija(inteligencija.N600:fs1q).NC:fs1q
	veštačka inteligencijo,veštačka(veštački.A2:aefs1g) inteligencija(inteligencija.N600:fs1q).NC:fs5q
	veštačke inteligencije,veštačka(veštački.A2:aefs1g) inteligencija(inteligencija.N600:fs1q).NC:fs2q
	veštačkoj inteligenciji,veštačka(veštački.A2:aefs1g) inteligencija(inteligencija.N600:fs1q).NC:fs3q:fs7q
	veštačkom inteligencijom,veštačka(veštački.A2:aefs1g) inteligencija(inteligencija.N600:fs1q).NC:fs6q
	veštačku inteligenciju,veštačka(veštački.A2:aefs1g) inteligencija(inteligencija.N600:fs1q).NC:fs4q

Cvetana Krstev, **Processing of Serbian – Automata, Texts and Electronic dictionaries** Faculty of Philology, University of Belgrade, Belgrade, 2008.

Биљана Рујевић, “Речници у дигиталном добу - информатичка подршка за српски језик” (2022), M70, [pdf](#)

Biljana Lazić, Mihailo Škorić, “**From DELA Based Dictionary to Leximirka Lexical Database**”, Infotheca (2020), Faculty of Philology, University of Belgrade. [pdf](#)

Ranka Stanković, Cvetana Krstev, Biljana Lazić, Mihailo Škorić, “**Electronic Dictionaries - from File System to lemon Based Lexical Database**”, W23 6th Workshop on Linked Data in Linguistics : Towards Linguistic Data Science (LDL-2018), LREC 2018, Miyazaki, Japan, May 7-12, 2018 (2018).[pdf](#)

# Inteligencija

N600 delas-im.dic

## NOUN

Relations:

- To [neinteligencija](#) using [negacija](#) ([\\_ne](#))

Check in dictionaries:

- show [DRJ](#)
- show [Stranih reči](#)
- show [WordNet](#)
- show [RSinonima](#)
- show [Terminološki](#)
- show [BI-lista](#)
- show [HŠR](#)

Check in external dictionaries: [Wiktionary](#) [Babelnet](#) [Termini](#) [Glosbi](#)

Frequencies:

- Top 10000 most frequent in [SrbCorp122M Corpus](#) by D.Vitas, M.Utvić (16.68 per million)
- Top 10000 most frequent in [matematika Corpus](#) by M.Radojičić (22.92 per million)
- Top 10000 most frequent in [RucniKor Corpus](#) by D. Vitas, C. Krstev, R. Stanković (13.12 per million)
- Top 10000 most frequent in [BiKes\\_sr Corpus](#) by (10.99 per million)
- Top 5000 most frequent in [WikiKorpus Corpus](#) by (17.56 per million)
- Top 5000 most frequent in [SrFudKo Corpus](#) by (3.73 per million)
- Top 10000 most frequent in [SrpKor2021 Corpus](#) by (3.60 per million)

Search corpora: [SrpKor2021](#) [A\(N\)](#) [Concordances](#) [Form Frequences](#) [Lemma Frequences](#)

- [SrpKorPRGF](#)

Senses (1):

### 1. +UPOS=NOUN

- Domains:
- Properties:
- Is a component of:
- veštačka inteligencija
  - tehnička inteligencija
  - vsještačka inteligencija

da Adali zamenjuje inteligenciju pojedina .	<b>samom</b> <b>Inteligencijom</b> .	
me je izlećio . Razume se, samoća je opasna za	<b>živu</b> <b>inteligenciju</b> .	Potrebna nam je okolina ljudi koji misle i
tehničke mogućnosti , upravljajska struktura ,	<b>veštačka</b> <b>inteligencija</b> .	zakonska regulativa , zaštita spomenika
( 17 ) , ućenje na daljinu ( 32 ) , veb sajt ( 5 ) ,	<b>veštačka</b> <b>inteligencija</b> .	( 4 ) itd. ) . Kod termina koji su sačinjeni od
4 1 NXZ grf03 sin veličina slova 4 1 AXN grf01 sin	<b>veštačka</b> <b>inteligencija</b> .	4 1 AXN grf01 sin visokoškolska biblioteka 95 1
znanjem 6 NXN grf06 sin veb sajt 5 AXN grf01 sin	<b>veštačka</b> <b>inteligencija</b> .	4 AXN grf01 sin visokoškolska ustanova 9 AXN
pojavnom obliku može da nastupi samo tamo gde se	<b>natprosćna</b> <b>inteligencija</b> .	spari s trajnim mladenaćim oduševljenjem (
nagone ostaci proraćunatosti i neka gotovo	<b>instinktivna</b> <b>inteligencija</b> .	A prvi uspeši već su se videli . Krać se
mu se nemo diviti , dobro kuvati i koja neće imati	<b>više</b> <b>inteligencije</b> .	nego što je to neophodno kad je čoveć pokaće
... A Drio , koga istina nije krasila neka	<b>izuzetna</b> <b>inteligencija</b> .	ali koji nije bio bać potpuna budala , shvatio
trgovaca , industrijalaca , i jaka	<b>jevrejska</b> <b>inteligencija</b> .	koja je proćzela ceo maćarski javni i kulturni
ali je bilo moguće interpretirati samo 4 . To su :	<b>moćorićka</b> <b>inteligencija</b> .	generalni faktor telesne snage , generalni
primer : znanja , intelektualne sposobnosti ,	<b>socijalna</b> <b>inteligencija</b> .	komunikacijska sposobnost , jezićka
kompetencije nadovezuje na konstrukt	<b>socijalne</b> <b>inteligencije</b> .	i socijalne kompetencije . U istraživanju
posmatrano dominiraju sledeći pojmovi :	<b>socijalna</b> <b>inteligencija</b> .	socijalna kompetentnost , komunikaciona
do sedamdesetih godina XX veka koristi pojam	<b>socijalna</b> <b>inteligencija</b> .	koja je proćzela ceo maćarski javni i kulturni
da se u budućnosti više neće ispitivati	<b>socijalna</b> <b>inteligencija</b> .	nego socijalna kompetentnost . Kao
se naglašava delatni aspekt široko shvaćene	<b>socijalne</b> <b>inteligencije</b> .	( Zlatić , Bjekić , 2004:325 ) . On socijalni

https://leximirkajerteh.rs/CQP/NoskeFreq

### A(N)

veštaćke inteligencije	581
veštaćka inteligencija	220
veštaćku inteligenciju	196
srpske inteligencije	158
veštaćkoj inteligenciji	106
poštena inteligencija	105
Veštaćka inteligencija	102
veštaćkom inteligencijom	75
poštene inteligencije	70
emocionalne inteligencije	65
poslovne inteligencije	55
srpska inteligencija	51
kritićke inteligencije	48
koeficijenta inteligencije	47
emocionalna inteligencija	46
nacionalne inteligencije	45
poštenu inteligenciju	45
tehnićke inteligencije	44
Emocionalna inteligencija	39
ljudske inteligencije	39
misionarske inteligencije	33
politićke inteligencije	31

veštaćki inteligencija	1307
pošten inteligencija	269
srpski inteligencija	256
emocionalan inteligencija	200
misionarski inteligencija	120
poslovan inteligencija	110
kritićki inteligencija	99
tehnićki inteligencija	92
socijalan inteligencija	86
ljudski inteligencija	82
nacionalan inteligencija	79
liberalan inteligencija	62
visok inteligencija	59
natprosćčan inteligencija	57
politićki inteligencija	55
koeficijent inteligencija	53
ruski inteligencija	50
prosćčan inteligencija	44
elementaran inteligencija	43
kolektivian inteligencija	41
seksualan inteligencija	40



# Поглед на корпус из терминолошке базе

← ↻ 🔒 https://termi.rgf.bg.ac.rs/Dodaj/Pretrazi/?term=ugalj

Prelistavanje Pretraga Pomoć

## Pretraga

ugalj

**Srpski**

- aktivan ugalj
- aktivni ugalj
- ugalj
- podsekačica za ugalj sa utovaračem
- utovarno mesto za ugalj

**ugalj** (Čvrsta, krta, manje ili više jasno stratifikovana zapaljiva karbonatna stena, formirana delimičnom ili potpunom dekompozicijom vegetacije; varira u boji od tamno braon do crne; ne može se rastopiti bez razlaganja i vrlo teško se rastvara.)

**Sinonimi:** ugljen

**Nadređeni:** ugalj > mineralne sirovine > materijal > rudarstvo

**Podređeni:** antracit, čiste tehnologije uglja, gasifikacija uglja, koks, kokсни ugalj, listasti ugalj, mrki ugalj, plameni ugalj, prašinst ugalj, ugljena prašina

Bilibsha 🧑 NaRa 🧑 Detaljno 🔍

Pretraži korpus:  Konkordance 🔍 Frekvencije oblika 🔍

© 2022 - Univerzitet u Beogradu, Rudarsko-geološki fakultet

https://termi.rgf.bg.ac.rs/CQP/NoskeFreq

### Plain lemma

ugalj	7001
ugalj	865
uglju	427
ugljem	223
Ugalj	164
UGLJA	12
ugalja	4
uglje	1
Ugljem	1
Uglja	1

**Synonyms:**

**Hyperonyms:** flame coal > mineral raw materials > raw material engineering

**Hyponyms:** anthracite, clean coal technologies, coal gasification, coking coal, paper coal, brown coal, flame

Concordances - Profil 1 - Microsoft Edge

https://termi.rgf.bg.ac.rs/CQP/Noske

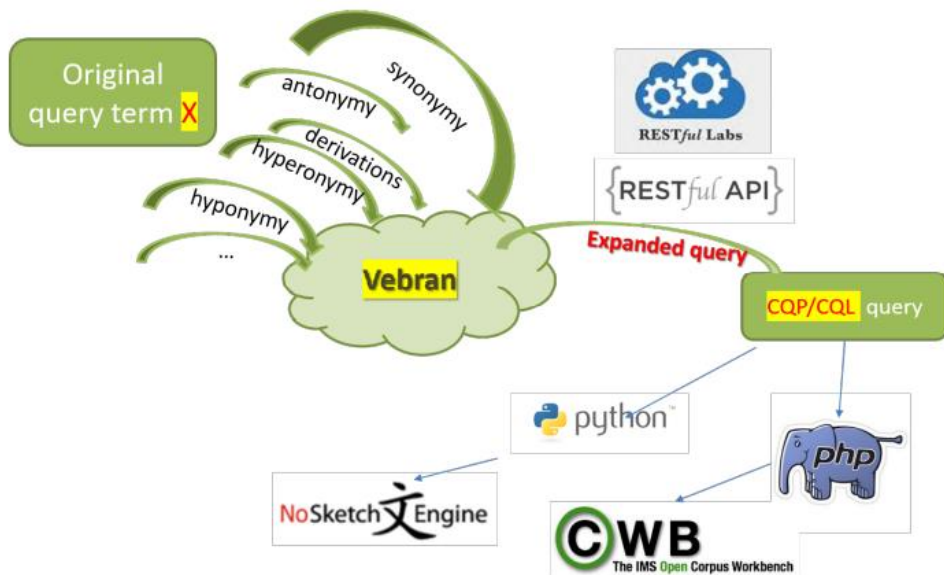
### Plain lemma

1. 2 prikazana je tehnološka šema pripreme ugalja za sagorevanje u termoelektranama .

1. 2. Tehnološka šema pripreme i deponovanja ugalja Eksploatacija siromašnijih ležišta zahteva transportuje kamionima i damperima, a ugalj transportnim trakama. Znatno ređe se sreću ; učešće štetne prašine - prilikom transporta ugalja na pretovarnim mestima obično dolazi do težine, kao što su : suvi šljunak, suvi kameni ugalj, suvi koks itd. Nepravilan zrnast ili grudvast itd. Komadasti materijali kao što su : kameni ugalj, šljaka, većina ruda, pepeo itd.

istraživanja je obavljao sa hidromešavinom ugalja mereći koncentraciju ( u cevovodu ND 350 mm . pri koncentracija, % Č Milerov broj Gips 50 19, 3 Ugalj 50 39, 7 Otpad od ugalja, bez pirita 25 53, 7 Č Milerov broj Gips 50 19, 3 Ugalj 50 39, 7 Otpad od ugalja, bez pirita 25 53, 7 Krečnjak 50 72, 0 Jalovina 50 72, 0 Jalovina rude bakra 50 80, 8 Otpad od ugalja, sa 13 % pirita 50 114, 0 Jalovina rude gvožđa 50 uložka rastresiti fini i laki materijali 2 1 W ugalj, pepeo, šljunak, pesak, fina ruda 2 do 4 2 W / Y, šljunak, pesak, fina ruda 2 do 4 2 W / Y komadni ugalj, stene, rude, stene otkrivke 4 do 8 2 do 3 Y do 16 3 do 4 X Trake sa karakasom od čeličnih užadi ugalj, pepeo, šljunak, pesak, fina ruda 4 do 8 4 do 6 W / Y šljunak, pesak, fina ruda 4 do 8 4 do 6 W / Y komadni ugalj, stene, rude, stene otkrivke 6 do 12 4 do 8 Y na stacionarnim transporterima za istovar ugalja u više bunkera . U tom slučaju se mogu koristiti i slični materijali sa   2000 kg / m 3 i amax  60 mm i amax  60 mm 1, 5 do 3, 0 2, 0 Srednje krupni kameni ugalj, briket i slični materijali sa   2000 kg / m 3 i kao što su : rovna i drobljena ruda, šljunak, ugalj i sl. Opšta šema člankastog transportera nasipni materijali ( prašinst ili drobljen ugalj, pesak, glina, cement itd ) . Utovar materijala, drobljena ruda, cementni klinker, krupni ugalj itd. ) teško se zahvataju vedrima jer se zbog

# Вебран у функцији проширења упита над корпусима



[flemma="kap" ] "kap(|i|ima|lju)"  
... & gram= "s" ] "kap(|i|lju)"

Stanković, Ranka; Utvić, Miloš. Vebran Web Services for Corpus Query Expansion. *Infotheca - Journal for Digital Humanities*, v. 19, n. 2, p. 99-118, 2020.

[\[pdf\]](#)

edno | obrada.jerteh.rs

JePTex – Друштво за језичке ресурсе и технологије



Paste Text   Upload File

Danas predstavljamo Jerteh na Seminaru za veštačku inteligenciju.

TreeTagger sr-lat  
 TreeTagger sr-cyr  
 TreeTagger en  
 Cyr to Lat  
 Lat to Cyr  
 s graf (lat)  
 s graf xml (lat)

Go

Danas   ADV   danas  
 predstavljamo   V   predstavljati  
 Jerteh   V   Jerteh  
 na   PREP   na  
 Seminaru   N   seminar  
 za   PREP   za  
 veštačku   A   veštački  
 inteligenciju   N   inteligencija  
 .   PUNCT   .

Sent. del.: 0   Tokens: 16   Simple forms: 8   Digits: 0   Errors: 1

Error percent: 12.5%

- Ranka Stanković, Mihailo Škorić, Branislava Šandrih Todorović, “Parallel Bidirectionally Pretrained Taggers as Feature Generators” Applied Sciences (2022), [\[pdf\]](#)
- <https://github.com/procesaur/BEaSTagger>
- Ranka Stanković, Branislava Šandrih, Cvetana Krstev, Miloš Utvić, Mihailo Škorić, “**Machine Learning and Deep Neural Network-Based Lemmatization and Morphosyntactic Tagging for Serbian**”, Proceedings of the 12th Language Resources and Evaluation Conference, 2020, Marseille, France (2020), ELRA [\[pdf\]](#).
- Duško Vitas and **Cvetana Krstev**. “Processing of Corpora of Serbian Using Electronic Dictionaries”. in *Prace Filologiczne*, vol. LXIII, pp. 279-292, Warszawa, 2012. ISSN 0138-0567 [PDF](#)

## Process plain text file, TMX, or a string

Upload a file here: Odabir datoteke Nije izabran... dna datoteka

Or type in/paste the text below:

Select processing Language: ca (if file is not a TMX)

Select parsing method: NER

Process



## About the project

**Project name:** It-Sr-NER: CLARIN compatible NER and geoparsing web services for parallel texts: case study Italian and Serbian 2022

**Call** "Bridging Gaps - A Call for Expressions of Interest to Connect CLARIN to External Language Technology Tools" 2022

**Project duration:** 1.6.2022-30.9.2022.

**Project leader:** Olja Perišić, Università degli Studi di Torino, Dipartimento di Lingue e Letterature Straniere e Culture Moderne, Italy

## Resources

txt\_0-it-sr-TMX.xml 595.78 KiB

Mediatype: application/xml

Language: Select language

+ Add another resource

## Matching Tools

Group by task  Search for tool

### Named Entity Recognition

- > spaCy bilingual NEL (for TMX) [Open](#)
- > spaCy bilingual NER (for TMX) [Open](#)
- > spaCy bilingual NER and Geoparsing (for TMX) [Open](#)
- > spaCy binlingual NER and NEL (for TMX) [Open](#)

PERS	LOC	ORG	DEMO	WORK	EVENT	Download XML		
n1	America	Europa	Ameriku	Evropu	country located mainly in North America			
n2	Nino	Pietro	Nina	Pjetra				
n3	Guido Airota	Gvido Ajrota						
n4	Strasbourg	Chartres	Bamberg	Parigi	Strazburu	Sartrni	Bambergu	Parizu
n5	Dobrun							
n6	Drina	Serbia						

- Olja Perišić, Ranka Stanković, Milica Ikonić Nešić and Mihailo Škorić, **It-Sr-NER: Веб сервиси за препознавање и повезивање именованих ентитета у тексту и њихово приказивање на веб карти**, Infoteka 2023 (prihvaćeno za publikovanje)
- Olja Perišić, Ranka Stanković, Milica Ikonić Nešić and Mihailo Škorić, **It-Sr-NER: CLARIN compatible NER and geoparsing web services for Italian and Serbian parallel text**, CLARIN2022: Selected papers in Post-Conference Proceedings (prihvaćeno za publikovanje)
- Olja Perišić, „**It-Sr-NER: CLARIN пројекат**“ – *Универзитет у Торину и ЈеРТех тим*, 20.10.2022. Семинар Јертеха [\[pdf\]](#)

## Терминологија / Terminology

Bilingual Terminology Extraction: BiTE

## Препознавање именованих ентитета / Named Entity Recognition

NER&Beyond

Serbian NER models

BRAT: Rapid Annotation Tool

Inception: A Semantic Annotation Platform

## Лексикографија и екстракција обележја / Lexicography & Feature Extraction

Good Dictionary EXamples

Feature Extraction

## Препроцесирање текста / Text Preprocessing

POS-tagger for Serbian and cyr-lat converter

KaMP: Towards Minimal Pairs

## Студентски пројекти / Student Projects

Baby WordNet: Python Wordnet Interface for English

Hate Speech Detection for Serbian

Text Generator for Serbian

- Tanja Ivanović, Ranka Stanković, Branislava Šandrih Todorović, Cvetana Krstev, “**Corpus-based bilingual terminology extraction in the power engineering domain**” у Terminology (2022), John Benjamins Publishing Company.
- Cvetana Krstev, Branislava Šandrih, Ranka Stanković, “**Using English Baits to Catch Serbian Multi-Word Terminology**” Proceedings of the 11th International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018 (2018), European Language Resources Association (ELRA). [\[pdf\]](#)
- Branislava Šandrih, Ranka Stanković, “**Extraction of Bilingual Terminology Using Graphs, Dictionaries and GIZA++**”, Infotheca (2020), Faculty of Philology, University of Belgrade
- Ranka Stanković, Branislava Šandrih, Rada Stijović, Cvetana Krstev, Duško Vitas, Aleksandra Marković, “**SASA Dictionary as the Gold Standard for Good Dictionary Examples for Serbian**” Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference (2019), Lexical Computing CZ, s.r.o. [\[pdf\]](#)
- Рада Стијовић, Цветана Крстев, Ранка Станковић, “**Аутоматска екстракција дефиниција – допринос убрзању израде речника**” у Лексикологија и лексикографија у светлу актуелних проблема (2021), Институт за српски језик САНУ.
- Danka Jokić, Ranka Stanković, Cvetana Krstev, Branislava Šandrih, “**A Twitter Corpus and Lexicon for Abusive Speech Detection in Serbian**”, 3rd Conference on Language, Data and Knowledge (LDK 2021) (2021), MDPI AG.



# Анотирани скупови података

Развијени су различити анотирани скупови података:

- за обележавање врстом речи,
- за лематизацију,
- за обележавање именованих ентитета,
- за анализу ставова изражених у текстовима,
- за препознавање говора мржње.

The screenshot displays the ELG website interface for the 'SrpELTeC-gold - Named Entity Recognition Training corpus for Serbian' dataset. The page includes the ELG logo, navigation links for 'Catalogue', 'Documentation & Media', and 'About', and a 'Go to catalogue' button. The main content area features the dataset title, version information (1.0.0, dated 26/06/2021), and a 'hosted in ELG' badge. Below this, there are sections for 'Keyword' (manually annotated corpus, named entity corpus, Serbian) and 'Corpus subclass' (annotated corpus). The right sidebar contains 'Cite resource' and 'Cite all versions' information, including the dataset title, authors (Stanković, Ranka; Ikonić Nešić, Milica; Šandrih Todorović, Branislava; Krstev, Cvetana), version, and DOI links. At the bottom, there is a 'Share' section with social media icons and a 'Read more' button. A descriptive paragraph at the bottom left explains that the dataset consists of 11 full novels and excerpts from 15 novels from the Serbian literary corpus, which have been automatically labeled using the SrpNER system.

# Корпуси

- [SrpELTeC-gold](#) – *Named Entity Recognition Training corpus for Serbian* – Корпус за *тернирање препознавања именованих ентитета* је поткорпус књижевног корпуса српског језика који садржи 11 романа у пуном тексту и исечке 15 романа написаних пре више од једног века. У првој фази припреме је обележен системом SrpNER. Садржи 330.119 токена, 7 класа: особе, организације, локације, догађаји, дела, демоници, улоге (професије), лиценца **CC-BY-NC-SA-4.0**.
- [SrpKor4Tagging](#) – Корпус је формиран комбинацијом књижевних (⅓) и административних (⅔) текстова на српском језику. Обележен је са два скупа ознака за врсте речи: Universal POS и SrpLemKor (скуп креиран на основу традиционалне, дескриптивне граматике српског језика) и лематован, 342,803 токена, лиценца **CC-BY-4.0**.
- [RudKorP](#) – Српски јавни корпус текстова из области рударства и обраде минералних сировина настао на Универзитету у Београду, Рударско-геолошком факултету. Садржи 2.34 милиона речи, лиценца **CC-BY-4.0**.
- [ELTEC](#) – Колекција европских романа у оквиру које је и српски скуп <https://distantreading.github.io/ELTeC/srp/index.html> (4.931.503 речи), такође доступно и на <https://github.com/COST-ELTeC/ELTeC-srp> ниво 2 укључује 100 романа: речима придружене леме и врсте речи, обележено 7 класа именованих ентитета, лиценца **CC-BY-4.0**.
- [INTERA Corpus – the Serbian-English part 1](#) – паралелни корпус од милион речи за српски и милион речи за енглески, упарено на нивоу реченице, лиценца **CC-BY-4.0**.
- [INTERA Corpus – the Serbian POS annotated part of the SR-EN pair](#) – милион речи, лиценца **CC BY 4.0**.
- [Орвелова 1984](#) – аутоматски аотиран и ручно коригован пуним граматичким категоријама, врстама речи и лемама, у оквиру вишејезичног пројекта, лиценца MULTEXT-East **CC BY-NC-SA 4.0**.
- [Пут око света за 80 дана](#) паралелни корпус са 3.700 упарених сегмената, углавном реченица, лиценца **CC-BY-NC-SA-4.0**.

# SrpELTeC

SrpELTeC је корпус старих српских романа први пут публикованих у периоду 1840-1920. године, дигитализованих у склопу COST акције *CA16204: Distant Reading for European Literary History, 2018-2022.*

Корпус обухвата 120 романа са 5.263.071 речи, 22700 страница, 2557 поглавља, 158.317 пасуса, 567 песама, 2972 стихова, 803 сегмента на страном језику и 949 поменутих дела.

[Информације о колекцији SrpELTeC добијене уз помоћ WIKIDATA-SPARQL упита](#)

[Приступ романима из колекције путем платформе Удаљено читање](#)

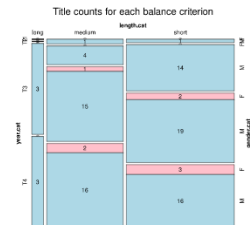
[Приступ романима путем дигиталне библиотеке Код Ауроре](#)

[Претрага корпуса путем платформе NoSketchEngineh](#)

[Подаци о 100 романа,](#)

[Подаци о 20 романа](#)

On 2022-10-05 ELTeC-srp contains 100 texts containing 4931593 words



Click on a column heading to sort. Click on a text identifier to read the text (may not work in older browsers).

Identifier	Encoding	Pages	Words	(Size)	Date (Slot)	Title	Author	Sex	Reprints
<a href="#">SRP18520</a>	eltec-1	204	44207	(short)	1852 (T1)	Два идола	Алавацковић, Соробој (1856-1888)	M	high
<a href="#">SRP18550</a>	eltec-1	332	97481	(medium)	1859 (T1)	Ђурађ Брањковић	Игњатовић, Јаков (1822-1889)	M	low
<a href="#">SRP18620</a>	eltec-1	97	20858	(short)	1862 (T2)	Једна женица	Игњатовић, Јаков (1822-1889)	M	high
<a href="#">SRP18630</a>	eltec-1	256	88117	(medium)	1863 (T2)	Милан Нарањчић	Игњатовић, Јаков (1822-1889)	M	high
<a href="#">SRP18631</a>	eltec-1	282	66347	(medium)	1863 (T2)	Кочина крајина	Ђорђевић, Владан (1844-1930)	M	low
<a href="#">SRP18680</a>	eltec-1	65	19320	(short)	1868 (T2)	У гостионици код "Полу-левака" на илеандан шаманског торбара	Грић, Јован Милењко (1846-1875)	M	high
<a href="#">SRP18690</a>	eltec-1	116	36543	(short)	1869 (T2)	Глушанско језеро	Ђорђевић, Владан (1844-1930)	M	low
<a href="#">SRP18691</a>	eltec-1	136	31620	(short)	1869 (T2)	Какав је ко оноко му и бива	Поповић, Панча (1843-1918)	M	low
<a href="#">SRP18730</a>	eltec-1	361	64724	(medium)	1874 (T2)	Калуђер	Суботић, Јован (1817-1886)	M	low
<a href="#">SRP18740</a>	eltec-1	52	14247	(short)	1874 (T2)	Сељаци	Захшић, Ђура (1832-1878)	M	high
<a href="#">SRP18741</a>	eltec-1	34	11223	(short)	1875 (T2)	Сирота Бананка	Захшић, Ђура (1832-1878)	M	high
<a href="#">SRP18750</a>	eltec-1	156	48661	(short)	1875 (T2)	Васа Рашчак	Игњатовић, Јаков (1822-1889)	M	high
<a href="#">SRP18751</a>	eltec-1	11035	(short)	1875 (T2)	Глава шећера	Глишић, Милован (1847-1908)	M	high	
<a href="#">SRP18752</a>	eltec-1	167	29321	(short)	1875 (T2)	Без оца и мајке	Ђузуловић, Пера (????-????)	M	low
<a href="#">SRP18760</a>	eltec-1	103	29793	(short)	1876 (T2)	Чича Тина	Захшић, Ђура (1832-1878)	M	low
<a href="#">SRP18780</a>	eltec-1	235	57165	(medium)	1878 (T2)	Пред зору	Милићевић, Бранко (Лукићева Брањковић) (1857-1918)	M	low
<a href="#">SRP18790</a>	eltec-1	110	18012	(short)	1879 (T2)	Зурмуса и Фатима или Турска сила сама себе једи	Милићевић, Милан Ђ. (1831-1908)	M	low
<a href="#">SRP18791</a>	eltec-1	227	32447	(short)	1879 (T2)	Пастир краљ или Ослобођење Србије	Барунић, Коста (????-????)	M	low
<a href="#">SRP18792</a>	eltec-1	61	13096	(short)	1879 (T2)	Жуђуци	Милићевић, Милан Ђ. (1831-1908)	M	low
<a href="#">SRP18793</a>	eltec-1	63	10304	(short)	1879 (T2)	Љубав и патриотизам	Зрењин, Јован (1860-1931)	M	low
<a href="#">SRP18800</a>	eltec-1	241	51104	(medium)	1880 (T3)	Драгољана огрлица	Комарчић, Лазар (1833-1909)	M	low
<a href="#">SRP18840</a>	eltec-1	34	18652	(short)	1884 (T3)	Из учитељског живота	Гаариловић, Драга (1854-1917)	F	high
<a href="#">SRP18860</a>	eltec-1	257	33284	(short)	1886 (T3)	Омер Челебија	Милићевић, Милан Ђ. (1831-1908)	M	low
<a href="#">SRP18870</a>	eltec-1	36	19350	(short)	1887 (T3)	Бабадевојка	Гаариловић, Драга (1854-1917)	F	high

- Aleksandra Trtovac, Vasilije Milnović, Cvetana Krstev, **The Serbian Part of the ELTeC - from the Empty List to the 100 Novels Collection**, [\[pdf\]](#)
- Cvetana Krstev, **The Serbian Part of the ELTeC Collection Through the Magnifying Glass of Metadata**, [\[pdf\]](#)
- Ranka Stanković, Cvetana Krstev, Branislava Šandrih Todorović, Mihailo Škorić, **Annotation of the Serbian ELTeC Collection**, [\[pdf\]](#)
- Milica Ikonić Nešić, Ranka Stanković, Biljana Rujević, **Serbian ELTeC Sub-Collection in Wikidata**, [\[pdf\]](#)
- Duško Vitas, **From Onions to Champagne – Food and Drink in the SrpELTeC Corpus**, [\[pdf\]](#)
- Cvetana Krstev, **White as Snow, Black as Night – Similes in Old Serbian Literary Texts**, [\[pdf\]](#)
- Ranka Stanković, Mihailo Škorić, Petar Popović, **SrpELTeC on Platforms: Udaljeno čitanje, Aurora, noSketch**, [\[pdf\]](#)

## Reviews

- Jelena Andonovski, **OCR and TEI for the Production of ELTeC - Würzburg Training School, 16-17 April 2018**, [\[pdf\]](#)
- Mihailo Škorić, **Methods and Tools of Distant Reading Adapted to Multiple European Languages**, Galway Training School, [\[pdf\]](#)
- Vasilije Milnović, Budapest Training School – **Canonization in Distant Reading Research**, [\[pdf\]](#)
- Mihailo Škorić, STSM Krakow: **Comparative Stylistic and Morphosyntactic Analysis of ELTeC Texts Using Stylo R Package**, [\[pdf\]](#)
- Ranka Stanković, Distant Reading Training School 2020: **Named Entity Recognition & Geo-Tagging for Literary Analysis**, [\[pdf\]](#)
- Cvetana Krstev, Ranka Stanković, **Novels and Authors of the Serbian ELTeC Collection**, [\[pdf\]](#)
- Cvetana Krstev, **Ideas and Observations from the Time of the ELTeC Corpus - a Selection of Quatations**, [\[pdf\]](#)

[Home](#) / [Archives](#) / Vol 21 No 2 (2021): Infotheca - Journal for Digital Humanities



The papers selected for this issue of the Infotheca journal present contributions of the Serbian team in the COST action CA16204 *Distant Reading for European Literary History*. The project started in 2017 with a plan to create a network of researchers which would develop resources and methods within the Distant Reading paradigm – usage of computational methods in the analysis of a large group of literary texts. One of the ultimate goals was to establish good, innovative methods and techniques of computational analysis of texts from this specific group. One of the most important achievements of this action is the development of a multilingual corpus, *European Literary Text Collection - ELTeC* which contains 100 novels first published in the period 1840-1920 for many European languages, including Serbian. All texts in this corpus were selected, prepared and annotated using the common principles.

Editors of this issue are Prof. Dr. Cvetana Krstev and Prof. Dr. Ranka Stanković





# Модели

- [SrpCNNER](#) – *Named Entity Recognizer for Serbian (7 classes)* – Препознавач именованих ентитета обучен да препознаје 7 типова именованих ентитета, са архитектуром Конволуцијске неуронске мреже (Convolutional Neural Network – CNN), са скором F1 од око 91% на тестном скупу података. Лиценца **CC-BY-NC-SA-4.0**.
  - Branislava Šandrih Todorović, Cvetana Krstev, Ranka Stanković, Milica Ikonić Nešić, “**Serbian NER& Beyond: The Archaic and the Modern Intertwinned**”, Proc. of the International Conference Recent Advances in Natural Language Processing - RANLP 2021, pp. 1252-1260, 2021.
- [SrpKor4Tagging-TreeTagger](#) – TreeTagger модел за обележавање скуповима ознака *Universal POS* и *SrpLemKor*, обучен коришћењем аотираног корпуса *SrpKor4Tagging* и речника *SrpMD4Tagging*.  
Лиценца **CC-BY-4.0**.
  - Ranka Stankovic, Branislava Šandrih, Cvetana Krstev, Miloš Utvić, and Mihailo Skoric. 2020. Machine Learning and Deep Neural Network-Based Lemmatization and Morphosyntactic Tagging for Serbian. LREC2020, Marseille, France.
- [SrpKor4Tagging-spaCy](#) – Модели *spaCy* за обележавање врстама речи коришћењем скупа ознака *Universal POS* и *SrpLemKor*, обучени коришћењем аотираног корпуса *SrpKor4Tagging*.  
Лиценца **CC-BY-4.0**.
  - Ranka Stanković, Mihailo Škorić, Branislava Šandrih Todorović, “Parallel Bidirectionally Pretrained Taggers as Feature Generators”, Applied Sciences (2022).

# Речници

- [SrpMD](#) – *Serbian Morphological Dictionaries* – Морфолошки речници за српски језик следе методологију и формат (познат као *DELAS/DELAf*) који су развијени у Лабораторији *LADL (Laboratoire d'Automatique Documentaire et Linguistique)*, 10.288 вишечланих речи, 88.753 речи и 3.753.750 флективних облика, лиценца **CC-BY-NC-SA-4.0**.
- [SrpMD4Tagging](#) – *Serbian Morphological Dictionaries for Tagging* – Морфолошки речници српског језика за обележавање изведени из Морфолошких речника српског језика (Krstev & Vitas) као референтни речници за придруживање леме и ознаке врсте речи облику речи пронађеном у тексту. Доступне су две датотеке са скупом ознака за обележавање врстама речи: *Univesal Dependencies* и традиционални скуп ознака *Serbian POS*, број анотираних речничких облика 935.466, лиценца **CC-BY-NC-SA-4.0**.
- [GeolISSTerm](#) – *Речник геолошких термина* је електронски речник у виду таксономије посебне намене основних геолошких појмова и термина. GeolISSTerm је део Геолошког информационог система Србије (GeolISS) и користи се за валидацију, класификовање и спецификовање уочених и интерпретираних геолошких атрибута. Садржи: 2.631 двојезичних термина са дефиницијама и синонимима, **CC-BY-NC-SA-4.0**.



# Векторске репрезентације речи

- Иницијалне верзије модела:
  - Word2vec (величина вектора 100)
  - Fasttext (величина вектора 100)

# Језички модели засновани на дубоком учењу

- Иницијалне верзије модела:
  - ГПТ2 (122 милиона параметара)

# Михаило Шкорић, Композитне псеудограматике засноване на паралелним језичким моделима српског језика, докторска дисертација

Неки од резултата:

<https://github.com/procesaur/Parallel-language-models>

<https://huggingface.co/procesaur/gpt2-srlat>

<https://huggingface.co/procesaur/gpt2-srlat-sem>

<https://huggingface.co/procesaur/gpt2-srlat-synt>

<https://plma.jerteh.rs/>

The screenshot displays the PLMA web application interface. The main area is divided into two columns. The left column contains input fields and controls for text processing and generation. The right column provides information about the application and visualization options.

**Input text to process or to use as context** (with a question mark icon)

Type here...

Select language model\*: procesaur/gpt2-srlat

**Text perplexity visualisation**

Slider window size: [Slider] Visualize

**Generation (GPT-2 only)** (with a Generate button)

Text length (tokens): [Slider] 10

Temperature (creativity): [Slider] 0.7

Number of outputs: [Slider] 3

Select second model (for evaluation)\*: huggingface model name

Full Evaluation (Serbian only)

\* Will download for local instance! Some models require additional preprocessing files!

**About PaLMA**

Application is developed in support of the University of Belgrade doctoral dissertation *Composite pseudogramms based on parallel language models of Serbian* by Mihailo Skoric, under the supervision of Ranka Stankovic.

Usage instructions are shown below for each of three usage scenarios:

**Perplexity visualisation**

requirements:

- Inputted text
- Selection of language model
- Slider size definition

Any kind of text is accepted as input. Inputted text will be truncated if too long!

Easiest way to select language model is to paste its huggingface id, eg. `procesaur/gpt2-srlat`. The model will be downloaded to a local computer if it is not already!

For visualisation language models like GPT and BERT (and perhaps other) should work. Selected model should ideally be applicable for the inputted text, eg. the language of the model and text should match.

Slider size should be set between 2 and 10, and it will define the size of the token-n-grams being evaluated using the model, in order to produce the final values. The shorter the text, the higher impact of this variable will be.

In so, the parameters

When to show, do as the browser do:

Select language model\*: gpt2

Text perplexity visualisation

Slider window size: [Slider] 3 Visualize

will output:

Text perplexity using gpt2 31.841

Generation

# Актуелна и будућа истраживања

- велики језички модели: BERT, нова верзија GPT-а
- унапређење тагирања врстама речи,
- унапређење анотација именованих ентитета
- анотација синтаксе
- континуалан рад на
  - прикупљању и обради текстова за корпус
  - специјализовани доменски корпуси
  - паралелизација текстова
  - електронски речници
  - лексичка база
  - терминологија
  - ...

[http://lloed.jerteh.rs/ELTEC/](http://llod.jerteh.rs/ELTEC/)

RDF формат kopnyca: NLP Interchange Format (NIF)

Pristupna tačka <http://fuseki.jerteh.rs/>

The screenshot shows the Apache Jena Fuseki web interface. The browser address bar displays `http://fuseki.jerteh.rs/#/dataset/SrpELTECtest2/query`. The page title is `/SrpELTECtest2`. Below the title, there are options for `query`, `add data`, `edit`, and `info`. A section for `SPARQL Query` includes a text area with a SPARQL query and a `run` button. Below the query, there are options for `Prefixes` (rdf, rdfs, owl, xsd), `JSON` content type, and `Turtle` content type. The results section shows a table with 12 rows of data.

subject	entity	etype
1 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=85870,85877>	сержан	<http://www.wikidata.org/entity/Q28640>
2 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=85453,85460>	сержан	<http://www.wikidata.org/entity/Q28640>
3 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=86198,86205>	сержан	<http://www.wikidata.org/entity/Q28640>
4 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=67554,67561>	Матрози	<http://www.wikidata.org/entity/Q28640>
5 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=66822,66829>	Матрози	<http://www.wikidata.org/entity/Q28640>
6 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=33038,33044>	Србији	<http://www.wikidata.org/entity/Q7884789>
7 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=1366,1376>	Карловица	<http://www.wikidata.org/entity/Q7884789>
8 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=1004,1014>	Карловица	<http://www.wikidata.org/entity/Q7884789>
9 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=81115,81121>	Младен	<http://www.wikidata.org/entity/Q5>
10 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=99165,99171>	Младен	<http://www.wikidata.org/entity/Q5>
11 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=98954,98960>	Младен	<http://www.wikidata.org/entity/Q5>
12 <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP18520.txt#char=82409,82415>	Младен	<http://www.wikidata.org/entity/Q5>

The screenshot shows the website `lloed.jerteh.rs - /ELTEC/srp/NIF/`. It features a list of text files with columns for date, time, and filename. The filenames are `SRP18520.txt` through `SRP18751.txt`.

Date	Time	Filename
3/17/2023	3:00 AM	SRP18520.txt
3/17/2023	2:59 AM	SRP18520.txt
3/17/2023	3:01 AM	SRP18590.txt
3/17/2023	3:01 AM	SRP18590.txt
3/17/2023	3:02 AM	SRP18620.txt
3/17/2023	3:02 AM	SRP18620.txt
3/17/2023	3:02 AM	SRP18630.txt
3/17/2023	3:03 AM	SRP18630.txt
3/17/2023	3:05 AM	SRP18631.txt
3/17/2023	3:03 AM	SRP18631.txt
3/17/2023	3:06 AM	SRP18680.txt
3/17/2023	3:05 AM	SRP18680.txt
3/17/2023	3:07 AM	SRP18690.txt
3/17/2023	3:06 AM	SRP18690.txt
3/17/2023	3:10 AM	SRP18691.txt
3/17/2023	3:08 AM	SRP18691.txt
3/17/2023	3:11 AM	SRP18730.txt
3/17/2023	3:10 AM	SRP18730.txt
3/17/2023	3:12 AM	SRP18740.txt
3/17/2023	3:11 AM	SRP18740.txt
3/17/2023	3:12 AM	SRP18741.txt
3/17/2023	3:12 AM	SRP18741.txt
3/17/2023	3:13 AM	SRP18750.txt
3/17/2023	3:12 AM	SRP18750.txt
3/17/2023	3:14 AM	SRP18751.txt
3/17/2023	3:13 AM	SRP18751.txt

```
<http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP19203.txt#char=14876,14880> a nif:RFC5147string,
    nif:Word ;
nif:anchorOf "шам"^^xsd:string ;
nif:beginIndex "14876"^^xsd:nonNegativeInteger ;
nif:endIndex "14880"^^xsd:nonNegativeInteger ;
nif:lemma "шам"^^xsd:string ;
nif:nextWord <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP19203.txt#char=14880,14881> ;
nif:postTag "VERB"^^xsd:string ;
nif:previousWord <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP19203.txt#char=14873,14875> ;
nif:referenceContext eltec:SRP19203.txt ;
nif:sentence <http://lloed.jerteh.rs/ELTEC/srp/NIF/SRP19203.txt#char=14871,14994> .
```

# RDF лексички ресурси следе Ontolex-Lemon модел

The screenshot shows the Apache Jena Fuseki web interface. The browser address bar shows the URL: `fuseki.jeriteh.rs/#dataset/Senti-Pol-sr/query`. The page title is `/Senti-Pol-sr`. There are buttons for `query`, `add data`, `edit`, and `info`. Below these, there is a section for `SPARQL Query` with instructions: "To try out some SPARQL queries against the selected dataset, enter your query here." There are also buttons for `Selection of triples` and `Selection of classes`. The `SPARQL Endpoint` is `/Senti-Pol-sr/`. The `Content Type (SELECT)` is `JSON` and the `Content Type (GRAPH)` is `Turtle`. A SPARQL query is entered in the text area: 

```
1 SELECT ?subJect ?predicate ?objJect
2 WHERE {
3   ?subJect ?predicate ?objJect
4 }
5 LIMIT 25
```

 Below the query, there is a table with 25 results. The table has three columns: `subject`, `predicate`, and `obje`. The first few rows are: 

subject	predicate	obje
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Advent-n-0&gt;</code>	<code>&lt;http://www.lemon-model.net/lemon#lexicalForm&gt;</code>	<code>&lt;http://www.lemon-model.net/lemon#writtenRep "iznad"@sr&gt;</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Advent-n-0&gt;</code>	<code>&lt;http://www.lexinfo.net/ontology/2.0/lexinfo#partOfSpeech&gt;</code>	<code>"N"</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Advent-n-0&gt;</code>	<code>&lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#rType&gt;</code>	<code>&lt;http://www.w3.org/2000/01/rdf-schema#label&gt;</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Advent-n-0&gt;</code>	<code>&lt;http://www.w3.org/ns/lemon/ontolex#sense&gt;</code>	<code>"Adv"</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Advent-n-0-form&gt;</code>	<code>&lt;http://www.lemon-model.net/lemon#writtenRep&gt;</code>	<code>"Adv"</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Advent-n-0-sense1&gt;</code>	<code>&lt;http://www.gsi.dit.upm.es/ontologies/mar/lex#hasPolarity&gt;</code>	<code>hasP</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Advent-n-0-sense1&gt;</code>	<code>&lt;http://www.gsi.dit.upm.es/ontologies/mar/lex#hasValue&gt;</code>	<code>hasV</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Apostol-n-0&gt;</code>	<code>&lt;http://www.lemon-model.net/lemon#lexicalForm&gt;</code>	<code>&lt;http://www.lemon-model.net/lemon#writtenRep "iznutrice"@sr&gt;</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Apostol-n-0&gt;</code>	<code>&lt;http://www.lexinfo.net/ontology/2.0/lexinfo#partOfSpeech&gt;</code>	<code>"N"</code>
<code>&lt;http://lod.jeriteh.rs/SA/Senti-Pol-sr#Apostol-n-0&gt;</code>	<code>&lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#rType&gt;</code>	<code>&lt;http://www.w3.org/2000/01/rdf-schema#label&gt;</code>

```
jeriteh:iznad-prep-0 lemon:lexicalForm jeriteh:iznad-prep-0-form;
lexinfo:partOfSpeech "PREP"@sr;
a lemon:LexicalEntry;
rdfs:label "iznad"@sr;
ontolex:sense jeriteh:iznad-prep-0-sense1.
jeriteh:iznad-prep-0-form lemon:writtenRep "iznad"@sr.
jeriteh:iznad-prep-0-sense1 marl:hasPolarity "hasPolarity:Positive";
marl:hasValue "hasValue:1".
jeriteh:iznenada-adv-0 lemon:lexicalForm jeriteh:iznenada-adv-0-form;
lexinfo:partOfSpeech "ADV"@sr;
a lemon:LexicalEntry;
rdfs:label "iznenada"@sr;
ontolex:sense jeriteh:iznenada-adv-0-sense1.
jeriteh:iznenada-adv-0-form lemon:writtenRep "iznenada"@sr.
jeriteh:iznenada-adv-0-sense1 marl:hasPolarity "hasPolarity:Negative";
marl:hasValue "hasValue:-1".
jeriteh:iznenadxenxe-n-0 lemon:lexicalForm jeriteh:iznenadxenxe-n-0-form;
lexinfo:partOfSpeech "N"@sr;
a lemon:LexicalEntry;
rdfs:label "iznenadenje"@sr;
ontolex:sense jeriteh:iznenadxenxe-n-0-sense1.
jeriteh:iznenadxenxe-n-0-form lemon:writtenRep "iznenadenje"@sr.
jeriteh:iznenadxenxe-n-0-sense1 marl:hasPolarity "hasPolarity:Positive";
marl:hasValue "hasValue:1".
jeriteh:iznenadxujukxe-adv-0 lemon:lexicalForm jeriteh:iznenadxujukxe-adv-0-form;
lexinfo:partOfSpeech "ADV"@sr;
a lemon:LexicalEntry;
rdfs:label "iznenadujuće"@sr;
ontolex:sense jeriteh:iznenadxujukxe-adv-0-sense1.
jeriteh:iznenadxujukxe-adv-0-form lemon:writtenRep "iznenadujuće"@sr.
jeriteh:iznenadxujukxe-adv-0-sense1 marl:hasPolarity "hasPolarity:Positive";
marl:hasValue "hasValue:1".
jeriteh:iznuren-a-0 lemon:lexicalForm jeriteh:iznuren-a-0-form;
lexinfo:partOfSpeech "A"@sr;
a lemon:LexicalEntry;
rdfs:label "iznuren"@sr;
ontolex:sense jeriteh:iznuren-a-0-sense1.
jeriteh:iznuren-a-0-form lemon:writtenRep "iznuren"@sr.
jeriteh:iznuren-a-0-sense1 marl:hasPolarity "hasPolarity:Negative";
marl:hasValue "hasValue:-1".
jeriteh:iznutrice-n-0 lemon:lexicalForm jeriteh:iznutrice-n-0-form;
lexinfo:partOfSpeech "N"@sr;
a lemon:LexicalEntry;
rdfs:label "iznutrice"@sr;
ontolex:sense jeriteh:iznutrice-n-0-sense1.
jeriteh:iznutrice-n-0-form lemon:writtenRep "iznutrice"@sr.
jeriteh:iznutrice-n-0-sense1 marl:hasPolarity "hasPolarity:Negative";
marl:hasValue "hasValue:-1".
```

# Википодаци за СрпКор

---

После успешне викификације SrpELTeC корпуса следи повезивање романа који се налазе у корпусу српског - то су углавном најзначајнија дела српске књижевности

---

подаци о делу и његовим издањима

---

подаци о писцу

---

подаци о главним ликовима,

---

подаци о везама ликова

---

места радње

---

важни догађаји

---

# Планови

Отворени смо за:

- заједничко учешће на научним пројектима
- комерцијалне пројекте за језичку индустрију
- практичне пројекте које укључују коришћење језичких ресурса и технологија

