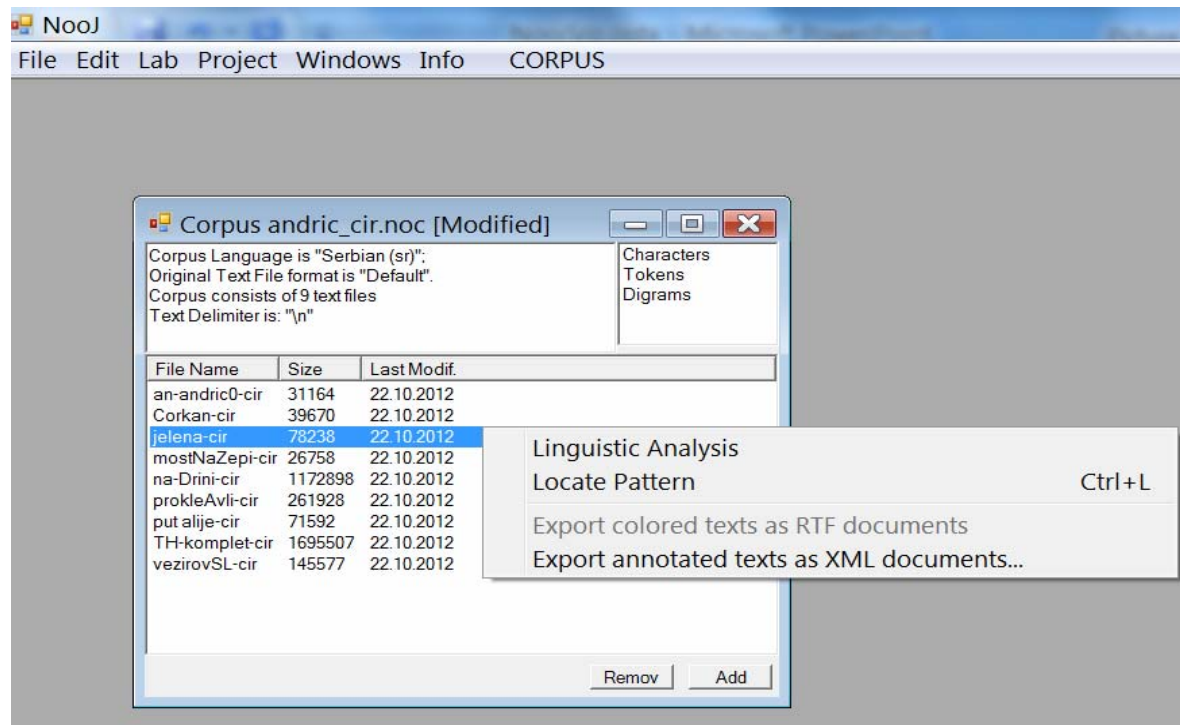


Различити нивои лингвистичких феномена

- ❑ Нуц, обрада природних језика заснована на језичким ресурсима
- ❑ СрпНуц, модул за српски језик је направљен у оквиру пројекта CESAR финансираног из фондова Европске Уније сачињава скуп ресурса на оба српска писма: ћирилицу и латиницу
 - Провера писања (Typography & Spelling)
 - Флективна и деривациона морфологија
 - Речници простих и сложених речи
 - Локалне граматике
 - Анотација и морфосинтаксичко обележавање
 - Екстракција информација
 - Проналажење информација

NooJ / Нуџ

- Нуџ је лингвистичко развојно окружење које применом речника и граматика анализира корпусе у реалном времену



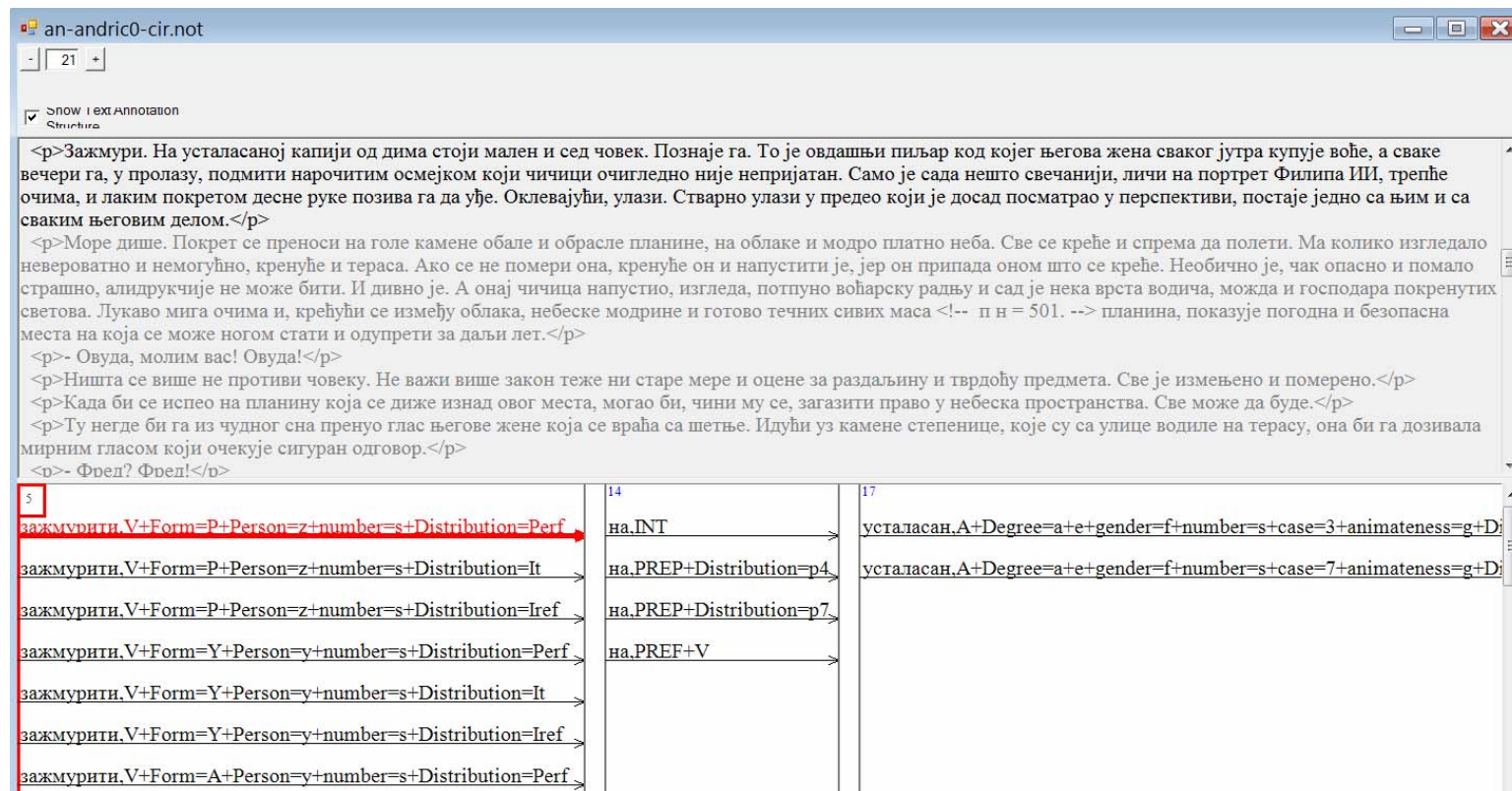
Креирање корпуса и лингвистичка анализа текста

<http://www.meta-net.eu>
<http://cesar.matf.bg.ac.rs> Дан језичких технологија Београд, Хотел Хајат 29. октобар 2012.

NooJ / Нуџ

- ❑ Нуџ садржи алате за креирање и одржавање лексичких ресурса, морфолошких и синтаксичких граматика
- ❑ Речници и граматике се примењују на текстове да би се идентификовали морфолошки, лексички и синтаксички обрасци и аотирале лексеме
- ❑ Нуџ омогућава креирање сложених конкорданси за све типове регуларних и контексно слободних образаца
- ❑ Екстракција семантичких јединица попут личних имена, локација, датума, техничких и финансијских израза и сл.
- ❑ Које су технологије и ресурси за то потребни?
 - Морфолошки речници
 - Коначни аутомати и транстуктори
 - Регуларни изрази
 - Локалне граматике

Морфолошка анализа



an-andric0-cir.not

21

SHOW TEXT Annotation Structure

<p>Зажмури. На усталасаној капији од дима стоји мален и сед човек. Познаје га. То је овдашњи пиљар код којег његова жена сваког јутра купује воће, а сваке вечери га, у пролазу, подмити нарочитим осмејком који чичици очигледно није непријатан. Само је сада нешто свечанији, личи на портрет Филипа III, трепће очима, и лаким покретом десне руке позива га да уђе. Оклевајући, улази. Стварно улази у предео који је досад посматрао у перспективи, постаје једно са њим и са сваким његовим делом.</p>

<p>Море дише. Покрет се преноси на голе камене обале и обрасле планине, на облаке и модро платно неба. Све се креће и спрема да полети. Ма колико изгледало невероватно и немогућно, кренуће и тераса. Ако се не помери она, кренуће он и напустити је, јер он припада оном што се креће. Необично је, чак опасно и помало страшно, алидрокције не може бити. И дивно је. А онај чичица напустио, изгледа, потпуно воћарску радњу и сад је нека врста водича, можда и господара покренутих светова. Лукаво мига очима и, крећући се између облака, небеске модрина и готово течних сивих маса <!-- п н = 501. --> планина, показује погодна и безопасна места на која се може ногом стати и одупрети за даљи лет.</p>

<p>- Овуда, молим вас! Овуда!</p>

<p>Ништа се више не противи човеку. Не важи више закон теже ни старе мере и оцене за раздаљину и тврдоћу предмета. Све је измењено и померено.</p>

<p>Када би се испео на планину која се диже изнад овог места, могао би, чини му се, загазити право у небеска пространства. Све може да буде.</p>

<p>Ту негде би га из чудног сна пренуо глас његове жене која се враћа са шетње. Идући уз камене степенице, које су са улице водиле на терасу, она би га дозивала мирним гласом који очекује сигуран одговор.</p>

<p>- Фелд? Фелд!</p>

Word	Form	Distribution
зажмури	V+Form=P+Person=z+number=s+Distribution=Perf	на.INT
зажмури	V+Form=P+Person=z+number=s+Distribution=It	на.PREP+Distribution=p4
зажмури	V+Form=P+Person=z+number=s+Distribution=Iref	на.PREP+Distribution=p7
зажмури	V+Form=Y+Person=y+number=s+Distribution=Perf	на.PREF+V
зажмури	V+Form=Y+Person=y+number=s+Distribution=It	
зажмури	V+Form=Y+Person=y+number=s+Distribution=Iref	
зажмури	V+Form=A+Person=y+number=s+Distribution=Perf	

- ❑ Резултат морфолошке анализе (врсте речи, деклинације, коњугације, род, број,...)
- ❑ Разрешавање вишезначности (дезамбигуација)

Претрага регуларним изразом

Изразом **<N+NProp+First+m>**

траже се мушка лична имена

The screenshot shows the 'Concordance for Text na-Drini-cir.not' application. A search window titled 'Locate a pattern in na-Drini-cir' is open, showing the pattern '<N+NProp+First+m>' selected under 'a NooJ regular expression:'. The main window displays a concordance table with three columns: 'Text', 'Before', 'Seq.', and 'After'. The 'Text' column contains a snippet of text in Cyrillic, and the 'Seq.' column lists names. The 'After' column contains the full text snippet.

Text	Before	Seq.	After
једног корита у друго. Мајстор	Антоније	је довео из Далмације нарочите	
Убрзо је стигао и мајстор	Антоније	. Млади црнац је, после прве	
очи. Састављених обрва, блед, мајстор	Антоније	је издавао наређења како да	
остало под каменом блоком. Мајстор	Антоније	му је подигао над гробом	
прашта са Тосунефендијом и мајстор	Антонијем	. Сви стадоше да се разилазе	
и крупније речи похвали. Само	Ахмед	-ага шета, богат житарски трговац	
је написао неки цариградски стихотворац	Бади	, и који су казивали име	
кад је она завршена. Тај	Бади	је за добре паре писао	
се троши у такве сврхе.	Бади	, који је видео ово, кад	
и онај који ни на	Бајрам	није. До у села око	
највећу цркву саградио!" "Помози нас,	Боже	, ти Једини, потри душманина и	
вику: - Не мене, не мене!	Вилу	хватајте! Не, Абидага! Из последњих	
по Романији, овако учио Дијете	Грујицу	, кад је требало да га	
од детета па до старца.	Дан	који свиће недеља је. Недељом	
оно што је дању саграђено.	Док	није "нешто" проговорило из воде	
и Остоји, и сличним чудима.	Док	су трајале свечаности, као и	
Раде, сине, немој мајке заборавит'...	Илија	! Илија! Илија! - викала је друга	
- сине, немој мајке заборавит'...	Илија	! Илија! - викала је друга жена	
немој мајке заборавит'...	Илија	! - викала је друга жена, тражећи	
онај с тобом! - Звао се	Јован	, а ја му не знам	
од онда кад је Краљевић	Марко	тамновао горе у Старом граду	
да то није био Краљевић	Марко	нит' је могао бити (јер	
и оног ко је подиже.	Мехмед	-паша, који се целог века	
српски господари, Вино служи провизор	Мијајло	, А свијетли сестра Кандосија Са	
да привуче њихову пажњу. Неки	Мурат	, звани Мута, малоуман младић из	

Претрага граматиком

Претрага граматиком
Im_Prez_FM_all_Cir.nog
проналази имена и презимена

Structure of Im... [Expanded]

- titule
- IP_M_en_6
- titule
- IP_M_en_7
- titule
- IM_Prez_FM_sr
 - Im_Prez_F_sr
 - IP_F_sr_1
 - titule
 - IP_F_sr_2
 - titule
 - IP_F_sr_3
 - titule
 - IP_F_sr_4
 - titule
 - IP_F_sr_5
 - titule
 - IP_F_sr_6
 - titule
 - IP_F_sr_7
 - titule
 - Im_Prez_M_sr
 - IP_M_sr_1
 - titule
 - IP_M_sr_2
 - titule
 - IP_M_sr_3
 - titule
 - IP_M_sr_4
 - titule
 - IP_M_sr_5
 - titule
 - IP_M_sr_6
 - titule
 - IP_M_sr_7
 - titule

IM_Prez_FM_sr

Serbian/Serbian syntactic grammar consists of 36 graphs.

IP_F_sr_1

Serbian/Serbian syntactic grammar consists of 36 graphs.

Structure of Im... [Expanded]

- titule
- IP_M_en_5
- titule
- IP_M_en_6
- titule
- IP_M_en_7
- titule
- IM_Prez_FM_sr
 - Im_Prez_F_sr
 - IP_F_sr_1
 - titule
 - IP_F_sr_2
 - titule
 - IP_F_sr_3
 - titule
 - IP_F_sr_4
 - titule
 - IP_F_sr_5
 - titule
 - IP_F_sr_6
 - titule
 - IP_F_sr_7
 - titule
 - Im_Prez_M_sr
 - IP_M_sr_1
 - titule
 - IP_M_sr_2
 - titule
 - IP_M_sr_3
 - titule
 - IP_M_sr_4
 - titule
 - IP_M_sr_5
 - titule
 - IP_M_sr_6
 - titule
 - IP_M_sr_7
 - titule

IP_F_sr_1

Serbian/Serbian syntactic grammar consists of 36 graphs.

Structure of Im... [Expanded]

- titule
- IP_M_en_5
- titule
- IP_M_en_6
- titule
- IP_M_en_7
- titule
- IM_Prez_FM_sr
 - Im_Prez_F_sr
 - IP_F_sr_1
 - titule
 - IP_F_sr_2
 - titule
 - IP_F_sr_3
 - titule
 - IP_F_sr_4
 - titule
 - IP_F_sr_5
 - titule
 - IP_F_sr_6
 - titule
 - IP_F_sr_7
 - titule
 - Im_Prez_M_sr
 - IP_M_sr_1
 - titule
 - IP_M_sr_2
 - titule
 - IP_M_sr_3
 - titule
 - IP_M_sr_4
 - titule
 - IP_M_sr_5
 - titule
 - IP_M_sr_6
 - titule
 - IP_M_sr_7
 - titule

IP_F_sr_1

Serbian/Serbian syntactic grammar consists of 36 graphs.

Претрага граматиком

Concordance for Text na-Drini-cir.not

Reset Display: 5 characters before, 5 after. Matche Output

Text	Before	Seq.	After
знају да то није био	Краљевић Марко	нит' је могао бити (јер	
од Турака. Суљага Османагић, газда	Петар Богдановић	, Мордо Папо, поп Михаило, крупни	
Мула Исмет, вишеградски хоца, и	Елиас Леви	, звани Хаџи Лиачо, јеврејски хамбаша	
свакојаке смешне и необичне згоде.	Поп Михаило	и Хаџи Лиачо су предњачили	
дође више њиховог села злогласни	Јован Мићић	, рујански сердар, чак из Ариља	
ономе што нам ти свјетлујеш.	Али Осман	ефендија је био глух за	
на незастрти део каменог седишта.	Поп Никола	извади велику кожну дуванкесу и	
били сви већ у годинама.	Поп Никола	и Мула Ибрахим стари, а	
и смирење и нове опасности.	Поп Никола	је од све четворице био	
миран живот и поштен рад.	Поп Никола	је говорио кратко и завршио	
су да неће моћи остати.	Поп Никола	је долазио и свештао му	
седе и рачунају нешто газда	Павле Ранковић	, један од млађих али богатијих	
долази у механу често и	Франц Фурлан	са својим хармоникама. Он је	
имају посла и Сумбо и	Франц Фурлан	и Ђоркан и Шаха Циганка	
година, и то као иноча.	Хаџи Омер	је већ преко тридесет година	
Дуго их је држала нада.	Хаџи Омер	је и на Ђабу ишао	
сносили своју злу срећу и	Хаџи Омер	и његова паметна ханума, али	
се све сврши у реду.	Хаџи Омер	се опирао и тврдио дуго	
и по младости, не би	Хаџи Омер	имао што има и не	
посла, прође преко моста онај	Петар Талијан	, мајстор Перо, у свом платненом	
по један стижу на капију.	Јанко Стиковић	, син једног абације са Мејдана	
ватрен дебатар на ђачким састанцима.	Велимир Стевановић	, здрав и крупан младић, под својче	
срца и крије свако осећање.	Ранко Михаиловић	, ћутљив и доброћудан младић, који	
касаби. Ту је, најпосле, и	Владо Марић	, бравар по занимању, весела и	
млађих другова. Тома Галус и	Фехим Бахтијаревић	. Њих двојица се држе мало	
једног од првих газда, Хаџи-	Томе Станковића	, једном поуселом и снажном девојком	
смелостима мисли и претераностима лектире.	Фехим Бахтијаревић	је само по мајци касабалија	
пролазних жеља и људске сујете."	Али Бахтијаревић	није изговорио ниједну од тих	
се пијана и гласна препирка.	Никола Пецикоза	се кладио у два литра	
сандуку од ексера чучи сељак	Ибро Ђемаловић	, из Узавнице. Газда Санто сабира	
радио. Ти знаш да је	Богдан Ђуровић	, мој друг са Околишта, већ	
Одлучили смо да бјежимо. Вечерас.	Владо Марић	, са још двојцом. Мислим да	
Сутрадан се проучио да су	Владо Марић	Гласници и још неки младићи	

Query 39/39

Претрага граматиком
Im_Prez_FM_all_Cir.nog
проналази имена и презимена

Морфолошки речници

- ❑ Леме простих речи су облика:
lema, K+FLX=Knnn [+SinSem]*
 - **lema** је канонски облик речи, као у традиционалним речницима
 - **K** означава врсту речи
 - **nnn** је број или нека друга врста кода који означава класу лема које деле иста флективна својства
 - Трансдуктор **Knnn** генерише све облике парадигме леме
 - Низ опционих ознака **+SinSem** за синтаксичка, семантичка, деривациона и друга својства леме
- ❑ **девојчин, A+FLX=A1cir+Pos+Ek**
 - Придев флективне класе **A1cir**, присвојни (**+Pos**), екавски (**+Ek**).
- ❑ **девојка, N+FLX=N618cir+Hum+Ek**
 - Именица флективне класе **N618cir**, означава људско биће (**+Hum**) и екавског је изговора (**+Ek**).

Морфолошки речници

- ❑ С обзиром да српски језик карактеришу два званична писма, ћирилична и латиница, речници и граматике се припремају за оба писма
- ❑ $\text{степен}, N+FLX=N5_{\text{cir}}+Sr+Mes$
- ❑ $\text{прстен}, N+FLX=N89_{\text{cir}}$
- ❑ $\text{глумац}, \text{glumac}, N+FLX=N42_{\text{cir}}+Hum$
- ...
- ❑ $\text{stepen}, N+FLX=N5_{\text{lat}}+Sr+Mes$
- ❑ $\text{prsten}, N+FLX=N89_{\text{lat}}$
- ❑ $\text{glumac}, N+FLX=N42_{\text{lat}}+Hum$

Морфолошки речници

Entry	S-Lemma	Category	Distribution	DRV	FLX	SynSem
ако	ако	CONJ				
баш	баш	PAR				
без	без	PREP	-			
блед	блед	A	-		A11cir	
близанац	близанац	N	Hum		N42cir	
брже	брже	ADV	Adj+Comp			
глумац	glumac	N	Hum		N42cir	
договорити	договорити	V	Perf+Tr+Iref		V151cir	
дотрчати	дотрчати	V	Perf+It+Iref		V641cir	
жутокљунац	жутокљунац	N	-		N42cir	
зазвечати	зазвечати	V	-		V641cir	
зајечати	зајечати	V	-		V641cir	
затим	затим	ADV	-			
заћутати	заћутати	V	-		V641cir	
изнад	изнад	PREP	-			
корен	корен	N	-		N89cir	
луд	луд	A	-		A11cir	
магарац	магарац	N	-		N42cir	
млад	млад	A	-		A11cir	
нарушити	нарушити	V	-		V151cir	
оборити	оборити	V	-		V151cir	
отворити	отворити	V	Perf+Tr+Iref+Ref		V151cir	
прамен	прамен	N	-		N89cir	
прећутати	прећутати	V	Perf+Tr+Iref		V641cir	
прстен	прстен	N	-		N89cir	
решити	решити	V	-		V151cir	
степен	степен	N	Sr		N5cir	Mes
тврђ	тврђ	A	-		A11cir	

<http://www.m>

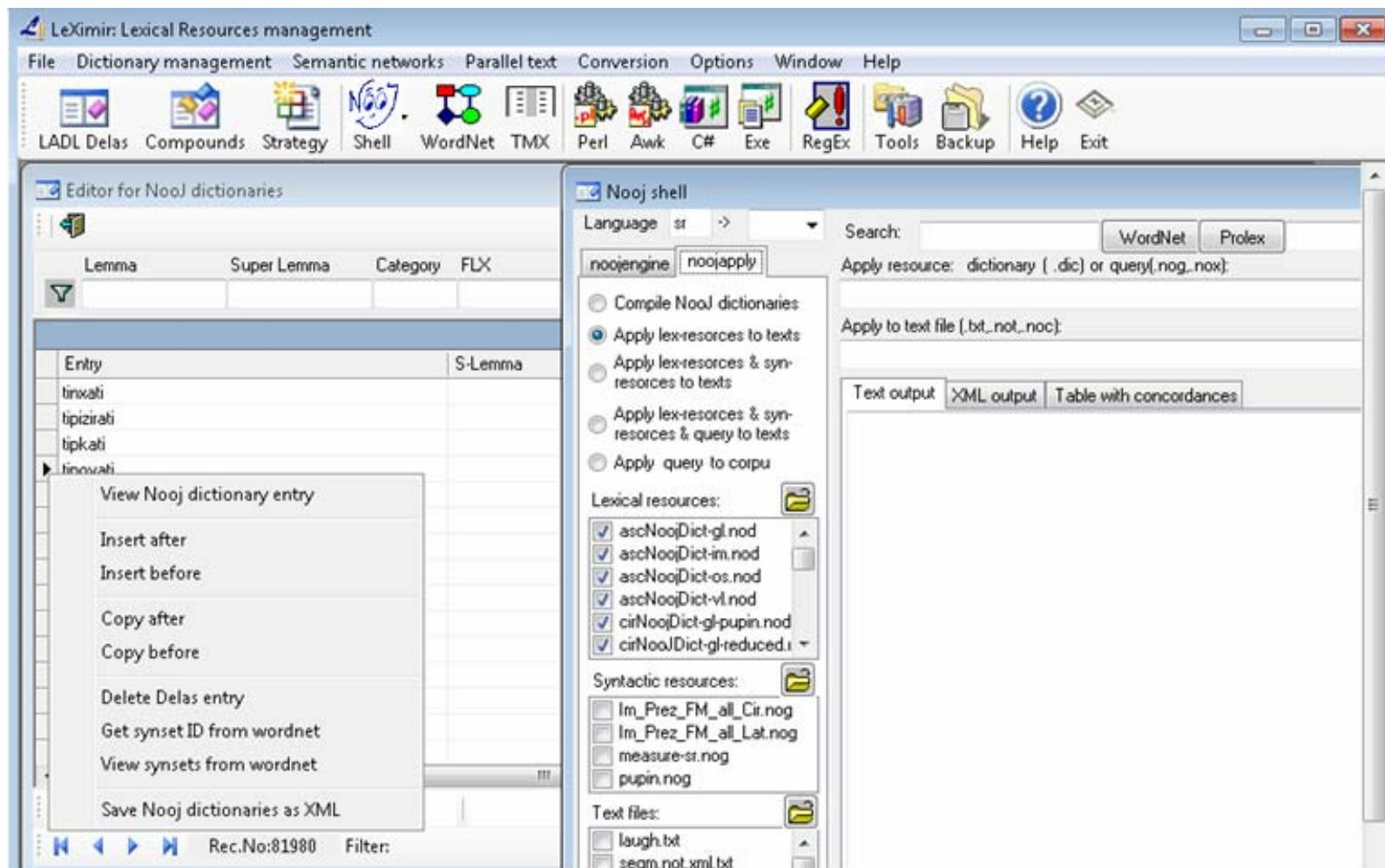
<http://cesar.matf.bg.ac.rs>

Дан језичких технологија Београд, Хотел Хајат 29. октобар 2012.

Нуџ и Лексимиr

- ❑ Нуџ обезбеђује рад са речницима, али не подржава управљање садржајем самих речника
- ❑ Речници се морају развијати независно, а коришћење уређивача попут Notepad-a је с обзиром на велики број ових ресурса и њихове димензије потпуно неадекватно
- ❑ У Лексимиру, алату за руковање лексичким ресурсима и за експанзију упита развијене су софтверске компоненте за унос, прегледање и ажурирање одредница које подржавају специфичности самих ресурса као и других софтверских алата који те ресурсе користе
- ❑ Лексимиr садржи и шкољку за коришћење Нуџаплај независне компоненте

Нуџ и Лексимиr



<http://www.meta-net.eu>
<http://cesar.matf.bg.ac.rs> Дан језичких технологија Београд, Хотел Хајат 29. октобар 2012.

Метаподаци

```
_properties.def

# NooJ V1
# Dictionary properties' definition
#
# Language is: sr
#
# Special Characters: '=' '+' '#' ' '
#
# List categories and properties associated with features
# Example: N_Number = m + f;
# Special KEYWORD: INFLECTION lists all inflectional features (used by variables)
# Example: INFLECTION = m + f + Present + Futur;
# ranka

INFLECTION = m | f | n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | s | p | w | v | q | g | i |

N_gender = m | f | n;
N_case = 1 | 2 | 3 | 4 | 5 | 6 | 7;
N_number = s | p | w;
N_animateness = v | q | g;

NUM_case = 1 | 2 | 3 | 4 | 5 | 6 | 7;
NUM_number = s | p | w;
NUM_gender = m | f | n;
NUM_animateness = v | q | g;

PRO_case = 1 | 2 | 3 | 4 | 5 | 6 | 7;
PRO_number = s | p | w;
PRO_gender = m | f | n;
PRO_animateness = v | q | g;
```

Морфолошке граматике

- $N_{5cir} = \langle E \rangle (\langle E \rangle / m+s+1+q + \langle E \rangle / m+s+4+q) +$
 $a(\langle E \rangle / m+s+2+q + \langle E \rangle / m+w+2+q + \langle E \rangle / m+w+4+q + \langle E \rangle / m+p+2+q) +$
 $y(\langle E \rangle / m+s+3+q + \langle E \rangle / m+s+7+q) +$
 $e(\langle E \rangle / m+s+5+q + \langle E \rangle / m+p+4+q) +$
 $om(\langle E \rangle / m+s+6+q) +$
 $i(\langle E \rangle / m+p+1+q + \langle E \rangle / m+p+2+q + \langle E \rangle / m+p+5+q) +$
 $ima(\langle E \rangle / m+p+3+q + \langle E \rangle / m+p+6+q + \langle E \rangle / m+p+7+q);$
 степен ...а ... у ... е ...ом ...и ...има
- $N_{5lat} = \langle E \rangle (\langle E \rangle / m+s+1+q + \langle E \rangle / m+s+4+q) +$
 $a(\langle E \rangle / m+s+2+q + \langle E \rangle / m+w+2+q + \langle E \rangle / m+w+4+q + \langle E \rangle / m+p+2+q) +$
 $u(\langle E \rangle / m+s+3+q + \langle E \rangle / m+s+7+q) +$
 $e(\langle E \rangle / m+s+5+q + \langle E \rangle / m+p+4+q) +$
 $om(\langle E \rangle / m+s+6+q) +$
 $i(\langle E \rangle / m+p+1+q + \langle E \rangle / m+p+2+q + \langle E \rangle / m+p+5+q) +$
 $ima(\langle E \rangle / m+p+3+q + \langle E \rangle / m+p+6+q + \langle E \rangle / m+p+7+q);$
 stepen ...а ... u ... е ...om ...i ...ima

Примери флективних класа

```
#####
# Samples of noun inflection classes (simple forms)
#####
N5cir = <E>(<E>/m+s+1+q + <E>/m+s+4+q)
+ a(<E>/m+s+2+q + <E>/m+w+2+q + <E>/m+w+4+q + <E>/m+p+2+q)
+ y(<E>/m+s+3+q + <E>/m+s+7+q)
+ e(<E>/m+s+5+q + <E>/m+p+4+q)
+ ом(<E>/m+s+6+q)
+ и(<E>/m+p+1+q + <E>/m+p+2+q + <E>/m+p+5+q)
+ има(<E>/m+p+3+q + <E>/m+p+6+q + <E>/m+p+7+q)
;

N89cir = <E>(<E>/m+s+1+q + <E>/m+s+4+q)
+ a(<E>/m+s+2+q + <E>/m+w+2+q + <E>/m+w+4+q + <E>/m+p+2+q)
+ y(<E>/m+s+3+q + <E>/m+s+7+q)
+ e(<E>/m+s+5+q + <E>/m+p+4+q)
+ ом(<E>/m+s+6+q)
+ ови(<E>/m+p+1+q + <E>/m+p+5+q)
+ ова(<E>/m+p+2+q)
+ ове(<E>/m+p+4+q)
+ овима(<E>/m+p+3+q + <E>/m+p+6+q + <E>/m+p+7+q)
+ и(<E>/m+p+1+q + <E>/m+p+5+q)
+ има(<E>/m+p+3+q + <E>/m+p+6+q + <E>/m+p+7+q)
;

N42cir = <E>(<E>/m+s+1+v)
+ <L><B><R>a(<E>/m+s+2+v + <E>/m+s+4+v + <E>/m+w+2+v + <E>/m+w+4+v)
+ <L><B><R>y(<E>/m+s+3+v + <E>/m+s+7+v)
+ <L><B><R>e(<E>/m+s+5+v + <E>/m+p+4+v)
+ <B2>че(<E>/m+s+5+v)
+ <L><B><R>ем(<E>/m+s+6+v)
+ <B2>ци(<E>/m+p+1+v + <E>/m+p+5+v)
+ a(<E>/m+p+2+v)
+ <B2>цима(<E>/m+p+3+v + <E>/m+p+6+v + <E>/m+p+7+v)
;
```

```
#####
# Samples of noun inflection classes (compounds)
#####
NC1cir = <P><B>a<N><E>/f+s+1+q
+ <P><B>e<N><B>e<E>/f+s+2+q
+ <P><B>o<N><B>и(<E>/f+s+3+q + <E>/f+s+7+q)
+ <P><B>y<N><B>y(<E>/f+s+4+q)
+ <P><B>a<N><B>o(<E>/f+s+5+q)
+ <P><B>ом<N><B>ом(<E>/f+s+6+q)
;
```

```
V641cir = <E>(<E>/W)
+ <B3>им(<E>/P+x+s)
+ <B3>иш(<E>/P+y+s)
+ <B3>и(<E>/P+z+s + <E>/Y+y+s)
+ <B3>имо(<E>/P+x+p + <E>/Y+x+p)
+ <B3>ите(<E>/P+y+p + <E>/Y+y+p)
+ <B3>е(<E>/P+z+p)
+ <B2>х(<E>/A+x+s)
+ <B2>(<E>/A+y+s + <E>/A+z+s)
+ <B2>смо(<E>/A+x+p)
+ <B2>сте(<E>/A+y+p)
+ <B2>ше(<E>/A+z+p)
+ <B2>вши(<E>/X)
+ <B2>о(<E>/G+m+s)
+ <B2>ла(<E>/G+f+s + <E>/G+n+p + <E>/G+w+m + <E>/G+w+n)
+ <B2>ло(<E>/G+n+s)
+ <B2>ли(<E>/G+m+p)
+ <B2>ле(<E>/G+f+p + <E>/G+w+f)
+ <B2>ху(<E>/F+x+s)
+ <B2>неш(<E>/F+y+s)
+ <B2>не(<E>/F+z+s + <E>/F+z+p)
+ <B2>немо(<E>/F+x+p)
+ <B2>хере(<E>/F+y+p)
;
```

```
#####
# Samples of noun derivation classes (derivation of simple forms from multiword unit)
#####
```

```
CGDrv = <P><B>o<S><R><C><RW><B>ац/N
```

;http://www.meta-net.eu

;http://cesar.matf.bg.ac.rs Дан језичких технологија Београд, Хотел Хајат 29. октобар 2012.

Сложене речи

- ❑ Црна Гора, N+FLX=NC1cir+DRV=CGDrv:N42cir
- ❑ Црна Трава, N+FLX=NC1cir+DRV=CGDrv:N42cir

CGDrv = <P>o<S><R><C><RW>aц/N

Оператори:

- ❑ <P> претходна реч
- ❑ брисање једног знака уназад
- ❑ <S> брисање текућег знака
- ❑ <R> померање у један знак десно
- ❑ <C> промена малог у велико слово и обрнуто
- ❑ < RW> померање на крај речи
- ❑ ... У тексту препознаје **црногорац...**

Сложене речи

❑ Претрага леме <Црна Гора>

na-Drini-cir.not [Modified]

5276 / 10216 TUs

Characters
Tokens
Digrams
Annotations

Language is "Serbian (sr)".
Text Delimiter is: \n (NEWLINE)
Text contains 10216 Text Units (TUs).

SHOW | EXT Annotation Structure

вођа, муслимана и Срба, избегло је у Црну Гору или у Турску. Остало је у овим крајевима још неколико хајдука, који у ствари нису ни имали праве везе са устанком због регрутације, него су хајдуковали за свој рачун. Па су и они похватани или растурени. Херцеговина се умирила. Босна је дала регруте без отпора. Али одлазак првих регрута из касабe није био лак ни једноставан. Из целог среза није узето више од стотињак младића, али оног дана кад су искупљени пред Конаком сељаци са својим торбама и ретки грађани са својим дрвеним сандуцима, изгледало је као да је у касаби помор и узбуна. Многи регрути су пили немилице још од раног јутра, мешајући пића. Сељаци су у чистим, белим кошуљама. Ретки су који не пију, него седе поред својих ствари, уза зид, и дремучкају. Већина је узбудена, румена од пића и знојна од врелог дана. По четири-пет момака из истог села загрле се, саставе главе и љуљајући се тако као жива љеса, пуштају своје грубо и отегнуто ојкање, као да су сами на свету:

37

+Distribution=p2 Црна Гора.N+gender=f+number=s+case=4+animateness=q

+Distribution=p4 црн.А+Degree=a+definiteness=k+gender=m+number=s+case=7+animateness=g+Distribution=Col

+Distribution=p7 црн.А+Degree=a+definiteness=k+gender=n+number=s+case=3+animateness=g+Distribution=Col

+V црн.А+Degree=a+definiteness=k+gender=n+number=s+case=7+animateness=g+Distribution=Col

црн.А+Degree=a+e+gender=f+number=s+case=4+animateness=g+Distribution=Col

Locate a pattern in na-Drini-cir

Pattern is:

☐ a string of characters:

☐ a PERL regular expression:

☒ a NooJ regular expression:

<Црна Гора>

☐ a NooJ grammar:

C:\Users\ranka\Documents\NooJ\

☐ Syntactic Analysis

Index

☐ Shortest matches

☒ Longest matches

☐ All matches

☒ Reset Concordance

Concordance for Text na-Drini-cir.not

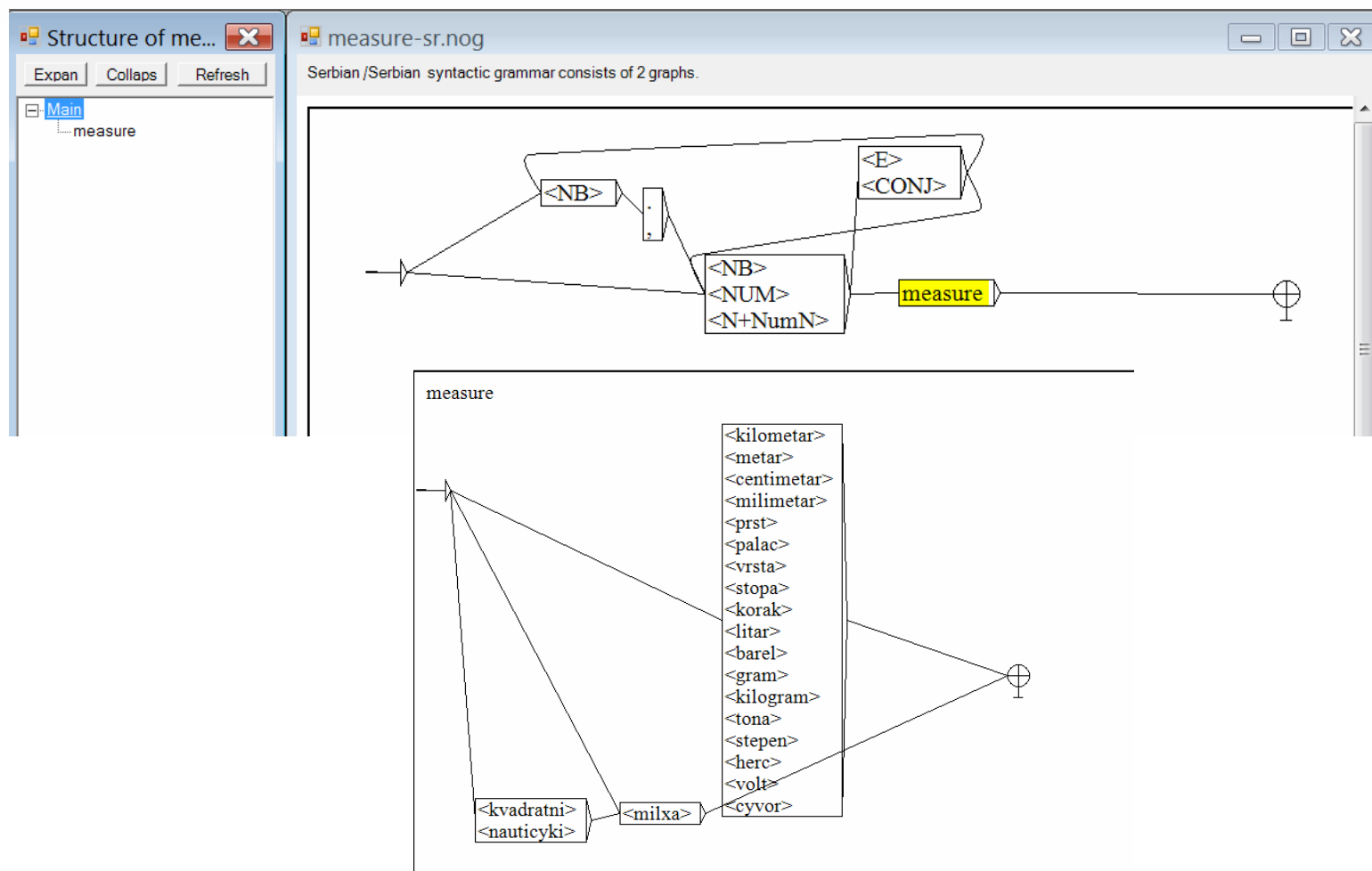
Reset Display: 5 characters before, 5 after. Matche Output

Text	Before	Seq.	After
њима се нашао однекуд неки	Црногорац		, ухватили га сејмени на друму
дубоког цепа свога сивог гуња	Црногорац		извлачи гусле, неутледне и малене
од Турака. Сви гледају у	Црногорца		као да га сад први
и нераван. Узбуђење расте. А	Црногорац		подешава и почиње кроз нос
причу. А у једном тренутку	Црногорац		заиста, пошто је како-тако
и мишићи на вилицама стежу.	Црногорац		везе и кити све брже
у Русији, а други на	Црној Гори		. Алихоца је још млад човек
и Срба, избегло је у	Црну Гору		или у Турску. Остало је
деси да наиђе гуслар, обично	Црногорац		неки, испоснички мршав, сиротињски одевен
Па се дигоше Србија и	Црна Гора		, дојучерашња раја, и узеше пола

http://www.meta-net.eu
http://cesar.matf.bg.ac.rs

Дан језичких технологија Београд, Хотел Хајат 29. октобар 2012.

Пример граматике



Шта има СрпНуџ за латиницу и ћирилицу?

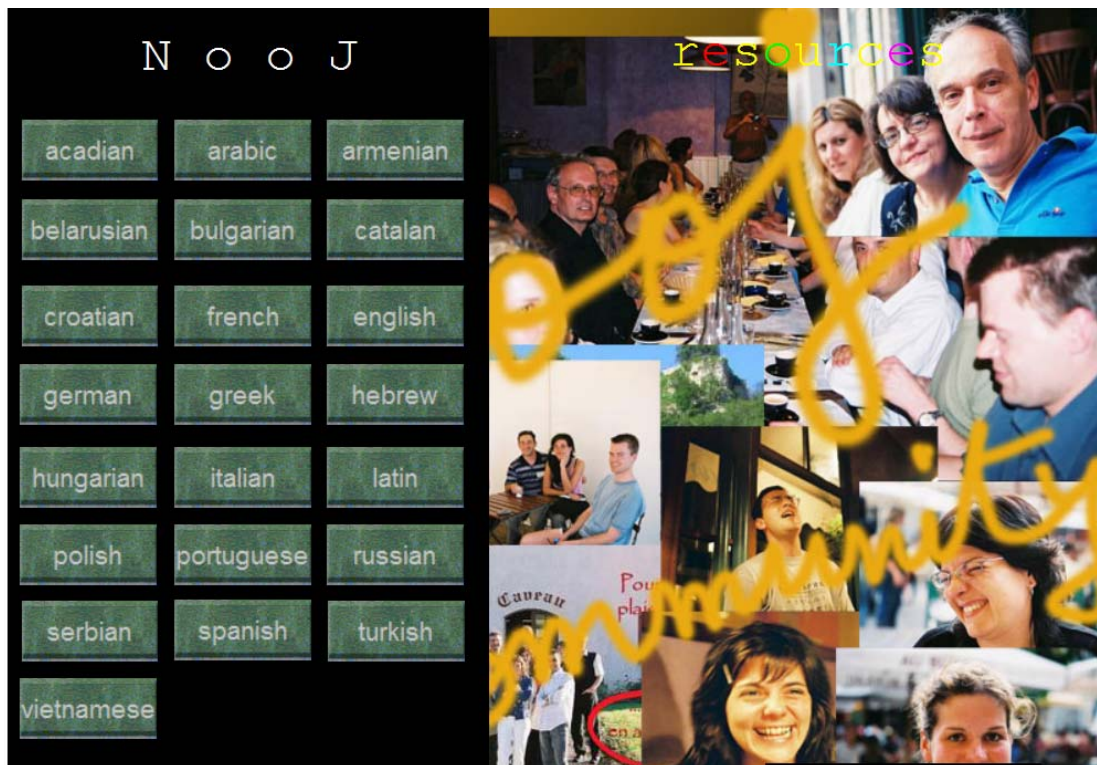
- ❑ Текст романа Два царства српског аутора Бранимира Ћосића са укупно 106684 токена
- ❑ Комплетан речник са 85868 одредница: именице (40886), придеви (25558), глаголи (15366) и остало (4058).
- ❑ Узорак речника у читљивом формату са 35 лема које представљају 9 врста речи, као и са примерима вишечланих лексема и регуларне деривације
- ❑ Узорак речника у читљивом формату са флективним облицима произведеним на основу лема и морфолошких граматика
- ❑ Узорак морфолошких граматика које се користе за аутоматско генерисање флективних облика лема датих у узорку речника
- ❑ Синтаксичка граматика која препознаје једну класу именованих ентитета – пуна лична имена са њиховим улогама и функцијама
- ❑ Метаподаци о граматичким и семантичким категоријама коришћеним у речницима и локалним граматикама

Шта још увек нема СрпНуџ?

- ❑ Лексикон-граматичке табеле
- ❑ Речник сложених речи (у припреми)
- ❑ Деривациону морфологију (у припреми)
- ❑ Фразе (дисконтинуалне)

Нуџ заједница

- Српски модул за Нуџ можете преузети са веб адресе <http://www.nooj4nlp.net/pages/resources.html>



Демонстрације информатичке подршке у обради српског језика

Hvala na pažnji!

Нуц ћете наћи на адреси

◆ <http://www.nooj4nlp.net>

Српски модул за Нуџ ћете наћи на адреси

◆ <http://www.nooj4nlp.net/pages/resources.html>

