

СРПСКА АКАДЕМИЈА НАУКА И УМЕТНОСТИ
Циклус предавања: ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА

Предавање

**ЈЕЗИЧКЕ ТЕХНОЛОГИЈЕ КАО КЉУЧНИ ФАКТОР ВЕШТАЧКЕ ИНТЕЛИГЕНЦИЈЕ:
ПРОШЛОСТ, САДАШЊОСТ И БУДУЋНОСТ**

које ће одржати

ПРОФ. ДР РАНКА СТАНКОВИЋ

Рударско-геолошки факултет, Универзитет у Београду

Уторак, 3. октобар 2023. године у 11.00 сати, Свечана сала САНУ, Кнеза Михаила 35/II, Београд

Језичке технологије као кључни фактор вештачке интелигенције: прошлост, садашњост и будућност

проф. др Ранка Станковић



Универзитет у Београду, Рударско-геолошки факултет

Друштво за језичке ресурсе и технологије - ЈеРТех



Увод

Језичке технологије играју кључну улогу у развоју ВИ и имају богату прошлост, значајну садашњост и узбудљиву будућност.

На МП се ради практично од настанка првих рачунара, а касније је уследио напредак у анализи текста и гласовној комуникацији.

У двадестом веку су се смењивале парадигме засноване на статистичким моделима и правилима.

Шта је доступно данас?

У данашње време ЈТ су достигле значајан ниво напретка захваљујући машинском и дубоком учењу:

- развијене су методе анализе и обележавања засноване на неуронским мрежама.

Ове технике омогућавају моделима да боље обрађују и трансформишу текст на природном језику.

Реч-две о терминима

Сродни термини

- Обрада природних језика (ОПЈ) – *Natural language Processing (NLP)*
- Рачунарска лингвистика – *Computational Linguistics (CL)*
- Рачунарска обрада језика – *Computational Language Processing*
- **Језичке технологије – *Language Technology***
- Технологија за обраду људског/природног језика – *Human Language Technology*
- Језичко инжењерство – *Language Engineering*
- Анализа текста – *Text Analysis*
- Ископавање (информација) из текста – *Text Mining*



EUROPEAN
LANGUAGE
GRID

European Language Grid

- ... provides access to **Language Technology** resources from all over Europe.
- ... contains **tools** and **services**, **language resources** and information on European LT companies and research organisations as well as their projects.

кровни термин који бирамо: **Језички ресурси и технологије**

Шта су NLP, NLU, NLG у свету језичких технологија?

Natural language processing (NLP)

- ОПЈ конвертује неструктуриране језичке податке у формат структурираних података како би машине могле да обраде говорни и писани текст и формулишу одговоре.
- ОПЈ обухвата обраду, разумевање и генерисање природног језика.

Natural language understanding (NLU)

- Разумевање природних језика усмерено је на разумевање машински прочитаног текста кроз анализу граматике и контекста, са циљем да се одреди значење реченице.

Natural language generation (NLG)

- Генерисање природног језика се фокусира на генерисање текста, или конструкцију текста на (природним) језицима, на основу задатог улазног текста.

Обрада природних језика

NLP

Natural Language Processing

обележавање
врста речи

лематизација

препознавање
именованих
ентитета

синтаксичка
анализа

категоризација
текста

конвертовање
текста у говор

резолюција
(разрешавање)
кореференци

аутоматско
препознавање
говора

Разумевање природних језика

NLU

Natural Language Understanding

екстракција
релација

семантичка
анализа

анализа
осећања
(емоција,
мишљења,
сентимента)

текст у слику
слика у текст

Генерисање природних језика

NLG

Natural Language Generation

парафразирање и
закључивање на
природном језику

одговарање на питања

машинско превођење

сумаризација
(сажимање)

планирање текста

трансформација података
у текст

реализација реченица

агенти за дијалог
(чет ботови)

- 1 Данас је 3. октобар.
- 2 Налазимо се у Свечаној сали САНУ.

1 Данас је 3. октобар .

ADV данас
 VERB јесам
 NUM 3.
 NOUN октобар
 PUNCT .

2 Налазимо се у Свечаној сали САНУ .

VERB налазити
 PART се
 ADP у
 ADJ свечан
 NOUN сала
 PROPN САНУ
 PUNCT .

ID:4487
 :idata.org/entity/Q1132636
 es and Arts
 demic institution in Serbia

1 Данас је 3. октобар

ADV данас
 VERB јесам
 NOUN октобар
 DATE 3.

1 Данас је 3. октобар.

2 Налазимо се у Свечаној сали САНУ .

Serbian Academy of Sciences and Arts | ORG



Језичке технологије

За “велике” светске језике доступне су моћне апликације за:

- аутоматско превођење
- генерисање текста
- одговарање на питања
- анализу ставова и осећања исказаних у тексту
- екстракцију информација
 - издвајање релевантних података из неструктурираних извора података,
 - мерење задовољства корисника
- класификација и карактеризација текста
 - препознавање говора мржње
- и друге.

Језичке технологије

Са порастом употребе чет ботова и виртуелних асистената, потреба да машине разумеју природни језик постала је пресуднија:

- први корак укључује претходну обраду текстуалних података како би се припремили за анализу
- токенизација
- нормализација текста, одређивање граматичких информација
- препознавање ентитета
- извлачење значења из текста:
 - препознавање намере
 - идентификација циља у основи датог текста

Пример корисничког захтева:

- „Какво ће време бити сутра у Београду?“
 - идентификује се намера корисника да добије информације о времену и издвоји параметар „сутра“, пронађе локацију и одговарајућу временску прогнозу.

Позиција српског језика

Српски језик се упркос заостатку бори да достигне развијеност ресурса и алата “великих” језика.

ЈТ имају велики потенцијал за будућност ВИ очекује се напредак у области :

- превођења (нарочито за специфичне домене)
- генерисања текста са могућношћу одговарања на питања
- гласовне комуникацију са уређајима.

Напредак се очекује и у разумевању

- значења текста у контексту
- прилагођавању система за
 - употребу у различитим ситуацијама и
 - за различите профиле корисника (узраст, стручност, интересовања)

Реч-две о историји

четрдесете године XX века:

- војска заинтересована за двојезичне речнике и МП за разбијање шифара током рата.

педесете године XX века:

- Алан Тјуринг: „Рачунарска машина и интелигенција“ који чини основу Тјуринговог теста.
- Ноам Чомски: „Синтаксичке структуре“ дају основу правила која су рачунарски системи могли да имплементирају.

шездесете године XX века:

- SHRDLU, MIT, Тери Виноград: корисници могли да комуницирају с рачунарима, уз бројна ограничења, али то је била одскачна даска у правом смеру.
- ELIZA, MIT, Визембаум: „психотерапеут“ који је „разговарао“ користећи информације које су дали корисници и нудио одговоре на основу програмираних.

седамдесете године XX века:

- Р. Шанк, Станфорд, теорија концептуалне зависности за разумевање природних језика која се користи у системима вештачке интелигенције.
- В. Вудс: АТН заснован на коначним аутоматима, преводио је информације из стварног света у податке које је рачунар разумео (четбот **PARRY**).

осамдесете године XX века:

- Након обраде засноване на правилима, уз повећање рачунарских (хардверских) могућности почиње коришћење машинског учења за разумевање задатог текста.

деведесете године XX века:

- IBM предњачи у истраживању, све већи успех система машинског учења, креирање модела који побољшавају комуникацију између човека и рачунара.

прва деценија XXI века:

- IBM систем **Ватсон** 2006: заснован на ВИ који одговара на питања; победио у такмичењу *Jeopardy!* против најбољих играча квиза у 2011.

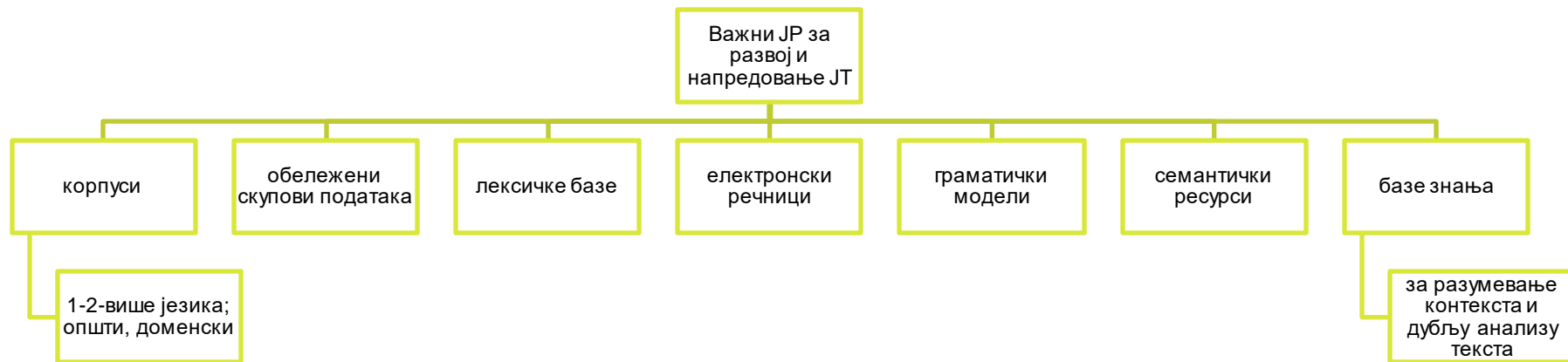
друга деценија XXI века:

- Сири (2011) и Алекса (2014): лични асистенти који користе интерфејс на природном језику за одговарање на питања, давање препорука и инструкција

трећа деценија XXI века:

- Све више предузећа усваја ботове у пословању, а са развојем технологија, верује се да ће се све више окретати ка клијентима (купцима роба и услуга).

Језички ресурси



Начин креирања

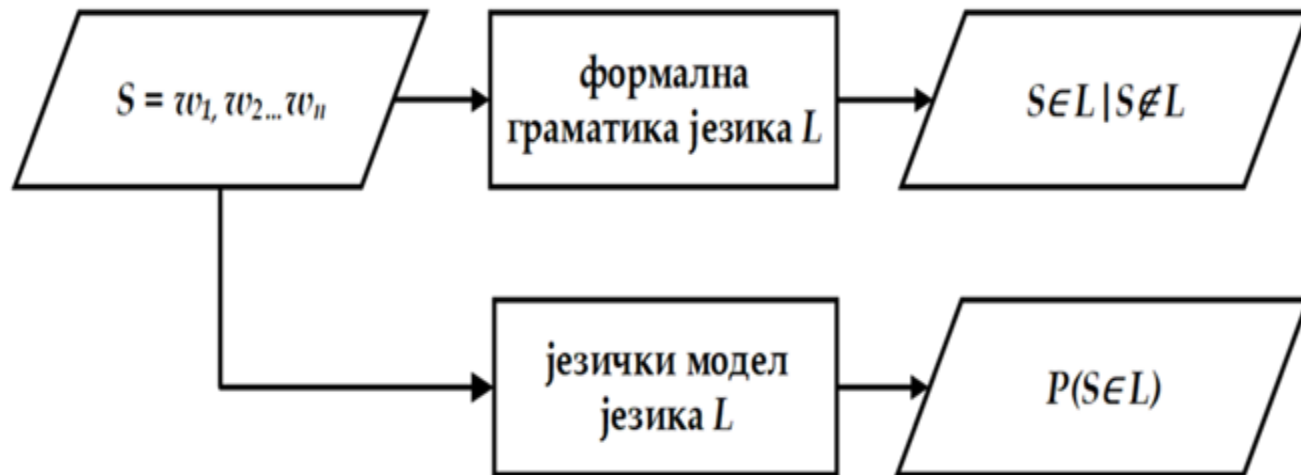
Раније

- су ресурси били мањег обима и углавном ручно креирани.
- за обучавање модела коришћени су мањи корпуси различитих домена и регистра.
- традиционални речници и енциклопедије трансформисани су у лексичке базе података да би се могли користити у системима заснованим на обради језика.

Данас

- захваљујући интернету, постале су доступне огромне количине језичких ресурса
- велики корпуси који се могу користити за обучавање и евалуацију језичких модела.

Однос граматика и модела?

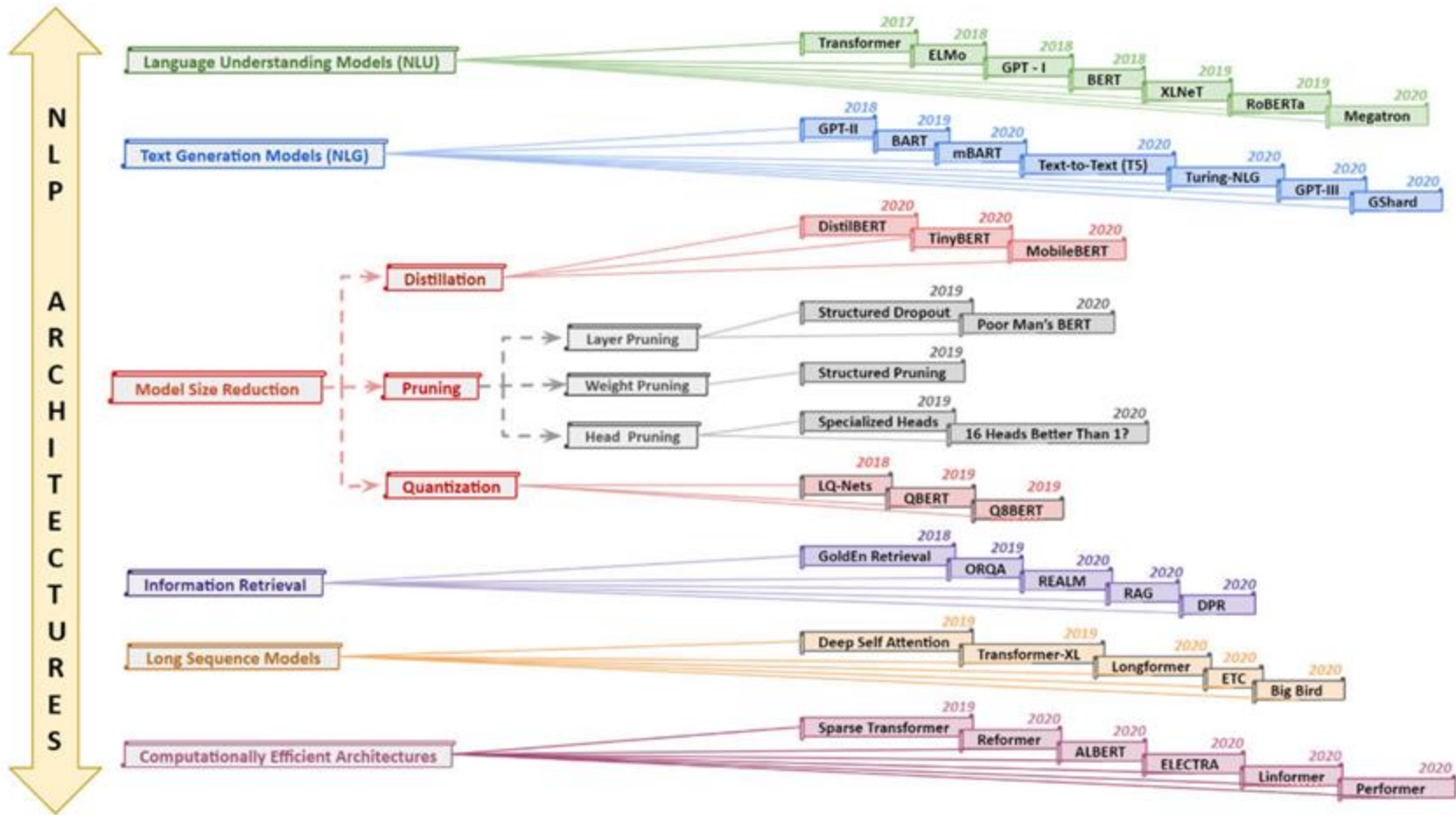


„Композитне псеудограматике засноване на паралелним језичким моделима српског језика“ – др Михаило Шкорић [\[pdf\]](#)

Однос граматика и модела?



„Композитне псевдограматике засноване на паралелним језичким моделима српског језика“ – др Михаило Шкорић [\[pdf\]](#)



Izvor: Singh, S. and Mahmood, A. "The NLP cookbook: modern recipes for transformer based deep learning architectures." *IEEE Access* 9 (2021): 68675-68702. [PDF](#)

Важност језичких ресурса

ЈР омогућавају развој нових апликација и модела, и незаменљиви за неке задатке.

Развој ЈТ ће се у будућности заснивати на великим и богатим језичким ресурсима.

Напредни алгоритми и модели захтевају велике, репрезентативне и разноврсне корпусе.

Развој бољих метода за аутоматску анотацију и обогаћивање корпуса различитим информацијама.

Чланови Друштва ЈеРТех већ више деценија развијају различите ресурсе и технологије за српски језик

JePTex: ресурси и технологије за српски језик

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.

Ресурси семантичког веба.

Ресурси

Међу ресурсима које Друштво развија и одржава издвајају се:

- једнојезични и вишејезични корпуси
 - платформа No Sketch Engine <https://noske.jerteh.rs/>
 - Библиша, <http://biblisha.jerteh.rs/>
 - Текстометрија [ТХМ](#)
- електронски речници
 - [Unitex](#), [Leximirka](#)
- Терминолошке базе
 - [Терми](#), [GeollSSTerm](#)
- повезани лингвистички отворени подаци.

Корпуси

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, ... а и српско-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		
Укупно	964.054.672	808.528.712	617.198	35.732.944	2.934.786

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
Укупно	674.486.541	1.916	100,0

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
Укупно	878.388.177	616.535	100

(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 INTELIgENCJE	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"/> roboŭ	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: 1-20 of 377

[<](#) [<](#) [1](#) / 19 [>](#) [>](#)

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

SrpKor2021

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

	Lemma	Frequency	Relative [?]
1	<input type="checkbox"/> veštački inteligencija	1,307	1.82
2	<input type="checkbox"/> pošten inteligencija	269	0.38
3	<input type="checkbox"/> srpski inteligencija	256	0.36
4	<input type="checkbox"/> emocionalan inteligencija	200	0.28
5	<input type="checkbox"/> misionarski inteligencija	120	0.17
6	<input type="checkbox"/> poslovan inteligencija	110	0.15
7	<input type="checkbox"/> kritičko inteligencija	99	0.14
8	<input type="checkbox"/> tehnički inteligencija	92	0.13
9	<input type="checkbox"/> socijalan inteligencija	86	0.12
10	<input type="checkbox"/> ljudski inteligencija	82	0.11
11	<input type="checkbox"/> nacionalan inteligencija	79	0.11
12	<input type="checkbox"/> liberalan inteligencija	62	0.09
13	<input type="checkbox"/> visok inteligencija	59	0.08
14	<input type="checkbox"/> natprosečan inteligencija	57	0.08
15	<input type="checkbox"/> politički inteligencija	55	0.08
16	<input type="checkbox"/> koeficijent inteligencija	53	0.07
17	<input type="checkbox"/> ruski inteligencija	50	0.07
18	<input type="checkbox"/> prosečan inteligencija	44	0.06
19	<input type="checkbox"/> elementaran inteligencija	43	0.06
20	<input type="checkbox"/> kolektivan inteligencija	41	0.06

SrpKor2013

(603 items, 1,149 total frequency)

	Lemma	Frequency	Relative [?]
1	<input type="checkbox"/> veštački inteligencija	94	0.65
2	<input type="checkbox"/> srpski inteligencija	58	0.40
3	<input type="checkbox"/> emocionalan inteligencija	32	0.22
4	<input type="checkbox"/> socijalan inteligencija	29	0.20
5	<input type="checkbox"/> pošten inteligencija	27	0.19
6	<input type="checkbox"/> stvaralački inteligencija	23	0.16
7	<input type="checkbox"/> ljudski inteligencija	21	0.14
8	<input type="checkbox"/> koeficijent inteligencija	21	0.14
9	<input type="checkbox"/> visok inteligencija	19	0.13
10	<input type="checkbox"/> test inteligencija	17	0.12
11	<input type="checkbox"/> velik inteligencija	15	0.10
12	<input type="checkbox"/> koeficijentom inteligencija	13	0.09
13	<input type="checkbox"/> francuski inteligencija	13	0.09
14	<input type="checkbox"/> kritički inteligencija	13	0.09
15	<input type="checkbox"/> evropski inteligencija	13	0.09
16	<input type="checkbox"/> ruski inteligencija	12	0.08
17	<input type="checkbox"/> srpski stvaralački inteligenc	12	0.08
18	<input type="checkbox"/> tehnički inteligencija	10	0.07
19	<input type="checkbox"/> natprosečan inteligencija	10	0.07
20	<input type="checkbox"/> gradanski inteligencija	10	0.07

SrpELTeC

(26 items, 33 total frequency)

	Lemma	Frequency	Relative [?]
1	<input type="checkbox"/> slab inteligencija	3	0.51
2	<input type="checkbox"/> viša inteligencija	3	0.51
3	<input type="checkbox"/> seoski inteligencija	2	0.34
4	<input type="checkbox"/> visok inteligencija	2	0.34
5	<input type="checkbox"/> srpski inteligencija	2	0.34
6	<input type="checkbox"/> narodni inteligencija	1	0.17
7	<input type="checkbox"/> prirodan inteligencija	1	0.17
8	<input type="checkbox"/> svet inteligencija	1	0.17
9	<input type="checkbox"/> energiju brajkovački inteligencija	1	0.17
10	<input type="checkbox"/> borila inteligencija	1	0.17
11	<input type="checkbox"/> svedok inteligencija	1	0.17
12	<input type="checkbox"/> velik inteligencija	1	0.17
13	<input type="checkbox"/> varoški inteligencija	1	0.17
14	<input type="checkbox"/> jak muzički inteligencija	1	0.17
15	<input type="checkbox"/> oskuđevala prirodan inteligencija	1	0.17
16	<input type="checkbox"/> brajkovački inteligencija	1	0.17
17	<input type="checkbox"/> nov inteligencija	1	0.17
18	<input type="checkbox"/> sakrovište inteligencija	1	0.17
19	<input type="checkbox"/> korisnu inteligencija	1	0.17
20	<input type="checkbox"/> ceo sakrovište inteligencija	1	0.17

Kapetan kraljeve garde Žarko Belimarković
stupa u „Ujedinjenje ili smrt“.

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

SrpELTeC-NERfull

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

CONCORDANCE

DOWNLOAD

Download first _____ rows

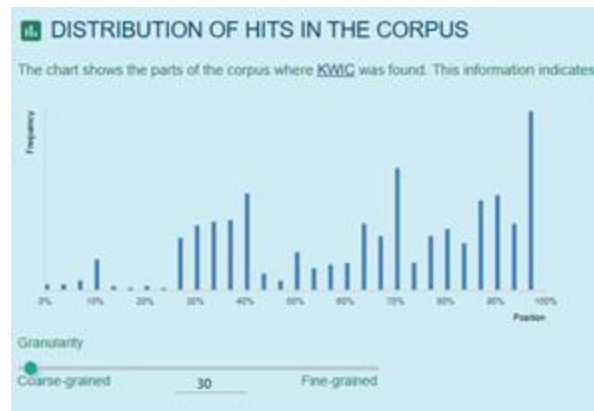
Or save the current view as

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

Frequency

Show frequency per million

	Lemma	↓ Frequency	Frequency per million	
61	komandant Ugrinović	2	2.12	***
62	kapetan krajev garda Žarko Belmarković	2	2.12	***
63	kapetan krajev garda Žarko	2	2.12	***
64	kapetan Pribičević	2	2.12	***
65	harmonikašica Mara	2	2.12	***
66	harambaša Marinka	2	2.12	***
67	gospodica Amalija Bertoldina	2	2.12	***



Претрага паралелног корпуса

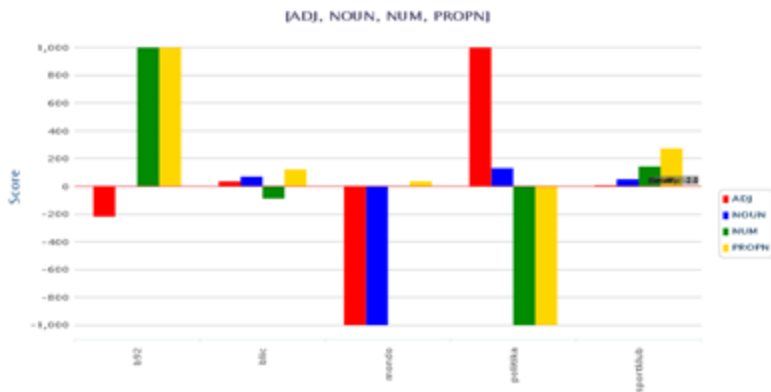
`[tag="V.*"][]{{,2}}[lemma="zakon"] within </>`

The screenshot displays the PARALLEL CONCORDANCE interface. At the top, the search query is entered: `simple [tag="V.*"][]{{,2}}[lemma="zakon"] within </>`. The results show 564 matches (305.38 per million) with no translations found. The interface is split into two columns: BiKes_sr (Serbian) on the left and BiKes_en (English) on the right. The search results are presented in a table with 10 rows. Each row contains a snippet from the Serbian text and its corresponding English translation. The search results are filtered to show only instances where the tag `</>` is used, and the lemma is `zakon`. The results are sorted by relevance, with the most relevant results appearing first. The interface includes a search bar, a list of filters, and a pagination control at the bottom.

BiKes_sr	BiKes_en
<p></> Odredbe o utvrđivanju poreza primenjuju se i na utvrđivanje sporednih poreskih davanja , osim ako je ovim zakonom drukčije propisano . </></p>	<p></> The provisions pertaining to the tax assessment shall be applied to the assessment of secondary tax duties , unless otherwise provided for by this Law . </></p>
<p></> Odredbe o naplati poreza primenjuju se i na naplatu sporednih poreskih davanja , osim ako je ovim zakonom drukčije propisano . </></p>	<p></> Provisions pertaining to the tax collections shall be applied to the collection of secondary tax duties , unless otherwise provided for by this law . </></p>
<p></> U postupku prinudne naplate poreza Poreska uprava primenjuje radnje uređene ovim zakonom , vodeći računa o ekonomičnosti postupka . </></p>	<p></> In the enforced collection procedure , the Tax Administration shall apply the actions regulated by this Law , taking into account the financial worth of the procedure . </></p>
<p></> Rešenje iz stava 1. ovog člana izvršava se na način uređen Zakonom o platnom prometu iz člana 69. stav 2. ovog zakona . </></p>	<p></> The decision indicated in paragraph 1 of this Article shall be implemented in the manner regulated by the Law on Payment Operations from Article 69 , paragraph 2 of this Law . </></p>
<p></> Na zaštitu kupca i njegovih prava , kao i na sve ostalo što nije posebno uređeno ovim zakonom u vezi sa prodajom nepokretnosti , shodno se primenjuju odredbe zakona kojim se uređuje izvršni postupak . </></p>	<p></> To the protection of the buyer and his rights , as well as to all other matters not specially regulated by this Law regarding the sale of real estate , the provisions of the Law regulating executive procedures shall duly be applied . </></p>
<p></> Na zaštitu kupca i njegovih prava , kao i na sve ostalo što nije posebno uređeno ovim zakonom u vezi sa prodajom nepokretnosti , shodno se primenjuju odredbe zakona kojim se uređuje izvršni postupak . </></p>	<p></> To the protection of the buyer and his rights , as well as to all other matters not specially regulated by this Law regarding the sale of real estate , the provisions of the Law regulating executive procedures shall duly be applied . </></p>
<p></> Rešenjem iz stava 1. ovog člana nalaže se poreskom obvezniku da u roku određenom rešenjem otkloni utvrđene povrede zakona , odnosno nepravilnosti u primeni propisa . </></p>	<p></> By the decision indicated in paragraph 1 of this Article , the taxpayer shall be instructed to remove the determined violations of the law or irregularities in the application of regulations within the term stated in the decision . </></p>
<p></> Ako drugostepeni poreski organ nađe da je prvostepeni poreski upravni akt zasnovan na zakonu , ali zbog drugih razloga , a ne zbog onih navedenih u aktu , on će u svom rešenju navesti te razloge , a žalbu odbiti . </></p>	<p></> When the second-level tax authority finds that the first-level tax administrative act is based on the law , but due to other reasons , and not those stated in the act , these reasons shall be stated in the ruling , and the appeal shall be rejected . </></p>
<p></> Postupak za izricanje novčanih kazni i zaštitnih mera za poreske prekršaje Poreska uprava vodi po zakonu kojim se uređuje prekršajni postupak , ako ovim zakonom nije drukčije propisano . </></p>	<p></> The procedures for sentencing of fines and precautionary measures for tax violations shall be conducted by the Tax Administration according to the Law by which the violations procedure is regulated , if not otherwise regulated by this Law . </></p>

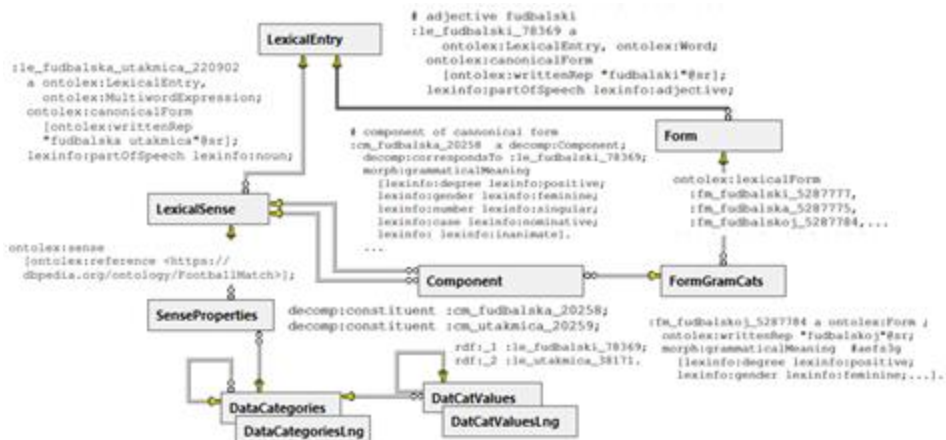
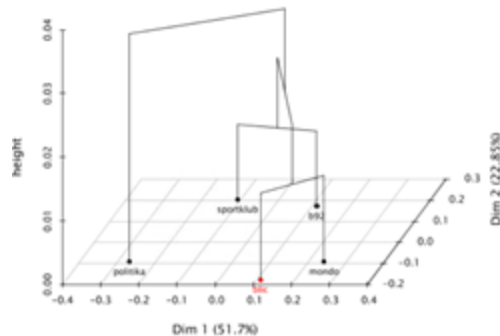
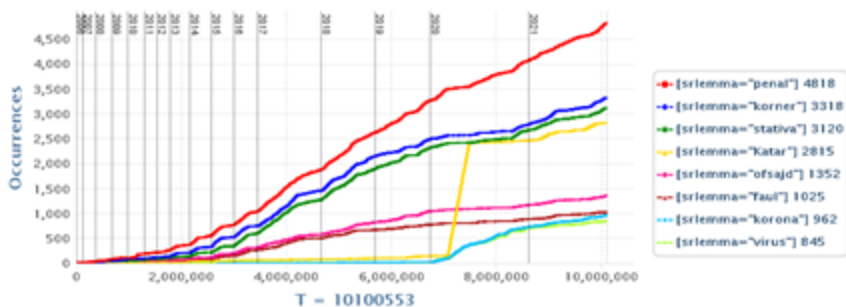
Rows per page: 10 31-40 of 564 4

Текстометрија на корпусу фудбала



Progression of [srlemma="penal"], [srlemma="korner"], [srlemma="stativa"], [srlemma="Katar"], [srlemma="ofsajd"], [srlemma="faul"], [srlemma="korona"], [srlemma="virus"] in SRFUDKOR

(structure: text, property: godina, filtering REGEX:)



Jelena Lazarević, Ranka Stanković, Mihailo Škorić, Biljana Rujević,
Football terminology: compilation and transformation into OntoLex-Lemon resource,
 TermTrends 2023: Terminology in the Era of Linguistic Data Science, Vienna, Austria,
 2023.

Апликације

Од развијених апликација издвајамо:

- апликације за претраживање колекције паралелних текстова
- апликације за управљање лексичком базом
- апликације за развој терминолошких ресурса
- апликације за екстракцију терминологије
- скуп веб сервиса за проширење упита за различите намене
- сервисе за аотацију текстова именованим ентитетима
- апликације за аотацију врста речи и лематизацију
- ...

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

BIBLIŠA: ALIGNED COLLECTION SEARCH TOOL [Log In] [Register]

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

WELCOME TO ALIGNED TEXT COLLECTION SEARCH TOOL!

Keyword **veštačka inteligencija** sr Text collection **All**

Synonyms	en	sr
<input checked="" type="checkbox"/> WordNet	AI,artificial intelligence	AI,veštačka inteligencija
<input checked="" type="checkbox"/> Dictionary of Librarianship	S: artificial intelligence - AI,AI - artificial	veštačka inteligencija
<input checked="" type="checkbox"/> Biblimir		
<input checked="" type="checkbox"/> GeolISSTerm		veštačka inteligencija
<input checked="" type="checkbox"/> RudOnto		veštačka inteligencija
<input checked="" type="checkbox"/> Termi		veštačka inteligencija

Include Hypernyms Include Hyponyms **Preview and modify terms for**

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

Morphological query expansion **Obtain concordance** **New Search**

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

Buridans Esel

SrpNemKor, Roman [pdf]

ID: 11.1.005 Number: 1 Volume: 1 Month: 1 Year: 1 UDC: 825.112.2-31 [bom] [bwl]

Günter de Bruyn [VIAF] [GND] [WikiData] [LCN]

Abstract
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 gefasste Metapher für die Erzählung eines Dreieckskonflikts.

Pages: 199

Publishing place: Frankfurt am Main
Publisher: Fischer-Taschenbuch
Publishing year: 1992

Translator:

Buridans magarac

SrpNemKor, roman [VIAF] [GND] [WikiData]

ID: 11.1.005 Number: 1 Volume: 1 Month: 1 Year: 1 UDC: 825.112.2-31 [bom] [bwl]

Günter de Bruyn [VIAF] [GND] [WikiData] [LCN]

Apstrakt
De Brojnovi juneci dobivaju svoje radost muku u Istočnom Berlinu sedeseth godina pozadini realizacije kao da se još jebe boje autorove bogate paleta. De Brojn nas rukom provodi kroz mizantropijsku palnu ogledalo i izmamljujući nam otmeha koji ne Ovo je prvi roman Gintara de Brojna koji je prevodu na srpski.

Strane: 272

Mesto izdanja: Beograd
Izdavač: Fabrika knjiga
Godina izdanja: 2007

Prevodilac: s nemackog previla Aleksand Vuben

BURDANS ESEL / GUNTER DE BRUYN = BURIDANOV MAGARAC / GINTER DE BROJN, ID: 11.1.005 [About]

En/De/Hr/Sr: (first 9 out of 2890 sentences) [pdf]	Srpski - (prvih 9 od 2890 rečenica) [pdf]
n1 Günter de Bruyn	n1 Günter de Brojn
n2 Buridans Esel: Roman	n2 BURIDANOV MAGARAC
n3 Anfängen hat es so:	n3 Evo kako je sve počelo:
n4 Karl Erp lächelte beim Erwachen und wußte nicht, warum.	n4 Karl Erp se probudio nasmijan 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 Fraulein Broder ein.	n6 Tek kasnije, ne baš mnogo kasnije, ali ipak kasnije, setio se gospođice Broder.
n7 Sagen wir lieber:	n7 Ili bolje je kažemo ovako:
n8 nachträglich schien ihm, daß er an jenem Morgen lächelnd erwacht sei.	n8 naknadno mu se učinilo da se tog jutra probudio nasmijan.
n9 [Und so erzählte er es Fraulein Broder auch]	n9 (Pa je to tako i ispričao gospođi Broder:

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

Јелена Андоновски, 2020
 Мрежа отворених података и је зички ресурси у пропе су
 изградње српско-не мачког литерарног корпуса

Бурдана магарац	roman	СрпNemKor издања	✓ избор
дане дане или имена	српски јазик	немање јазик	
датум издавања	2019	немање јазик	
место издавања	Београд	немање јазик	
+ 0 референца			• додати референца
Бурдана магарац	roman (издање из 1992. издање)	✓ избор	
дане дане или имена	српски јазик		
датум издавања	1992		
место издавања	Франкфурт на Мајни		
+ 0 референца			• додати референца
Бурдана магарац	roman (издање из 2007. издање)	✓ избор	
дане дане или имена	српски јазик		
датум издавања	2007		
место издавања	Београд		
+ 0 референца			• додати референца

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: polileksemska jedinica

Is composed of:

Form	Lemma	FST Code	Gram Cat	Separator
veštačka	veštački	A2	æfs1g	
inteligencija	inteligencija	N600	fs1q	

https://leximirka.jerteh.rs/LexicalEntry/Delaf?id=216789 - Profil 1 - Microsoft Edge

https://leximirka.jerteh.rs/LexicalEntry/Delaf?id=216789

veštačka inteligencija,veštačka(veštački.A2:æfs1g) inteligencija(inteligencija.N600:fs1q).NC:fs1q

veštačka inteligencijo,veštačka(veštački.A2:æfs1g) inteligencija(inteligencija.N600:fs1q).NC:fs5q

veštačke inteligencije,veštačka(veštački.A2:æfs1g) inteligencija(inteligencija.N600:fs1q).NC:fs2q

veštačkoj inteligenciji,veštačka(veštački.A2:æfs1g) inteligencija(inteligencija.N600:fs1q).NC:fs3q:fs7q

veštačkom inteligencijom,veštačka(veštački.A2:æfs1g) inteligencija(inteligencija.N600:fs1q).NC:fs6q

veštačku inteligenciju,veštačka(veštački.A2:æfs1g) 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 Iemon 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-lm.dic

NOUN

Relations: To *neinteligencija* using *negacija* [L_{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 Termi Glosbi

- Frequencies:
- Top 10000 most frequent in SrbCorp122M Corpus by D. Vitas, M. Ubrvić (16.68 per million)
 - Top 10000 most frequent in maletalka Corpus by M. Radojčić (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 (3.00 per million)

Search corpora: SrpKor2021 A(N) Concordanca Form Frequencies Lemma Frequencies

Senses (1):

1. +UPOS=NOUN

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

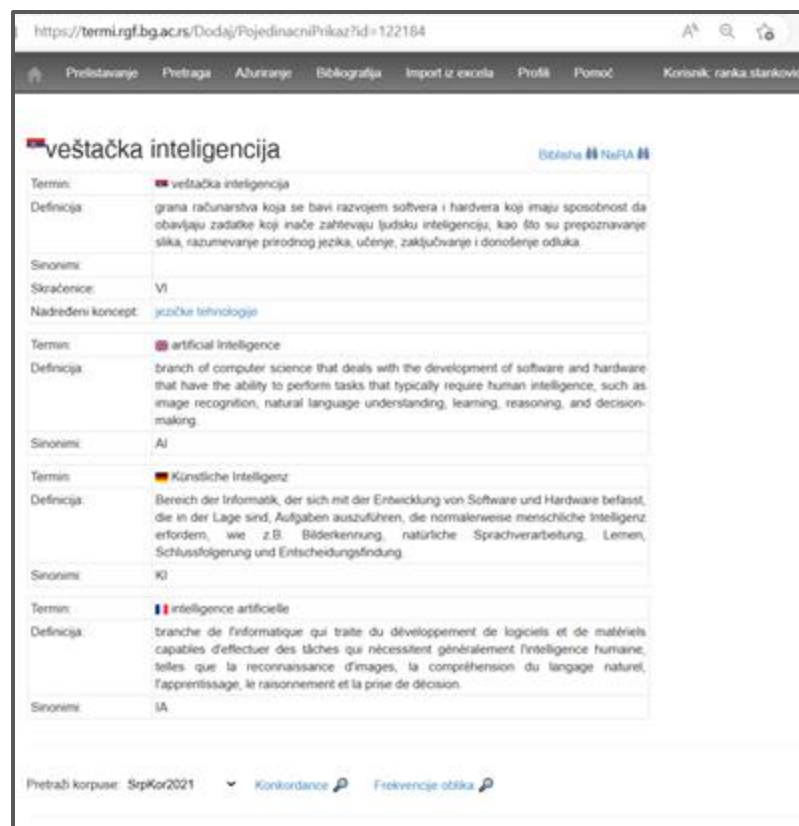
de Adali znanostje inteligencija pojedina .	osnovni inteligencijom	
me je izložio . Razume se , samoća je opasna za	diva inteligencija	Potreba nam je okolina ljudi koji misle i
težakih mogućnosti , sprječava struktura	veštačka inteligencija	, znanstvena regulativa , zaštita sponzora
(17) , učenje na daljinu (32) , web saj (5) ,	veštačka inteligencija	(4) od . Kod termina koji su sačinjeni od
* 1 NZN grf1 sin veština sleva 4 1 ANN grf1 sin	veštačka inteligencija	* 1 ANN grf1 sin visokofilofska biblioteka 95 1
znajem 6 NZN grf6 sin web saj 5 ANN grf1 sin	veštačka inteligencija	4 ANN grf1 sin visokofilofska ustanova 9 ANN
pojavom oblika može da nastupi samo tako gde se	katolička inteligencija	spri i trajni mladačkan odličijem i
napise ostaci prevačkanosti i neka govore	konstaktrna inteligencija	. A prvi opesi vel su se videli . Kod se
mu se omo diviti , dobro koriti i koja seče izmi	vile inteligencija	narje bio je to nepobno kad je čovek pokaze
... A Dio , koga istina nije kralja neka	izovna inteligencija	, ali koji nije bio bit potpna budala , stvarno
tegneca , indostrijalca , i juka	jevrejska inteligencija	koja je proba čeo suadarki javni i kulturni
ali je bilo moguće interpretirati samo 4 . To su	masovna inteligencija	, generalni faktor telesne snage , generalni
primer : muzja , intelektualne sposobnosti ,	socijalna inteligencija	, komunikacijska sposobnost , jezička
kompetencije nadovezje na kometridi	ocijalna inteligencija	i socijalne kompetencije . U istražvanju
posmatrano domonacija sileki pojnovi	ocijalna inteligencija	, socijalna kompetencija , komunikacijska
do sedamdesetih godina XX veka koriti pojnovi	ocijalna inteligencija	, s od sedamdesetih godina domonacija pojnovi :
da se u budućnosti vile neće ispraviti	ocijalna inteligencija	, teje socijalna kompetencija . Kao
se naglavava delatni aspekti laičko stravece	ocijalna inteligencija	(Zarić , Bjokić , 2004 : 325) . On socijalni

интеграција лексичке базе и корпуса

A(N)

veštačke inteligencije	581
veštačka inteligencija	220
veštačku inteligenciju	196
srpske inteligencije	158
veštačkoj inteligenciji	106
politena inteligencija	105
Veštačka inteligencija	102
veštačkom inteligencijom	75
politene inteligencije	70
emocionalne inteligencije	65
poslovne inteligencije	55
srpska inteligencija	51
kritičke inteligencije	48
koefficijenta inteligencije	47
emociionalna inteligencija	46
nacionalne inteligencije	45
politeni inteligencija	45
tehničke inteligencije	44
Emociionalna inteligencija	39
ljudske inteligencije	39
misioarske inteligencije	33
političke inteligencije	31

veštački inteligencija	1307
politni inteligencija	269
srpski inteligencija	256
emocionalni inteligencija	200
misioarski inteligencija	120
poslovan inteligencija	110
kritički inteligencija	99
tehnički inteligencija	92
socijalan inteligencija	86
ljudski inteligencija	82
nacionalan inteligencija	79
liberalan inteligencija	62
visok inteligencija	59
natprosečan inteligencija	57
politički inteligencija	55
koefficijent inteligencija	53
ruski inteligencija	50
prosečan inteligencija	44
elementaran inteligencija	43
kollektivan inteligencija	41
seksualan inteligencija	40



Olivera Kitanović, Ranka Stanković, Aleksandra Tomašević, Mihailo Škorić, Ivan Babić, Ljiljana Kolonja, “A Data Driven Approach for Raw Material Terminology”, Applied Sciences (2021), [pdf](#)

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, kokeni ugalj, listasti ugalj, mrki ugalj, plameni ugalj, prašinst ugalj, ugljena prašina

Biblisha 🗣️ NaFla 🗣️ Detaljno 🗣️

Pretraži korpus: Konkordance 🗣️ Frekvencije oblika 🗣️

© 2022 - Univerzitet u Beogradu, Rudarsko-geološki fakultet

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

Plain lemma

uglja	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 coal

Concordance - Prof 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 ; učelće 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

uloška 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 karakosom od čeličnih uzadi **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

1 , 5 1 , 0 Pesak , cement , sitniji kameni **ugalj** i šlični materijali sa 2000 kg / m 3 i amax 60 mm

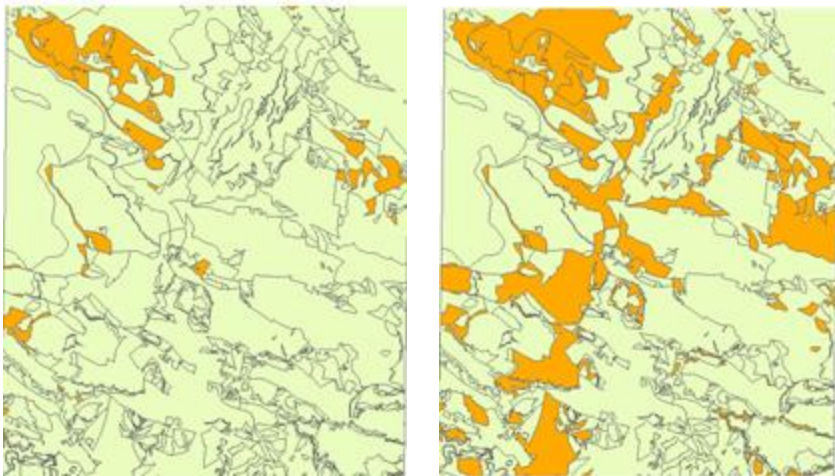
i amax 60 mm 1 , 5 do 3 , 0 2 , 0 Srednje krupni kameni **ugalj** , briket i šlični materijali sa 2000 kg / m 3 i

kao što su : rovna i drobljena ruda , šljunak , **ugalj** i sl . Opšta šema člankastog transportera nasipni materijali (prašinst ugalj , pesak , glina , cement itd) . Utovar materijala , drobljena ruda , cementni klinker , krupni **ugalj** itd .) teško se zahvataju vedrima jer se zbog

Унапређење ГИС-а лексичким и термиолошким ресурсима

Морфолошко проширење:

кречњак --> кречњак, кречњака, кречњаку,
кречњаком, кречњаче, кречњаџи,
кречњаџима, кречњаке



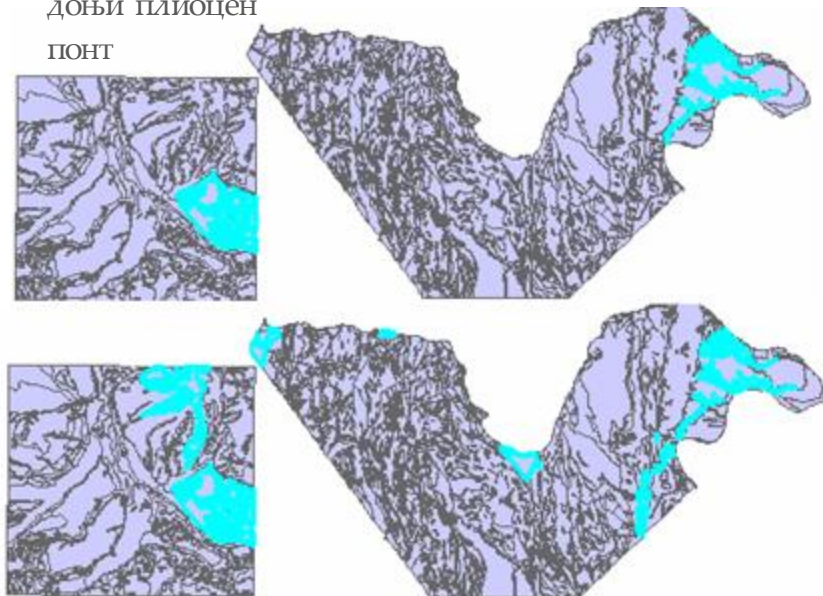
Ranka Stanković, Ivan Obradović, Olivera Kitanović, "GIS Application Improvement with Multilingual Lexical and Terminological Resources", 5th International Conference on Language Resources and Evaluation, LREC 2010, Valetta, Malta, 2010

Семантичко проширење:

доњи плиоцен -->

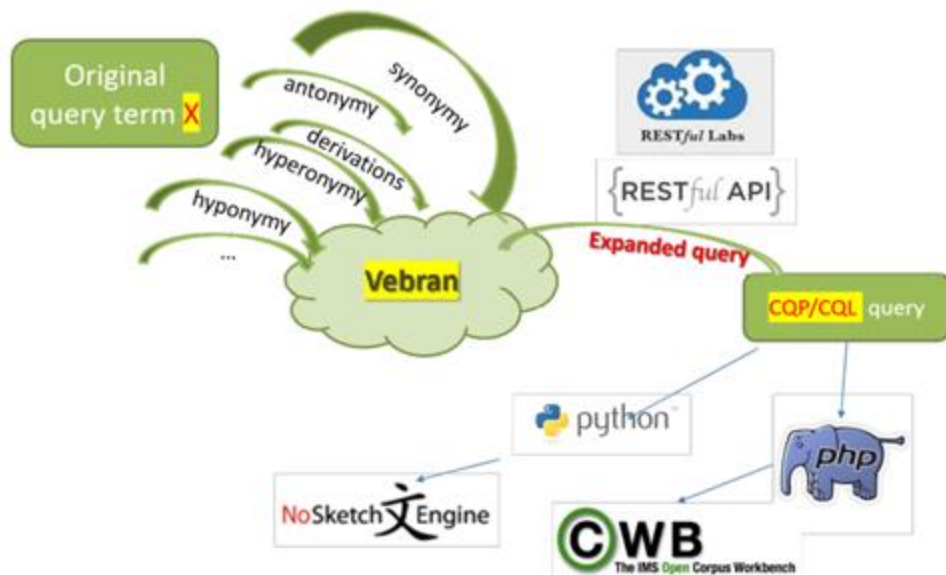
доњи плиоцен

понт



Ranka Stanković, "Improvement of geodatabase queries within GeoIIS", Review of the National Center for Digitization, Beograd: Faculty of Mathematics, Belgrade

Вебран у функцији проширења упита над корпусима



[flemma="kap"]
"kap(lilimalju)"
... & gram= "s"] "kap(lilju)"

Stanković, Ranka; Utvić, Miloš. Vebran Web Services for Corpus Query Expansion. *Infotheca - Journal for Digital Humanities*, v. 19, n. 2, p. 99–118, 2020. [\[pdf\]](#)

Природнојезичка подршка репозиторијумима и дигиталним библиотекама

- индексирање по пуном тексту, укључујући све флективне облике ушита
- рангирање према релевантности и (сагласности са уштом)
- исечци текста (конкорданце) са обојеним речима из ушита које потврђују релевантност поготка

Упит: **veštačka inteligencija ontologije** ×

Active facets

Катедра: Катедра за примењену математику и информатику x

1 од 1 < > 1-9 од 9 Per page Results by 20 Sort by Релевантност export citations

Тип

Рад у зборнику (4)
Рад у часопису (3)
Докторска дисертација (1)
Монографија (1)

Медија

pdf (9)

Одсек

Катедре и кабинети општих предмета (9)
Рударски одсек (4)
Геолошки одсек (2)

Депарتمان

Депарتمان за регионалну геологију (1)
Централна библиотека (1)

Катедра

Катедра за примењену математику информатику (9)
Катедра за заштиту на раду и заштиту животне средине (3)
Катедра за динамичку геологију (1)
Катедра за историју геологију (1)
Катедра за рударска мерења (1)

Аутор

Ранка Станковић (9)
Радови истраживача (8)
Иван Обрадовић (4)
Никола Лилић (3)
Александар Цвјетић (1)
Александра Томашевић (1)
Биљана Рујевић (1)
Бранислав Тривић (1)
Игор Миљановић (1)
Михаило Шкорић (1)
Ненад Бањаци (1)
Оливера Китановић (1)

Ontološki model upravljanja rizikom u rudarstvu cite
Olivera Kitanović (2021)
Rudarska proizvodnja obuhvata kompleksne tehnološke sisteme, što nameće potrebu za uspostavljanjem i unapređivanjem sistema upravljanja rizikom. Heterogenost i obim podataka neophodnih za upravljanje rizikom zahtevaju sistem koji ih na fleksibilan način integriše i omogućava njihovo optimalno korišćenje. Osnovni cilj ove disertacije je razvoj ontologije za domen rudarstva i na njoj zasnovanog modela za upravljanje rizikom. Njegova realizacija podrazumeva i implementaciju algoritama ekstrakcije informacija za popunjavanje ontologije, kao i odgovarajuće softversko rešenje. Razvoj modela obuhvata i značajno proširenje rudarskog korpusa, kao ...

rudarstvo, rizik, upravljanje rizikom, procena rizika, ontologija, semantička mreža, ekstrakcija informacija, upravljanje znanjem, računarska lingvistika

... radnog točka. 35 3. METODOLOGIJA I RESURSI ZA RAZVOJ ONTOLOGIJE 3.1. Šta je ontologija? Termin **ontologija** potiče iz filozofije gde se odnosi na granu metafizike koja se bavi postojanjem bića, dok u informatici i **veštačkoj inteligenciji** se odnosi na „eksplicitnu specifikaciju konceptualizacije“ ...

... u vidu standardizovanog jezika, pa se smatra da **ontologije** pružaju infrastrukturu za transformaciju veb informacija i podataka u veb znanja – semantički veb“ (Gađević, Burić, and Devedžić 2006). Razvoj **ontologija** se vezuje za **veštačku inteligenciju** i semantički veb, ali se primena može naći u mnogim ...

... klasifikacija povreda, terminologija iz oblasti nafte i gasa i sl. Glavni cilj **ontologije** RudOnto jeste strukturiranje i efikasna upotreba znanja u oblasti rudarstva, u okviru sistema poslovne **inteligencije** za upravljanje zaštitom na radu u rudarstvu (Kolonia 2016). RudOnto sadrži blizu 7.000 termina na srpskom ...

Hibridni sistem za planiranje i analizu ventilacije rudnika cite
Nikola Lilić, Ranka Stanković, Ivan Obradović (2000)
Ventilacija rudnika, informacioni sistemi
... da izvršava slične zadatke kao od svih zadržavaju korišćenje **inteligencije** napravi se inteligentni sistem, a običaj koji se bavi radom ovakvih sistema nalazi se **veštačkoj inteligenciji**. Počevši razvojem **veštačke inteligencije** običaj se vezuje za 1956. godinu kada se grupa istraživača sastala ...
... god radovim od matematičkih modela (u **veštačkoj inteligenciji** S.A.D.R.Z.A.) predložio ...
... 20-2 O inpart Metodologija planiranja ...
... predložio primenu računarskih metoda (tehniku **veštačke inteligencije** u planiranju i analizi rudničkih ventilacionih sistema. Cilj ove knjige je da prikaže rezultate napravljenog hibridnog sistema INVENTS, u kome su određene metode **veštačke inteligencije** kombinovane sa klasičnim proceduralnim postupkom ...

Razvoj geološkog informacionog sistema Republike Srbije cite
Branislav Enešević, Branislav Trivić, Miroslav Banjac, Ranka Stanković, Veljko Nikolić (2005)
Geološki informacioni sistemi Srbije (GeolMIS) je projektovan, pravstvenno, sa namernom efikasnom digitalnom arhiviranim geološkim i nimia sredstva podataka. U ovom radu je prikazana struktura baze podataka kao osnova za razvoj geološki koncipiranog GIS-a. Novi, objektivno orijentisani (O-O) način modeliranja omogućio je definisanje samostalnih tipova objekata, hierarhijski povezanih kroz topologije i druge relacije, čime je obezbeđena nizova međusobna interakcija. Objektivno orijentisano modeliranje izvršeno je koristeći jezike modeliranja (UML) i CASE alata, kroz konceptualni i logički nivo. Filozofski model Te ...
GeolMIS, geološki podaci, upravljanje bazama podataka, konceptualni model, logički model, GIS
... (Enešević Miroslav Languages) Admaltan metodološki pristup, preko konceptualnog i logičkog strukturiranja GeolMIS-a, obezbedio je kompletno truljene **ontologije** i semantike. Stoga su u radu kritički razmotrene prednosti i nedostaci ovakvog načina modeliranja geoloških podataka, odnočno osnovni principi objektivno ...
... Tehnologija neutralno, organizovanu digitalnu bazu geoloških podataka i interpretaciju. U informacionom učuju konceptualni model su primeni kao **inteligenciji** (Guzlino, 1996) klasične kroz hierarhijski klasifikacione sisteme, topologije i vokabulare, zasnovane na teoriji kvantitativne opserviranja, te ...
... „klasifikacije“ St. 1. Osnovni elementi konceptualnog modela GeolMIS-a. Fig. 1. Basic elements of the Geologic Information System Conceptual model. **Ontologizacija** – konceptom se obezbeđuje deduktivno (suzbijanje) i logično uređivanje opserviranja kontrolisano Geološkim Rešenjem, dok se klasično opserviranje omogućava ...

ool cite

do 12. februara 2021. godine sa ciljem i polaznici su se upoznali sa širokim ih podataka pomoću najsavremenijih ovaj ju je virtualno (onlajn) nekoliko

usLinguarum, COST akcija, škola za obuku

gent systems cite


the solution of complex problems that ent systems (HIS) can be successfully analyzed and a classification of hybrid engineering ...

nene veštačke inteligencije

R. (2000). Sprezanje metoda **veštačke inteligencije** u cilju

Nebezbedno obrada.jerteh.rs

JePTex – Друштво за jezičke resurse i tehnologije



Paste Text Upload File

Циклус предавања: ВЕШТАЧКА ИНТЕЛИГЕНЦИЈА посећују бројни стручњаци.

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

Go

Циклус N циклус
 предавања N предавање
 : PUNCT :
 ВЕШТАЧКА A вештачки
 ИНТЕЛИГЕНЦИЈА N интелигенција
 посећују V посећивати
 бројни A бројни
 стручњаци N стручњак
 . PUNCT .

Sent. del.: 1 Tokens: 16 Simple forms: 7 Digits: 0
 Errors: 0 Error percent: 0.0%

- 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](#)

← ↻ 🔒 https://ners.jerteh.rs

Process plain text file, TMX, or a string

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

Or type in/paste the text below:

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

Select parsing method:



It-Sr-NER

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

Language Resource Switchboard Upload Tool Inventory Help CLARIN

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)
- > spaCy bilingual NER (for TMX)
- > spaCy bilingual NER and Geoparsing (for TMX)
- > spaCy binlingual NER and NEL (for TMX)

PERS LOC ORG DEMO WORK EVENT Download All

<p>#1 Abilissimi agenti di polizia sono stati inviati in America e in Europa in tutti i principali porti d'imbocco e di sbarco, e sarà ben difficile che questi signori possano sfuggire alla loro caccia.</p>	<p>#1 Policički inspektori vrlo veliki ljudi, poslani su u Ameriku i Evropu u svim glavnim lukama za ulazak i izlazak, i taj će gospodin teško moći pobjeći od njihove ruke.</p>
<p>#2 Perciò respinsi a forza il pensiero di Nata, presi Lucia sottobraccio, mossimoci al più presto, voglio andar via di casa, voglio prendere la patente, voglio viaggiare, voglio avere il telefono, la televisione, non ho mai avuto niente.</p>	<p>#2 Zato sklopili snagu i izbacih Nata iz misli, starih Lucia pod ruku, posumnjali: da, hajde da se vrsavamo što pre. Želim da odem od kuće. Želim da polučim voznački list, da putujem, da imam telefon, televizor, nikada nista nisam imala.</p>
<p>#3 Il professor Guido Agosti provò a persuadere il figlio, ci furono lunghe telefonate nel corso delle quali con un certo stupore sentii che il vecchio pedeva la calza, alzava la voce.</p>	<p>#3 Profesor Guido Agosti pokušao je da ubedi sina da povuče spazbu, vodili su duge razgovore tokom kojih sam zapazeno shvatila kako njegovog oca izdaje stjepljenje, vikao je s druge strane slušalice.</p>
<p>#4 La chiesa non era maiestosa come altre che vidi in seguito a Venezia e Chianci e Bamburzi e a Parisi.</p>	<p>#4 Crkva nije bila veličanstvena poput onih koje sam kasnije video u Sredozemju, Saraju, Bamburzi i Parizu.</p>
<p>#5 Si raccontava che l'uciprete, vedendo quello che accadeva nella lanaba, avesse tenuto nella chiesa di Dobruša una veglia di meditazione, indossando abiti rivoltati e con le candele accese alla rovescia.</p>	<p>#5 Pričalo se da je proto, videći šta se dešava u kasabi, držao u dobrušanskoj crkvi kletveno bdeenje, u izvratnoj odehli i sa svećama zapaljenim unopako.</p>
<p>#6 Crebbe allora l'importanza del ponte sulla Tisza, quale unico collegamento sicuro tra il pascialato bosniaco e la Saraha.</p>	<p>#6 U tim vremenima važnost mosta kao jedine sigurne veze između bosničkog pašaluka i Saraja neobitno je porasla.</p>

- Oља Перишић, Ранка Станковић, Милица Иконић Нешић, Михаило Шкорић, **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 Ikončić 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)
- Oља Перишић, „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

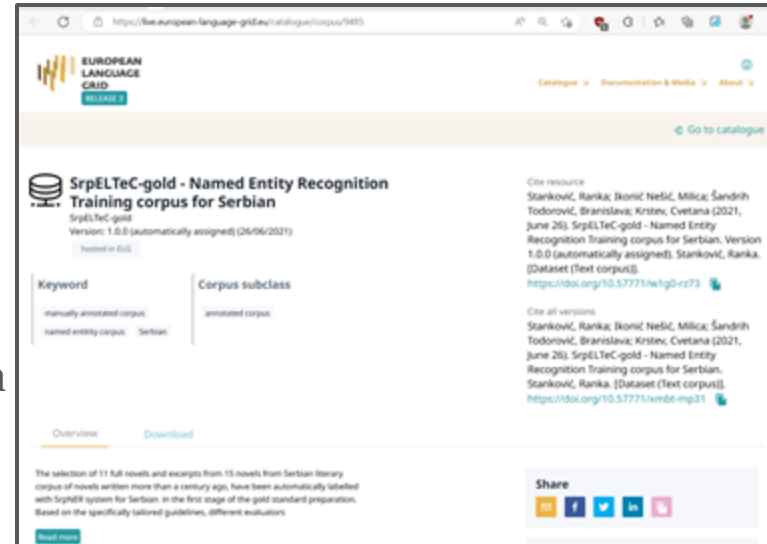
Text Generator for Serbian

- Tanja Ivanović, Ranka Stanković, Branislava Šandrih Todorović, Cvetana Krstev, “**Corpus-based bilingual terminology extraction in the power engineering domain**”, 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**“, 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 (сентимента)
- за препознавање говора мржње.



Корпуси у отвореном приступу

- [SrpELTeC-gold](#) – Корпус за тренирање препознавања именованих ентитета - поткорпус SrpELTeC 11 романа у пуном тексту и исечке из 15 романа. У првој фази обележен системом SrpNER. (330.119 токена, 7 класа).
- [SrpKor4Tagging](#) – Корпус формиран комбинацијом књижевних ($\frac{1}{3}$) и административних ($\frac{2}{3}$) текстова на српском језику. Обележен са два скупа ознака за врсте речи: Universal POS и SrpLemKor и лематизован, 342.803 токена.
- [RudKorP](#) – Српски јавни корпус текстова из области рударства и обраде минералних сировина (2.34 милиона речи).
- [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.

Корпуси у отвореном приступу


- [INTERA Corpus – the Serbian-English part](#) – паралелни корпус од милион речи за српски и милион речи за енглески, упарено на нивоу реченице.
- [INTERA Corpus – the Serbian POS annotated part of the SR-EN pair](#) – милион речи.
- [Орвелова 1984.](#) – аутоматски аотиран и ручно коригован пуним граматичким категоријама, врстама речи и лемама, у оквиру вишејезичног пројекта.
- [Пут око света за 80 дана](#) – паралелни корпус са 3.700 упарених сегмената, углавном реченица.

- 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

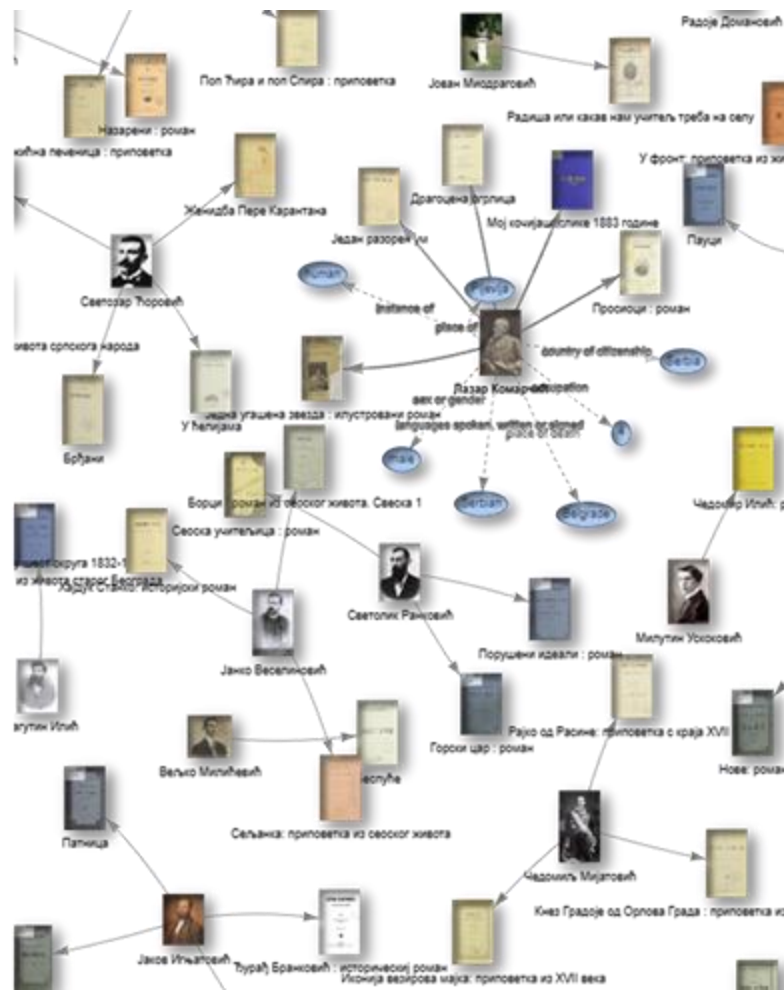
- Jelena Andonovski, **OCR and TEI for the Production of ELTeC - Würzburg Training School, 16-17 April 2018**, [\[pdf\]](#)
- Mihailo Škorić, **Methods and Tools of Distant Reading Adapted to Multiple European Languages**, Galway Training School, [\[pdf\]](#)
- Vasilije Milnović, Budapest Training School – **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



„Причање је души посланица“
Његуш



`<quote>`
`<p>`„Причање је души посланица“`</p>`
`</quote>`
`<p>`Његуш`</p>`



После успешне викификације *SrpELTeC* корпуса уследило је повезивање романа који се налазе у корпусима *SrpKor2013/2021*.

Модели

- [SrpCNNER](#) – препознаје 7 типова именованих ентитета, са скором F1 од око 91% на тестном скупу.
 - 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](#) – обележава скуповима ознака *UPOS* и *SrpLemKor*, обучен коришћењем аотираног корпуса *SrpKor4Tagging* и речника *SrpMD4Tagging*.
 - 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 *UPOS* и *SrpLemKor*, обучени коришћењем аотираног корпуса *SrpKor4Tagging*.
 - 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)* а у оквиру пројекта ЕуроТра (1982.-1997.)
 - 10.288 вишечланих речи, 88.753 речи и 3.753.750 флективних облика у јавно доступној дистрибуцији
 - 23.000 вишечланих речи, 211.000 речи и 8.417.000 флективних облика
- [SrpMD4Tagging](#) – *Serbian Morphological Dictionaries for Tagging* – Морфолошки речници српског језика за обележавање изведени из SrpMD (Krstev & Vitas); придруживање леме и ознаке врсте речи облику речи (2 датотеке: *UPOS* и традиционални *POS*), број аотираних речничких облика 935.466.
- [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) (2631 термина са дефиницијама, синонимима и преводима на енглески).

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)
- Напредни модели
 - ГПТ2 (136 и 800 милиона параметара)
 - RoBERTa (150 и 300 милиона параметара)

Михаило Шкорић, Композитне псеудограматике засноване на паралелним језичким моделима српског језика, докторска дисертација 2023

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, which is divided into two main panels. The left panel is titled "Input text to process or to use as context" and contains a text input field, a "Select language model" dropdown menu (set to "procesaur/gpt2-srlat"), a "Text perplexity visualisation" section with a "Slider window size" set to 3 and a "Visualize" button, a "Generation (GPT-2 only)" section with a "Generate" button, and a "Full Evaluation (Serbian only)" button. The right panel is titled "About PaLMA" and provides information about the application's development, usage instructions, and a "Perplexity visualisation" section. This section includes requirements for input text, a list of supported language models (GPT and BERT), and a line graph showing the perplexity of the generated text. The graph has a y-axis labeled "Perplexity" ranging from 0 to 10 and an x-axis labeled "Tokens" ranging from 0 to 10. The curve shows a peak in perplexity around token 5, followed by a decline and then a slight rise again.

Input text to process or to use as context

Type here, ...

Select language model:
procesaur/gpt2-srlat

Text perplexity visualisation

Slider window size: 3 Visualize

Generation (GPT-2 only) Generate

Text length (tokens): 10

Temperature (creativity): 0.7

Number of outputs: 3

Select second model (for evaluation):
huggingface's 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 pseudogramars based on parallel language models of Serbian by Mihaljo Skocic, 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 others) 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:

will output:

Perplexity

Generation

ГПТ2-ОРАО

Највећи генеративни модел за српски језик

Заснован на GPT2-large архитектури (~800 милиона параметара)

Генерише нови текст, или наставља започети текстуални унос.

НОВО: Једнака подршка уноса и на ћирилици и на латиници!

обучаван на отвореним корпусима Друштва за језичке ресурсе и технологије.

Осим тога, коришћени су и остали доступни корпуси: [PDRS 1.0](#), који је развио Институт за српски језик САНУ, [cc100\(sr\)](#) i [SRWAC](#).

ГПТ2-ВРАБАЦ

мањи модел обучен на истом корпусу

Модели су обучавани на [Националној платформи за вештачку интелигенцију Србије](#) (систем који се базира на nVidia DGX системима).

RDF формат корпуса: NLP Interchange Format (NIF)

Приступна тачка (end point) <http://fuseki.jerteh.rs/>

The screenshot shows the Apache Jena Fuseki web interface. The browser address bar displays `http://fuseki.jerteh.rs/datasets/SrpELTeCTest2/query`. The page title is `/SrpELTeCTest2`. Below the title, there are options for 'query', 'add data', 'edit', and 'info'. A section titled 'SPARQL Query' contains a text area for entering queries and a 'Prefixes' section with radio buttons for 'off', 'nif', and 'all'. Below this, there are dropdown menus for 'Content Type (SELECT)' and 'Content Type (GRAPH)'. A 'Turbo' checkbox is also visible. The main area shows a SPARQL query: `SELECT ?uri ?label WHERE { ?uri nif:isAnchorOf ?label }`. Below the query, there is a 'Table' section showing 13 results in a table with columns 'uri', 'label', and 'etype'. The table contains rows of URIs and labels like 'српски', 'магров', 'Црљур', 'Српски', 'Капичуви', 'Младш', and 'Младш'.

- осим корпуса, доступни су и лексички ресурси који следе Ontolex-lemon модел за онтологије
- W3C Ontology Lexicon community group (OntoLex)

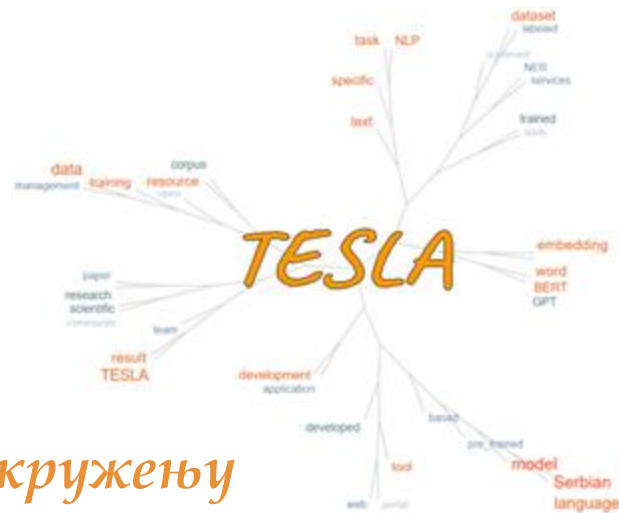
```
<http://lloj.jerteh.rs/ELTEC/srp/NIF/SRP19203.txt#char=14876,14880> a nif:RFC5147String,
  nif:Word ;
nif:anchorOf "шам"^^xsd:string ;
nif:beginIndex "14876"^^xsd:nonNegativeInteger ;
nif:endIndex "14880"^^xsd:nonNegativeInteger ;
nif:lemma "шати"^^xsd:string ;
nif:nextWord <http://lloj.jerteh.rs/ELTEC/srp/NIF/SRP19203.txt#char=14880,14881> ;
nif:posTag "VERB"^^xsd:string ;
nif:previousWord <http://lloj.jerteh.rs/ELTEC/srp/NIF/SRP19203.txt#char=14873,14875> ;
nif:referenceContext eltec:SRP19203.txt ;
nif:sentence <http://lloj.jerteh.rs/ELTEC/srp/NIF/SRP19203.txt#char=14871,14994> .
```

Актуелна и будућа истраживања

- велики језички модели: BERT, нова верзија GPT-а, T5
- унапређење тагирања врстама речи
- унапређење анотација именованих ентитета
- анотација синтаксе
- континуалан рад на унапређењу ресурса

- Конференција у Београду, април 2024.

Јужнословенски језици у дигиталном окружењу



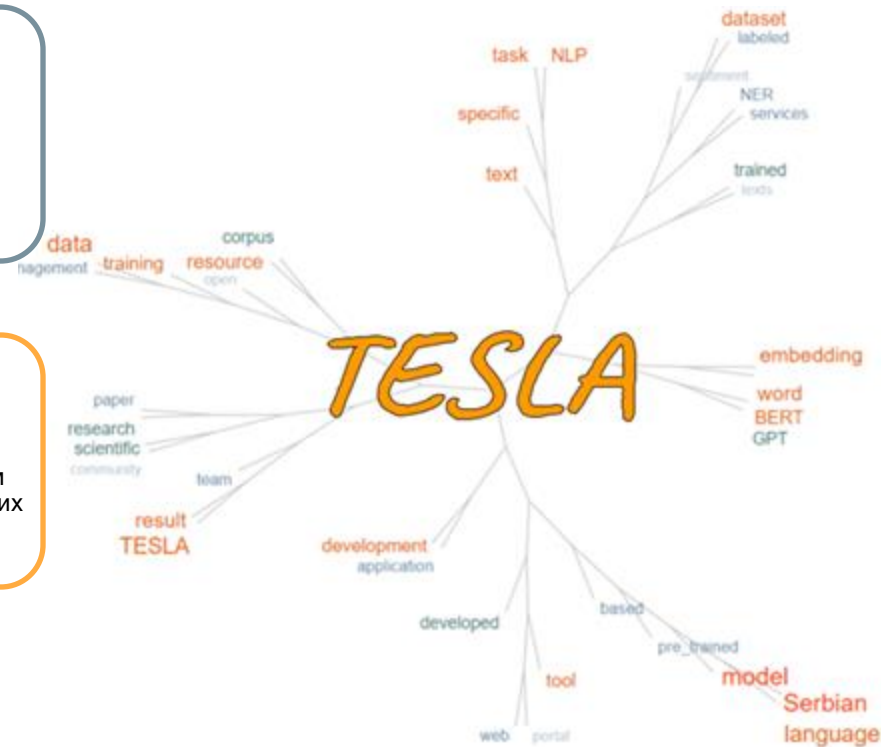
Text Embeddings – Serbian Language Applications

Универзитет
у Београду

- Рударско-геолошки факултет
- Филолошки факултет

Очекивани
резултати

- скуп предобучених језичких модела за српски језик
- алати и сервиси (укључујући и екстракцију релација, генерисање текста и чет ботове)
- веб портал са јавно доступним алатима, са корисним визуелизацијама које олакшавају разумевање језичких образаца и феномена детектованих у текстовима.



Шта можемо очекивати?



да НЛУ отвори нове
могућности за
привреду и
појединце,

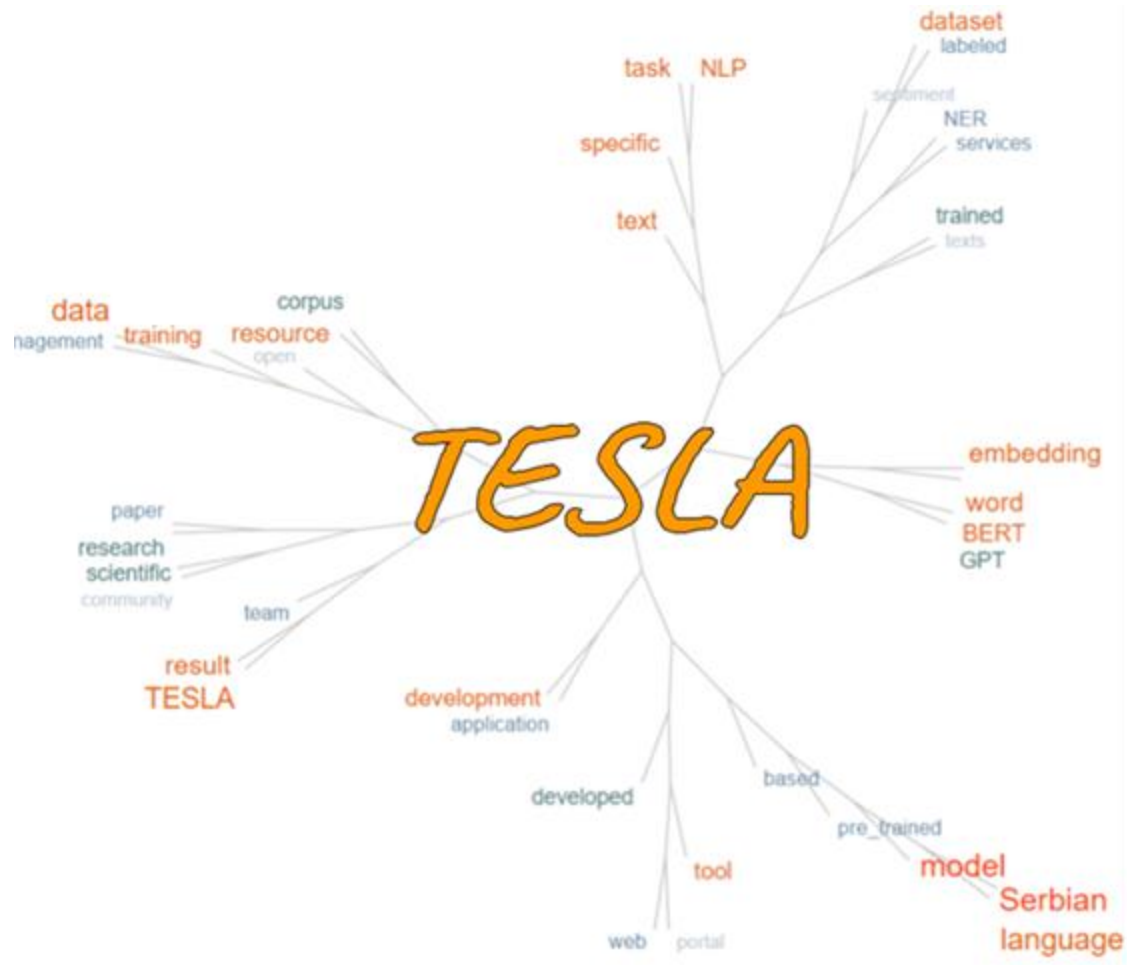
Како технологија
напредује, можемо
очекивати да
видимо
софистицираније
НЛУ апликације које
ће наставити да
побољшавају наш
свакодневни живот.



Хвала на пажњи!

Питања?

ranka@rgf.rs



Историја развоја векторских репрезентација

Bag of Words (BOW) [1954]: Врећа речи: број појављивања речи у документу се користи као својство текста. **TF-IDF** [1972]: модификована врећа речи где речи које се нађу у малом броју докумената имају високу вредност, а речи које се појављују у великом броју докумената имају ниске вредности.

Word2Vec [2013]: реч је представљена векторском репрезентацијом велике димензије који се зове ембединг (уграђња, уграђивање речи) и која тежи да ухвати њену семантику, обучава се НМ на великом корпусу.

RNN [1986]: рачунају ембединг докумената користећи контекст речи у реченицама, што није било могуће само са ембединзима речи. Касније је еволуирао помоћу **LSTM**-а [1997] за памћење дугих зависности и до **Bidirectional RNN**-а [1997] за леве и десне контекст, и коначно су се појавили **Encoder-Decoder RNNs** [2014], где једна **RNN** кодира документ у вектор, друга **RNN** га декодира у текст.

Transformer [2017]: код ер-декодер модел који користи механизме пажње за израчунавање бољих уграђња и боље усклађивање излаза са улазом.

BERT [2018]: предобучен двосмерни **T**рансформер коришћењем комбинације циљева моделирања маскираног језика и предвиђања следеће реченице, коришћењем глобалне пажње.

GPT [2018]: први ауторегресивни модел заснован на архитектури **T**рансформера, еволуирао у **v**ећу и оптимизовану верзију **GPT-2** [2019], предобучен на тексту са веба, и **GPT-3** [2020], још већа и оптимизована верзију **GPT-2**, предобучена на **Common Crawl**.

CTRL [2019]: сличан **GPT**-у, али са контролним кодовима за условно генерисање текста.

Transformer-XL [2019]: ауторегресивни **T**рансформер који може поново да користи претходно израчуната скривена стања да би обрадио дужи контекст.

ALBERT [2019]: лакша (упрошћена) верзија **BERT**-а: предвиђање следеће реченице замењено предвиђањем редоследа реченица, и технике смањења параметара за мање меморије и брже обучавање.

RoBERTa [2019]: боља верзија **BERT**-а, где (1) циљ моделирања маскирањем се адаптира, (2) циљ предвиђања следеће реченице је избачен, (3) користи се **BPE** токенизатор и (4) бољи хиперпараметри.

XLNet [2019]: **T**рансформер предобучен на вишејезичном корпусу за моделирање: узрочног језика (Causal Language), маскираног језика (Masked LM) и превођења (Translation LM).

XLNet [2019]: **T**рансформер-**XL** са генерализованом ауторегресивном методом предобучавања која омогућава учење двосмерних зависности.

PEGASUS [2019]: двосмерни кодер и декодер лево-десно унапред обучени са циљевима моделирања маскираног језика и генерисања недостајућих реченица (Gap Sentence Generation).

DistilBERT [2019]: исти као **BERT**, али мањи и бржи, уз очување преко 95% **BERT**-ових перформанси. Обучен дестилацијом претходно обученог **BERT** модела.

XLNet-RoBERTa [2019]: **RoBERTa** је обучен на вишејезичном корпусу са циљем моделирања маскираног језика.

BART [2019]: двосмерни кодер и декодер слева надесно обучени кварирањем текста са произвољном функцијом буке и учењем модела за реконструкцију оригиналног текста.

ConvBERT [2019]: боља верзија **BERT**-а, где су блокови самопажње (self-attention) замењени новим који користе конволуције за бољи модел глобалног и локалног контекста.

Funnel Transformer [2020]: тип трансформера који постепено компримује низ скривених стања на крају низ стања и самим тим смањује трошкове израчунавања.

Reformer [2020]: ефикаснији трансформер захваљујући локално осетљивој пажњи на хеширање, кодирању аксијалног положаја и другим оптимизацијама.

T5 [2020]: двосмерни енкодер и декодер лево-десно унапред обучени за комбинацију задатака без надгледања и са надгледањем.

Longformer [2020]: модел трансформера који замењује матрице пажње ретким матрицама за већу ефикасност обучавања.

ProphetNet [2020]: модел трансформера обучен за предикцију енграма са новим механизмом самопажње.

ELECTRA [2020]: исто као **BERT**, али лакши и бољи, обучен да открије одговарајућу замену токена (Replaced Token Detection objective).

Switch Transformers [2021]: ретко активан модел експертског трансформера који има за циљ да поједностави и побољша обучавање у односу на комбинацију експертских знања.

Дисертације

- Михаило Шкорић, Композитне псеудограматике засноване на паралелним језичким моделима српског језика, 2023.
- Биљана Рујевић, Речници у дигиталном добу – информатичка подршка за српски језик, 2022.
- Тања Ивановић, Развој терминологије из области електроенергетике коришћењем метода обраде природних језика, 2022.
- Милена Милинковић, Развој библиотечких и језичких ресурса за организовање и проналажење информација о просторном планирању, 2022.
- Оливера Китановић, Онтолошки модел управљања ризиком у рударству, 2021.
- Бранислава Шандрић, Impact of text classification on natural language processing applications, 2020.
- Јелена Андоновски, Мрежа отворених података и језички ресурси у процесу изградње српско-немачког литерарног корпуса, 2020.
- Јелена Митровић, Електронски језички ресурси и алати за обраду српског језика и њихово унапређивање путем групне расподеле рада, 2018.
- Александра Томашевић, Развој модела за управљање рударском пројектном документацијом, 2018.
- Бојана Ђорђевић, Израда основа формалне граматике српског језика употребом метаграматике, 2017.
- Зоран Ристовић, Кумулативни ефекти експлоатације вишејезичних корпуса у настави страних језика, 2016.
- Миљана Младеновић, Информатички модели у анализи осећања засновани на језичким ресурсима, 2016.
- Александра Тртовац, Дескриптори метаподатака и дескриптори садржаја у проналажењу информација у дигиталним библиотекама, 2016.
- Јелена Јаћимовић, Аутоматско препознавање и нормализација временских израза у неструктурираним новинским и медицинским текстовима на српском језику, 2016.
- Стаца Вујичић-Станковић, Екстракција информација вођена онтологијама: (модел за српски језик), 2016.
- Небојша Васиљевић, Аутоматска обрада правних текстова на српском језику, Филолошки факултет, ментор, 2015.
- Милош Утвић, Изградња референтног корпуса савременог српског језика, Филолошки факултет, ментор, 2014.

У току: Милица Иконић Нешић, Јелена Лазаревић, Милена Шошић, Данка Јокић, Марина Баги,...

COST акције



- Završene
 - PARSEME - Parsing and multi-word expressions. Towards linguistic precision and computational efficiency in natural language processing <https://typo.uni-konstanz.de/parseme/>
 - D-Reading - “Distant Reading for European Literary History”, <https://www.distant-reading.net>
- U toku:
 - NexusLinguarum – “European network for Web-centred linguistic data science” <https://nexuslinguarum.eu/>
 - LIHTME – “Language In Human Machine Era” <https://lithme.eu/>
 - UniDive - “Universality, diversity and idiosyncrasy in language technology” <https://unidive.lisn.upsaclay.fr/>
 - PhraConRep - “A Multilingual Repository of Phraseme Constructions in Central and Eastern European Languages” <https://www.cost.eu/actions/CA22115/>

Улагања у језичке технологије

Држава	Величина језичке заједнице	Улагања
Естонија	1,16 милиона	€811.000 годишње
Израел	6,1 милион	€2,5 милиона
Данска	5,5 милиона	€4 милиона
Исланд	0,33 милиона	€13 милиона
Словенија	2,5 милиона	€4 милиона

Састав ЈеРТеха

- Друштво за језичке ресурсе и технологије ЈеРТех окупља истраживаче различитих профила:
 - математичаре
 - информатичаре
 - лингвисте
 - лексикографе
 - терминологе
 - друге који су заинтересовани за развој и примену језичких технологија на научном, стручном и практичном нивоу.
- Отворено друштво за све заинтересоване сличних интересовања.

<https://jerteh.rs/>

<https://jerteh.rs/index.php/pristupnica/>