



UNITEX

à l'Université de Belgrade

**Duško Vitas, Faculté de Mathématiques,
Cvetana Krstev, Faculté de Philologie,
Université de Belgrade
(vitas|cvetana)@matf.bg.ac.rs**



Le plan

- Quelques mots sur le groupe de Belgrade
- Quelques mots sur la langue serbe
- Unitex et les ressources pour le traitement du serbe
- Quelques applications utilisant Unitex



Le groupe



Le Groupe TAL à Belgrade

... est rassemblé à la Faculté de
Mathématiques de l'Université de
Belgrade





Les membres

... sont soit des chercheurs, soit des étudiants des facultés au sein de l'Université, surtout de la Faculté de Philologie



et les étudiants (de différents niveaux d'études)



Les recherches

... en Serbie sont organisées de telle manière que les chercheurs peuvent collaborer dans des projets hors de leurs institutions

Les projets du groupe TAL au niveau de la Serbie:

- ***Langue serbe et ses ressources***
- ***Infrastructure pour la formation en ligne en Serbie***



La langue serbe



Le serbe

Un principe d'orthographe phonétique:

graphème <--> phonème

a, A <--> /a/ comme [lak] dans *lac*

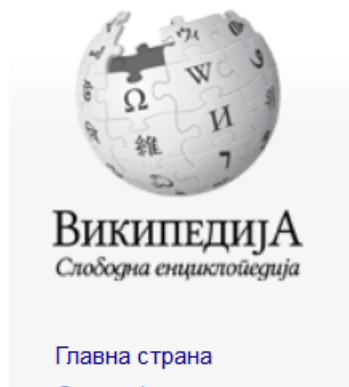
e, E <--> /e/ comme [de] dans *dé*



L'alphabet

Le serbe utilise deux alphabets:

- **cyrillique** (officiel) et
- **latin** (plus répandu)



Чланак [Разговор](#) [Кирилица](#) [Latinica](#)

Електронско учење

Из Википедије, слободне енциклопедије

Електронско учење се може дефинисати као процес преношења знања и вештина електронским апликацијама и процеси обухватају учење преко **веб-а**, **рачунара**, у дигиталним учионицама,



смјела унијети = smjela unijeti

[index - Tvorac Grada](#)

www.tvorac-grada.com/knjige/copic/sprava.html

Ni pod kakvim izgovorom u našu kuću nije se **смјела унијети** ни пуна ни празna, ни покварена, ни растављена. Од погледа на термометар дјед је ...

[Branko Copic :: Biblioteka ~ Library ~ Burek.mobi - Burek Forum](#)

forum.burek.com/branko-copic-t26098.5.wap2.html

Ni pod kakvim izgovorom u našu kuću nije se **smjela unijeti** ни пуна ни празna, ни покварена, ни растављена. Од погледа на термометар дјед је ...

[Cudna sprava Citati i po neka priča](#)

www.ivanminic.com/94-cudna-sprava/

10.03.2007. - Ni pod kakvim izgovorom u našu kuću nije se **smjela unijeti** ни пуна ни празna, ни покварена, ни растављена. Од погледа на термометар дјед је ...

[Kolači i slastice bez jaja - Stranica 6 - Forum.hr](#)

www.forum.hr > ... > [Kuhinja bez mesa](#) > Преведи ову страницу

16.07.2009. - ... to je bila trecina ukupnih masnoca koje trebao biti **smjela unijeti** ...

[\[PDF\] Export to PDF](#)

saturn.ffzg.hr/koha/index.../periodika.pdf?... > Преведи ову страницу

takva se niti ne bi **smjela unijeti** kao formula - tj. ako postoje bugovi u računanju formule, oni se trebaju pokušati ispraviti. • u slučaju da je numeracija različita za ...

[\[RTF\] Problemi s trenutnim stanjem obrade časopisa](#)

saturn.ffzg.hr/koha/index.cgi/periodika.rtf?... > Преведи ову страницу

takva se niti ne bi **smjela unijeti** kao formula - tj. ako postoje bugovi u računanju formule, oni se trebaju pokušati ispraviti. • u slučaju da je numeracija različita za ...

[Vinogradska - Stranica 294 - Forum.hr](#)

www.forum.hr > ... > [Roditelji i djeca](#) > [Trudnice](#) > Преведи ову страницу

22.01.2013. - Prvi put kad sam tamo radila (prije 2 i pol god) nisam **smjela unijeti** ništa, a muž je uspio samo fotic unjeti. Totalno glupo jer kad sam rodila i ...

[pitanja: pas i kreditna kartica za duza putovanja - Forum.hr](#)

www.forum.hr > [Lifestyle](#) > [Turizam](#) > Преведи ову страницу

10.04.2008. - ... su moji postovi gluposti. moja poznanica NIJE **smjela unijeti** psa od 2,5kg u vlak. mozda ima razlicitih konduktera pa samo nek bude spremna ...



Les dialectes (?)

A la base du BCMS est le dialecte **chtokavien** (prononciation du pronom interrogatif « *quoi* » : *što/šta*) avec trois variantes (les reflets du « *iat* »)

chtokavien (<i>la fille</i>)	
ékavien	<i>de</i> vojka
iékavien	<i>dje</i> vojka; <i>de</i> vojka
ikavien	<i>di</i> vojka



Les variants serbes

Les Serbes utilisent ékavien et iékavien.

	chtokavien	ékavien	iékavien
<i>La fleur</i>	singulier	<i>cvet (e longue)</i>	<i>cvijet</i>
	pluriel	<i>cvetovi (e courte)</i>	<i>cvjetovi</i>

Les dictionnaires (et les textes) du serbe reproduisent toutes ces variations.

13



Dérivation régulière

profesor,N:m

(le professeur)



lemmes dérivés



profesor+k+a,N:f

profesor+ic+a,N:f

profesor+k+in,A+Poss

profesor+ič+in,A+Poss

profesor+k+ica,N+Dem

profesor+k+ič+in,A+Poss



profesor,N:m

profesor+ov,A+Poss

profesor+ski,A+Rel

profesor+čić,N+Dem:m

profesor+če,N+Dem:n

profesor+če+t+ov,A+Pos

profesor+čina,N+Aug:f



Unitex et les ressources pour le traitement du serbe



Les dictionnaires



Les dictionnaires

DELAS

otac, **N118a**+Hum

muslinski, **A2**+PosQ+Mat

otpozdraviti, **V651**+Perf+It+Iref

sporiye, **ADV**+Adj+Comp

ispod, **PREP**+p2

dvojica, **N623**+NumN+MG+Pl

ponešto, **PRO13**+Indef+ProN

zanovijet, **N688**+Bot+Ijk

eksportovati, **V18**+Imperf+Perf+Tr+Iref+DerOvIr

Les dictionnaires

DELAS

otac,N118a+Hum

muslinski, **A2+**PosQ+Mat

otpozdraviti, V651+Perf+It+Iref

sporije, **ADV**+**Adj**+**Comp**

ispod, **PREP**+p2

dvojica, **N623**+NumN+MG+PI

ponesto, **PRO13**+Indef+ProN

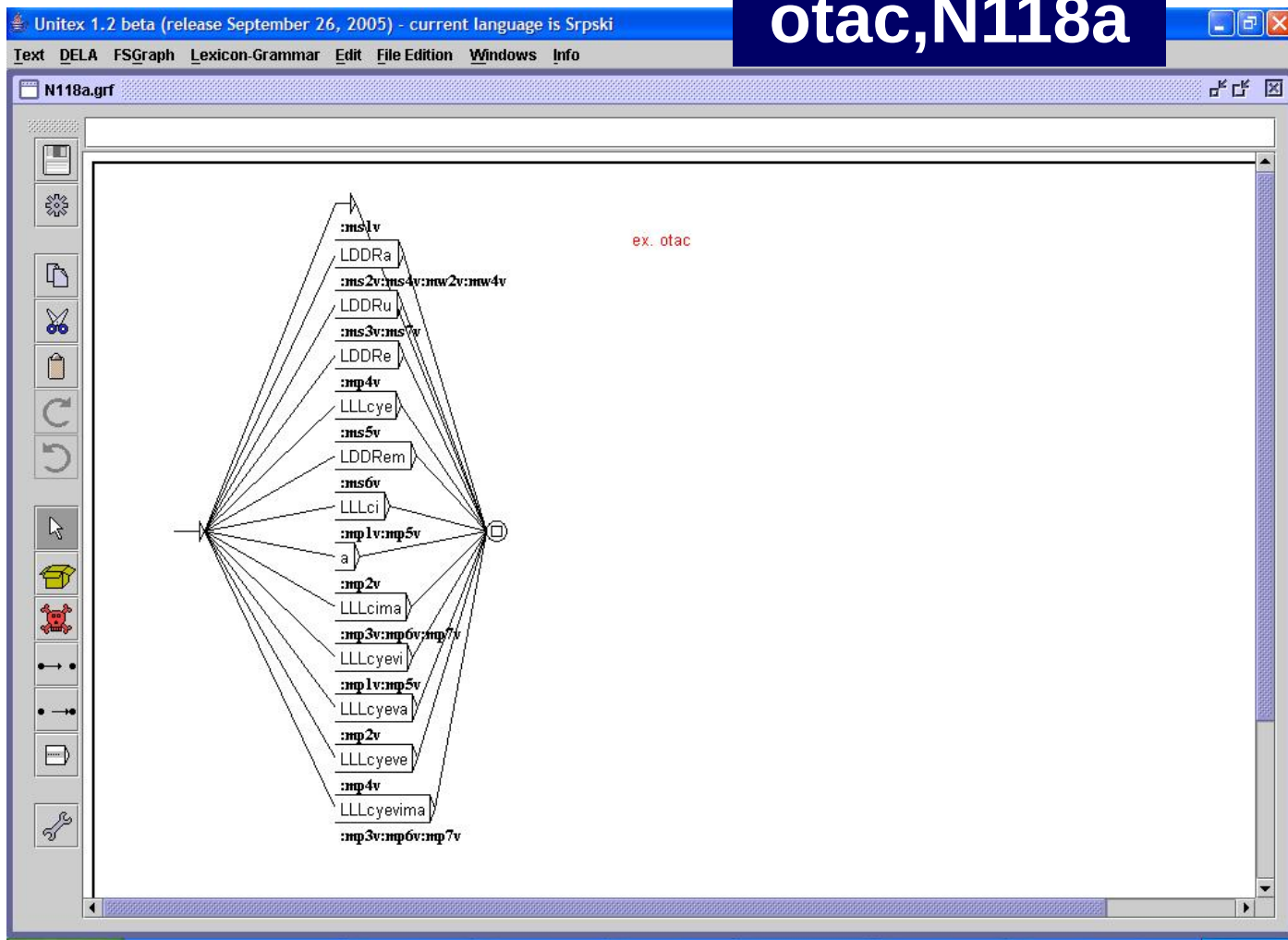
zanovijet,N688+Bot+Ijk

eksportovati, **V18**+Imperf+Perf+Tr+Iref+DerOvIr



Description formalisée

otac, N118a





Flexion des mots simples

- 337 graphes pour les substantifs
 - 348 graphes pour les verbes
 - 63 graphes pour les adjectifs
-
- **130,500 lemmes**
 - **4.500.000 formes fléchis**



DELA

glavama, glava.N: **fp3q:fp6q...**

pripove**ci**, pripovet**ka**.N: **fs3q...**

očev, očev.A: **akms1g:aems4q**

tvrđih, tvrd.A: **bemp2g:befp2g:...**

gdekoja, gdekoji.PRO+ProA+Indef+Ek: **fs1g:np1g:np**

devetih, deveti.A+Ord: **aemp2g:aefp2g:aenp2g...**



Exemple de *Locate pattern*

<N+Ijk-Ek>

kandidat bio je izložen prim*i*eni grube sile, a u Centru bezbjed
ala i izvršene ustavne prom*i*ene i, što je još značajnije, od
ore u našoj zemlji. "R*i*ecy je o parlamentarcima iz onih drz
leđa - tradicije, jezika i v*i*ere, odnosno svega što nas vjekovi
Mišković, iz Nikšićke zxel*i*ezare, istakao je da je zbog nedost



Les couleurs (A+Col)

Je voudrais voir **une espèce si rare**, dit
Martin = *corbeau blanc* = *mouton noir*

{S}Voleo bih da vidim tu belu vranu reče Martin. {S}Kandid
odmah

otidu koje su čuvali dva crna evnuha i dvadeset vojnika.

{S}Neobi

o svi mi deca Adamova, i crni i beli. {S}Ja se ne razumem u rodo
da taj ovan mora da bude crven i da uquine od ovčjih boginja.

{S}A

k bile ovako izgrizene i crvenih ivica, niti mi se nos spuštao
eče, opazi da nešto jako crveno pliva nedaleko od njegovog
broda.

ti, kad sam ponovo našao crvenog ovna i paketu, da opet vidim
e, koji su se nalazili u najcrnijoj bedi, pristaju uz obalu kraj
e krčme. {S}Dva čoveka u plavom primetiše ga: {S}Druže, reče
jeda

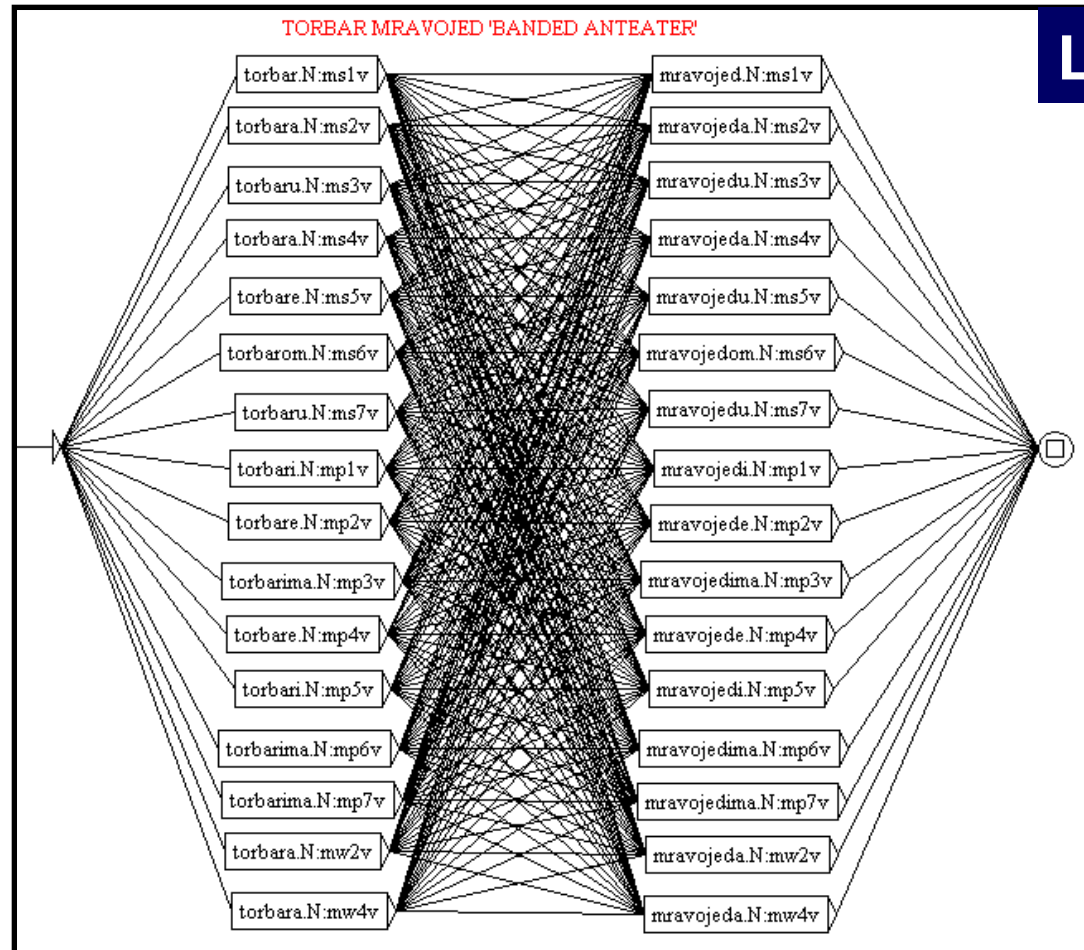
crnci protiv polucrnih, polucrni protiv zelenih, melezi proti
kamičci, žuti, crveni i zeleni, i čudno demi-noir stali.

{S}Putni

vrlo lepim stubovima od zelenog i zlatnog mramora i rešetkama u



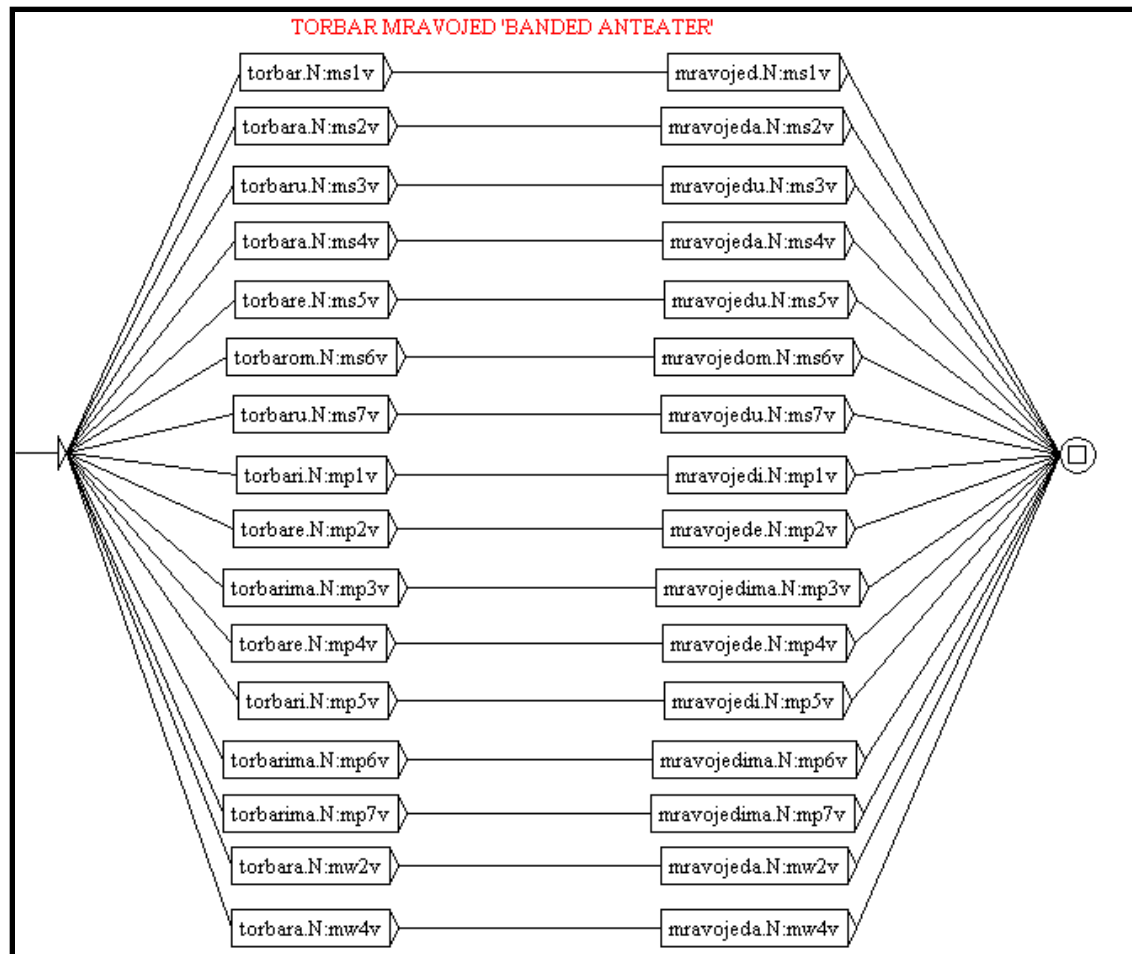
Les mots composés (N:N)



Le numbat

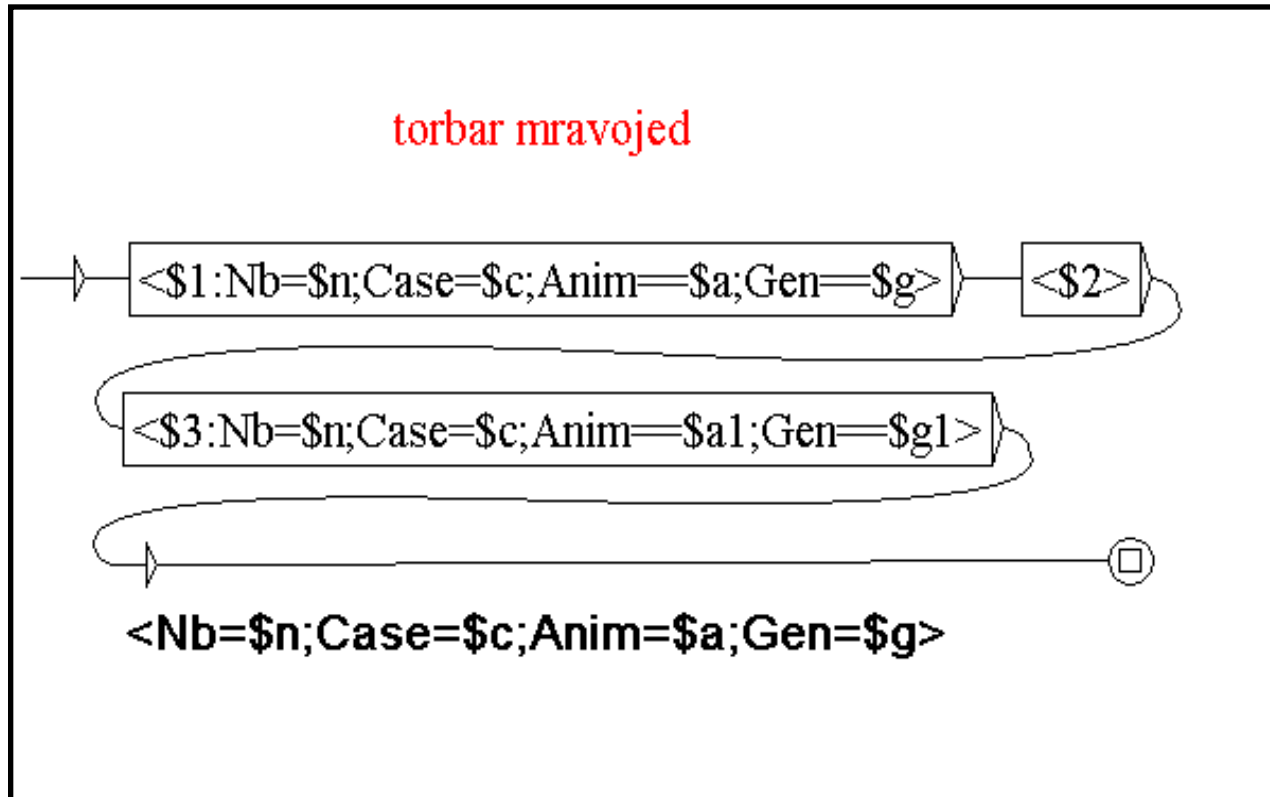


Les mots composés





Une solution (A. Savary)

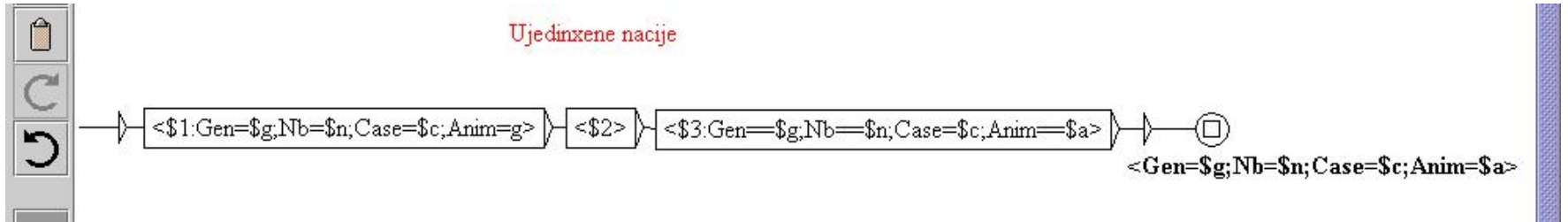




Les mots composés

Ujedinjene nacije A:p N:p (Nations unis)

Ujedinjene nacije





Le résultat

Unitex 1.2 beta (release September 26, 2005) - current language is Srpski

Text DELA FSGraph Lexicon-Grammar Edit File Edition Windows Info

Locate Pattern

Locate pattern in the form of:

☒ Regular expression:

<Ujedinxene nacije.N>

☐ Graph:

Index

☐ Shortest matches

☒ Longest matches

☐ All matches

Grammar outputs

☒ Are not taken

☐ Merge with in

☐ Replace recog

Search limitation

☒ Stop after 200 matches

☐ Index all utterances in text

Concordance: C:\Cvetana\MojUnitex\Srpski\Corpus\ekonomist_all_snt\concord.html

20 matches

☒ Enable links ☐ Allow concordance edition

ericykih saveznih drzava. {S}NxUJORK _ [Ujedinxene nacije](#) su uputile hitan apel za pomocx od 50 ericykih saveznih drzava. {S}NxUJORK _ [Ujedinxene nacije](#) su uputile hitan apel za pomocx od 50 ure. {S}Svi materijalne obaveze za koje [Ujedinxene nacije](#) terete bivsxu SFR Jugoslaviju, do dan ektivama 2004. koje su u petak objavile [Ujedinxene nacije](#). {S}Jak tempo oporavka, zapocyet u dr onstrukciju i veliku pomocx od programa [Ujedinxenih nacija](#) za zxivotnu sredinu _ UNEP-a koji je abilnost jugoistocoyne Evrope i Programa [Ujedinxenih nacija](#) za razvoj (UNDP) u Beogradu javnosti j ratifikaciji kod generalnog sekretara [Ujedinxenih nacija](#). {S} Do sada su Sporazum ratifikovale [Ujedinxenih nacija](#) za trgovinu i razvoj, razmatraju nac [Ujedinxenih nacija](#), planiranih da budu ostvareni do 201 ink, na kome rade ekonomski strucymxaci [Ujedinxenih nacija](#) i Univerziteta Toronto, koja je prav li drugi dokument izdat na osnovu normi [Ujedinxenih nacija](#), odlucyila je crnogorska vlada. {S}0 . {S}2002. Ekonomija Vodecki ekonomisti [Ujedinxenih nacija](#) tvrde da je prosxle godine svetska e u istocynoj Evropi, pokazuje izvesxtaj [Ujedinxenih nacija](#), koji cxe ove nedelxe biti objavlxen , Milford Bejtmen iz Programa za razvoj [Ujedinxenih nacija](#) (UNDP) je na pomenutoj debati rekao anizator skupa bio je Program za razvoj [Ujedinxenih nacija](#). {S}Jugoslovenska vlada usvojila je lxavanxa, koje josx uvek pod patronatom [Ujedinxenih nacija](#) za razvoj priprema tim Ministarstva zasxtite prirode. {S}Generalni sekretar [Ujedinxenih nacija](#) Kofi Anan pozvao je sve Vlade, kompa u napusxta SCG Svetski program za hranu [Ujedinxenih nacija](#) najavio je kraj svoje misiye u Srbij ferenciji za sxtampu u becykom sedisxtu [Ujedinxenih nacija](#) rekao da je tek sada ICPDR "kompletn ferenciji za sxtampu u becykom sedisxtu [Ujedinxenih nacija](#) rekao da je tek sada ICPDR "kompletn

DLC: 80 compound lexical entries

Sremska Mitrovica,.NC_AXN3+

Sremskoj Mitrovici,Sremska

Sremskoj Mitrovici,Sremska

sudija okruzxnog suda,.NC_M

sudija okruzxnog suda,.NC_M

sudija okruzxnog suda,.NC_M

Ujedinxene nacije,.NC_AXN3+

Ujedinxene nacije,.NC_AXN3+

Ujedinxene nacije,.NC_AXN3+

Ujedinxenih nacija,Ujedinx

UN\ -a,UN.N+ABB:ms2q

vojno\ -tehnicyke,vojno-tehr

ABF

ABN

Abramovicx

ABS

Acquis

NC_AXN1.grf ...



Delac

- ~ **11.500** lemmes
- ~ **200,000** des formes fléchis
 - ~ **116** graphes de flexion
- Un système de prévision des classes!



Gestion des dictionnaires

LeXimir

LeXimir: Lexical Resources management - [Editor for LADL Delas dictionaries: sr]

File Dictionary management Semantic networks Parallel text Conversion Options Window Help

LADL Delas Compounds Strategy Shell WordNet TMX Perl Awk C# Exe RegEx Tools Backup Help Exit

New Before After Edit View Sort Dict Delete Reflect Statistics WordNet Save Save LMF Exit

POS Lemma FST SinSem Lemma inverted

cvijet FR All

POS	Lemma	FSTCode	SinSem	Comment	Lemmalnv	WN_Sinset	Dict.File	Di
N	cvijetnxak	N9	+ljk		kaxntejivc	(null)	delas-im.dic	4
N	cvijet	N2081	+ljk		tejivc	(null)	delas-im.dic	4

*

DelafTmp.dic - WordPa

Home View

Courier New 11

Paste

Clipboard

Font

1 2 3 4 5 6

cvijet, cvijet.N:ms4q
cvijet, cvijet.N:ms1q
cvijeta, cvijet.N:mw4q
cvijeta, cvijet.N:mw2q
cvijeta, cvijet.N:ms2q
cvijetu, cvijet.N:ms7q
cvijetu, cvijet.N:ms3q
cvijete, cvijet.N:ms5q
cvijetom, cvijet.N:ms6q
cvjetovi, cvijet.N:mp5q
cvjetovi, cvijet.N:mp1q
cvjetova, cvijet.N:mp2q
cvjetove, cvijet.N:mp4q
cvjetovima, cvijet.N:mp7q
cvjetovima, cvijet.N:mp6q
cvjetovima, cvijet.N:mp3q



Gestion des dictionnaires

LeXimir: Lexical Resources management - [Editor for LADL Delas dictionaries: sr]

File Dictionary management Semantic networks Parallel text Conversion Options Window Help

LADL Delas Compounds Strategy Shell WordNet TMX Perl Awk C# Exe RegEx Tools Backup Help Exit

New Before After Edit View Sort Dict Delete Inflect Statistics WordNet Save Save LMF Exit

POS Lemma FST SinSem Lemma inverted

cvijet FR All Exact Match

POS	Lemma	FSTCode	SinSem	Comment	LemmaInv	WN_Sinset	Dict.File	DictL	LemmaL
A	Anzxeankin	A1	+Pos+NProp+Hum+In		niknaexznA	(null)	delas-top.dic	13	13240
A	Anzxejcyev	A1	+Pos+NProp+Hum+In		veycjexznA	(null)	delas-top.dic	13	13250
A	Arzxantejcyev	A1	+Pos+NProp+Hum+In		veycjtnaxzrA	(null)	delas-top.dic	13	13330
A	Arzxantejkin	A1	+Pos+NProp+Hum+In		nikjetnaxzrA	(null)	delas-top.dic	13	13340
A	Bordovcyev	A1	+Pos+NProp+Hum+In		veycvodroB	(null)	delas-top.dic	13	13600
A	Brestlxaninov	A1	+Pos+NProp+Hum+In		voninaxltserB	(null)	delas-top.dic	13	13690
Francuzov		A1	+Pos+NProp+Top+Inh		vozucnarF	(null)	delas-top.dic		
Klermonferancyev		A1	+Pos+NProp+Hum+In		veycnarefnomrelK	(null)	delas-top.dic		
Klermonferankin		A1	+Pos+NProp+Hum+In		niknarefnomrelK	(null)	delas-top.dic		
N	Lemanac	N42	+NProp+Hum+Inh+F		canamel	(null)	delas-top.dic	13	14900
N	Lemanka	N661	+NProp+Hum+Inh+FR		aknameL	(null)	delas-top.dic	13	14910
A	Limoxyzankin	A1	+Pos+NProp+Hum+In		niknaxzomiL	(null)	delas-top.dic	13	15010
A	Limoxyzancyev	A1	+Pos+NProp+Hum+In		veycnaxzomiL	(null)	delas-top.dic	13	15020
A	Marsejcyev	A1	+Pos+NProp+Hum+In		veycjesraM	(null)	delas-top.dic	13	15220
A	Marsejkin	A1	+Pos+NProp+Hum+In		nikjesraM	(null)	delas-top.dic	13	15230
A	Miluzxaninov	A1	+Pos+NProp+Hum+In		voninaxzuliM	(null)	delas-top.dic	13	15340
A	Miluzxankin	A1	+Pos+NProp+Hum+In		niknaxzuliM	(null)	delas-top.dic	13	15350
A	Nansijcyev	A1	+Pos+NProp+Hum+In		veycjisnaN	(null)	delas-top.dic	13	15460
A	Nansijkin	A1	+Pos+NProp+Hum+In		nikjjsnaN	(null)	delas-top.dic	13	15470
A	Nicxaninov	A1	+Pos+NProp+Hum+In		voninaxciN	(null)	delas-top.dic	13	15510
A	Nicxankin	A1	+Pos+NProp+Hum+In		niknaxciN	(null)	delas-top.dic	13	15520
A	Rencyev	A1	+Pos+NProp+Hum+In		veycneR	(null)	delas-top.dic	13	15870
A	Renkinxin	A1	+Pos+NProp+Hum+In		nixnikneR	(null)	delas-top.dic	13	15880

SinSem LIKE FR and SinSem LIKE Inh

Rec.No:80

22:03 6.10.2013



Les entités nommées: Cassys



nedelju, 21. maja
11. maja
aprilu 2006

juna 2010.

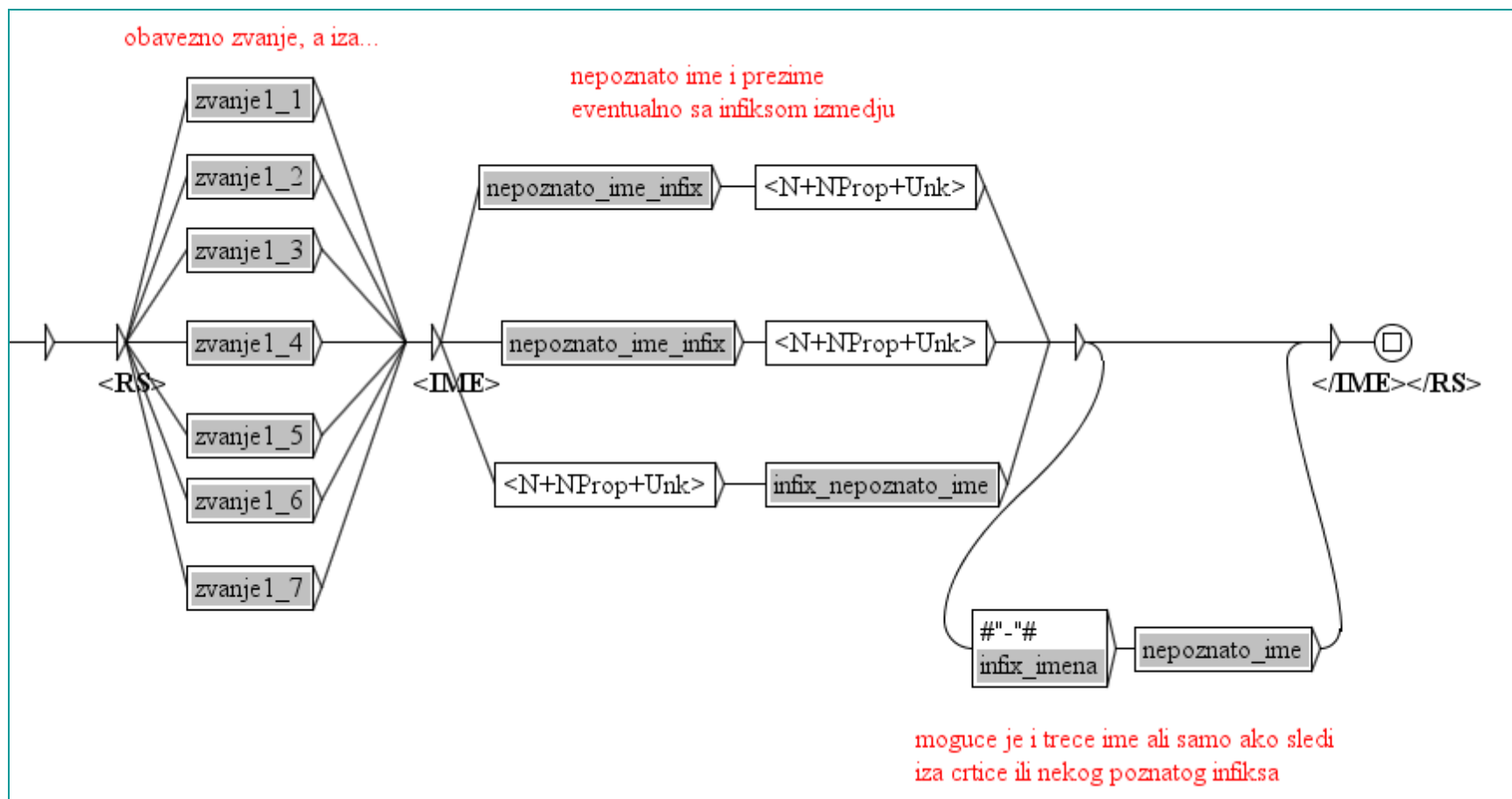
martu prošle godine
četvrtak, 17. juna 2010.

četvrtak, 27. juna ove godine

četvrtak (17. VI 2010.)



Les noms propres





Un exemple

<RS>Savetnici premijera i predsednika
 <TOP TYPE="Pays">Srbije</TOP>,
 <NP>Vladeta Janković</NP> i
 <NP> Dušan Bataković</NP>

</RS>,

ocenili su da su razgovori najviših predstavnika

<TOP TYPE="Ville">Beograda</TOP> i
<TOP TYPE=" Ville ">Prištine</TOP> u
<TOP TYPE=" Ville ">Beču</TOP>

o budućem statusu Kosova i Metohija važni jer se prethodnih

<TIME>šest meseci</TIME> razgovaralo samo o nestatusnim pitanjima,
prenosi Tanjug.

Advisors of the Prime Minister and the President of Serbia, Vladeta Janković and Dušan Bataković, assessed that the talks between the highest representatives of Belgrade and Pristina in Vienna on the future status of Kosovo and Metohija are important because during the previous six months only non-statutory issues were discussed, according to Tanjug.





Les corpus

Le corpus du serbe contemporain

<http://www.korpus.matf.bg.ac.rs/korpus/login.php>

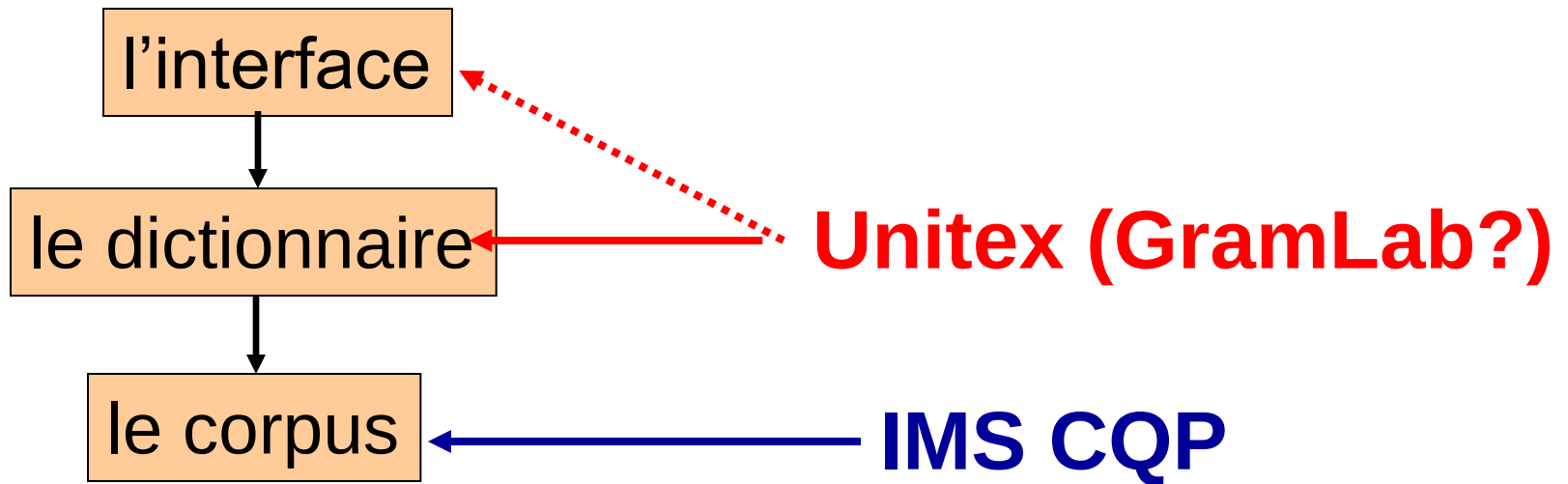
(« style belgradois »)

Logiciel: **IMS CWB/CQP** (Open CWB)

Étiquetage: **TreeTagger** avec **DELAS**



L'avenir (M. Utvić)





Les corpus et les textes alignés



Acide

ACIDE

Alignment Tools TMX Window Help

Concordance TMX

Loria

es-fr_fs.xml

File Edit View Help

Source ID's -	Target ID's -
n203	n203
n159	n159
n122	n122
n85	n85
n239	n239
n202	n202
n158	n158
n121	n121
n84	n84

391 concordances total, 391 displayed

Search

source:

target:

Operators: ☒ And ☐ Or ☐ Case sensitive



Acide → Unitex





Un exemple

C:\Cvetana\MojUnitex\XAlign\verne-fr-sr.xml

2554	Puis, des frères quêteurs, des pèlerins en longues robes, de simples civils, chevelure lisse et d'un noir d'ébène, tête grosse, buste long, jambes grêles, taille peu élevée, teint coloré depuis les sombres nuances du cuivre jusqu'au blanc mat, mais jamais jaune comme celui des Chinois, dont les Japonais diffèrent essentiellement.		Zatim kaludxeri koji mole za milostinxu, hodocynasnici u dugim halxinama, obicyni gradxani sa zalizanom kosom crnom kao abonos, velikom glavom, dugulxastim licem, tankih nogu, malog rasta, boje kozxe od bakarnosmedxe do svetlosmedxe, ali nikad zxute kao u Kineza, od kojih se Japanci bitno razlikuju.	2565
2662	Un autre, avec la fumée odorante de sa pipe, traçait rapidement dans l'air une série de mots bleuâtres, qui formaient un compliment à l'adresse de l'assemblée.		Drugi jedan, mirisavim dimom svoje lule brzo je pisao u vazduhu niz plavicvastih slova koja su izrazxavala pohvalu publici.	2673
2827	Plus de sombreros, plus de chemises rouges à la mode des coureurs de placers, plus d'Indiens emplumés, mais des chapeaux de soie et des habits noirs, que portaient un grand nombre de gentlemen doués d'une activité dévorante.		Nema visxe sombrera, ni crvenih kosxulxa koje su nosili kopacvi zlata, ni perjem, a i crnih c uzxurbani	
2835	Mr. Fogg et Mrs. Aouda s'installèrent devant une table et furent abondamment servis dans des plats lilliputiens par des Nègres du plus beau noir.		Gospodin sedosxe z najlepsxe posluzzxis tanxirime	
2874	Le nombre des chapeaux noirs diminuait à vue d'oeil, et la plupart semblaient avoir perdu de leur hauteur normale.		Crnih sxe a vecxine da je izo	
2901	Un énorme gaillard à barbiche rouge, au teint coloré, large d'épaules, qui paraissait être le chef de la bande, leva son formidable poing sur Mr.			

☐ All sentences/Plain text
☐ Matched sentences
☐ All sentences/HTML
☒ Aligned with target concordance

☐ Matched sentences
☐ All sentences/HTML
☐ Aligned with source concordance

Locate... Clear alignment Align Save alignment Save alignment as... Locate...

<A+Col>

crn - noir

bakarnosmed –

sombres nuances de cuivre

svetlosmed – blanc mat

žut - jaune



- # Textes sont alignés au

niveau des <seg>

Frankor)



Exemple de SrpFranKor (IMS)

publicit[a-z]*

SR: Zanat reklamiranja se toliko približio savršenstvu , da će biti teško usavršiti ga . Samuel Džonson , 1759

984476: CONTRAIREMENT aux idées reçues , la <publicité> et ses ficelles sont anciennes (* * * 1) .

SR: Nasuprot opšte usvojenom mišljenju , reklamiranje i njegova lukavstva su veoma stari

989860: La <publicité> joue donc le rôle essentiel de modéliser les besoins et les attentes des individus en fonction de la demande économique .

SR: Prema tome , reklama ima odlučujuću ulogu : ona oblikuje prohteve i očekivanja jedinki , stavljajući ih u funkciju ekonomskih zahteva

988143: La <publicité> la plus à la page admet en effet que quelque chose ne va pas dans notre existence , que le marché ne nous a pas donné tout ce qu ' il promettait , qu ' il n ' a pas résolu les problèmes issus du développement capitaliste .

SR: jer poslednji reklamni hitovi priznaju da nešto u našim životima nije kako bi trebalo da bude , da nam tržište nije dalo sve ono što je obećavalo , da nije rešilo probleme proistekle iz daljeg razvoja kapitalizma

984904: Le XXe siècle , avec la multiplication des médias électriques (cinéma , radio) , électroniques (télévision) et numériques (Internet) , a vu non seulement l ' explosion de la <publicité> mais aussi sa sophistication .

SR: XX vek , sa umnožavanjem električnih (kinematografija , radio) , elektronskih (televizija) i numeričkih medija (Internet) , doveo je ne samo do eksplozije reklame , nego i do njenog usložnjavanja

990218: La <publicité> mensongère est assurément interdite .

SR: U reklamama nije dozvoljeno lagati

992424: La <publicité> mène ainsi sa guerre de position sur le terrain de notre vie quotidienne .

SR: Tako reklamna delatnost vodi svoj pozicioni rat na terenu našeg svakodnevnog života

963594: Et font parfois passer pour des pisse - froid tous ceux qui , simplement , rappellent que , sous ses dehors aguichants , la <publicité> n ' est souvent qu ' une propagande , une véritable machine de guerre idéologique au service d ' un modèle de société fondée sur le capital , le marché , le commerce et la consommation .

SR: Ponekad deluju čarolizavo svi oni koji samo podsećaju da je ispod koketne spoljašnosti reklama često nuka propaganda prava mašinerija za



- 45





NERanka

<http://hlt.rgf.bg.ac.rs/VeBranka/NERanka.aspx>

HLT Group, language resources demo - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Named Entity Recognition x HLT Group, language resources demo x +

hlt.rgf.bg.ac.rs/VeBranka/NERanka.aspx

Most Visited Getting Started

Home > Cascade texts > Upload text/XML document

NERanka - Named Entity Recognition and Annotation Tool

Select language: serbian

Select name for annotation: ☒ All named entities

☐ person
☐ organization
☐ location
☐ amount
☐ time
☐ unknown

Click the button Browse to select a file for upload

Browse... No file selected.

Upload file and apply cascade graph

File URL (<http://mysite.com/myfile.txt>):
<http://www.rgf.rs/>
Download and parse web page

Download xml file

Časovnik tek što je označio ponoć, a Beograd se već ponosio najlepšim mogućim novogodišnjim poklonom - bratom i sestrom blizancima, prvim bebama rođenim u ovoj godini. Milka Ivković, srećna mama, rodila je prvo dečaka Aleksu minut posle ponoći, a potom i devojčicu Anđelu u nula časova i dva minuta.

Apply cascade graphs to text

Text Processed. You can download result.



NERanka

Untitled - Notepad

File Edit Format View Help

Časovnik tek što je označio ponoć, a
<top.gr>Beograd</top.gr> se već ponosio najlepšim
mogućim novogodišnjim poklonom - bratom i sestrom
blizancima, prvim bebama rođenim u <time.date.rel>ovoj
godini</time.date.rel>. {S} <pers><persName.full>Milica
Ivković</persName.full><role>, srećna
mama</role></pers>, rodila je prvo dečaka Aleksu
<time.hour.abs>minut posle ponoći</time.hour.abs>, a
potom i devojčicu Anđelu u nula časova i
<time.duration.abs>dva minuta</time.duration.abs>.
{S} Porodaj sa carskim rezom uradio je <persName.full>dr
Dragan Ivanović</persName.full> i, po rečima dežurnih
te večeri u <org>Ginekološko-akušerskoj klinici</org> u
Višegradske ulici, sve je proteklo rutinski {S} dečak
dugačak <measure.exact>47 santimetara</measure.exact>
težio je <measure.exact>dva i po
kilograma</measure.exact>, a devojčica je bila
<measure.exact>tri stotine grama</measure.exact> teža i
<measure.exact>dva santimetra</measure.exact> duža od
svog brata. {S} Trudnoća <persName.full>Milice
Ivković</persName.full> je, inače, rezultat vantelesne
oplodnje. {S} Milica ima <time.duration.abs>30
godina</time.duration.abs> i rodom je iz
<top.gr>Kruševca</top.gr>, a srećni tata se zove Ivan,
ističu u porodištu.

Belgrade
cette annee
deux minutes
47 cm
300 gr
Kruševac

Apply cascade graphs to text

HLT Group, language re...
File Edit View History
Named Entity Recognit...
hlt.rgf.bg.ac.rs/
Most Visited
1640
Untitled - Notepad
File Edit Format
Časovnik tek š...
<top.gr>Beogra...
mogućim novogo...
blizancima, pr...
godini</time.d...
Ivković</persN...
mama</role></p></div>

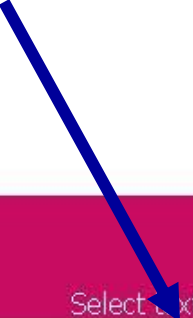




NERosetta

http://arhimed.matf.bg.ac.rs/~andjelka/paralel_extended/index.php

[SEARCH](#) [PARALEL SEARCH](#) [LIST OF ANNOTATION SCHEMAS](#) [LOGIN](#)



NERosetta
APPLICATION FOR NAMED ENTITY COMPARISON

Select text
Around the World in Eighty Days

Select language pair
en-sr
fr-sr
fr-en

Select annotation schemas

SHOW CATEGORIES



[SEARCH](#) [PARALEL SEARCH](#) [LIST OF ANNOTATION SCHEMAS](#) [LOGIN](#)

NERosetta

APPLICATION FOR NAMED ENTITY COMPARISON

Select text

Around the World in Eighty Days

Select language

en

fr

sr

Select annotation scheme

SHOW CATEGORIES



NERosetta

NERosetta
APPLICATION FOR NAMED ENTITY COMPARISON

Select schema:

sr - Krstev, Vitas

All the used entities are according to the Quaero guide recommendation.

Official web site: <http://quaero.org>

[Quaero annotation guide \(in French\)](#)

amount.measure → **measure.exact** 37 kilometara
amount.measure → **measure.range** od 500 metara do jednog kilometra

amount.money → **money.exact** 70.000 miliona evra
amount.money → **money.range** (izmedxu) 5.5 I svest miliona dolara

amount.obj → **amount.exact** jedanaest socijaldemokrata
amount.obj → **amount.range** sa dve na cyetiri godine

amount.percent → **percent.exact** 41,6 posto
amount.percent → **percent.range** 3.5 do 4 odsto

func.coll → **role.coll** bankari Dyon Sullivan i Samuel Falentin

func.ind → **role** ministar za kulturu i medije Srbije

loc.adm.town → **top.gr** Novi Sad

loc.adm.nat → **top.dr** Azerbejdyan

loc.phys.geo → **top.geo** Triglav

loc.phys.hydro → **top.hyd** Bileckko jezero



NERosetta

Named Entity Recognition - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Named Entity Recognition

www.korpus.matf.bg.ac.rs/nerosetta/



Google

Most Visited Getting Started

- subsubcategory time.hour.period is not supported
- subsubcategory time.hour.rel is not supported

- time.hour.period is time.hour.period
- time.hour.rel is time.hour.rel

tout le monde, et non sans raison, se mettait contre lui; on ne le prenait plus qu'à cent cinquante, à deux cents contre un, quand, **sept jours** après son départ, un incident, complètement inattendu, fit qu'on ne le prit plus du tout.

SEGID: N427

svi su bili protiv njega i ne bez razloga; već su ga uzimali i za sto pedeset i dve stotine prema jedan, kad ga usled jednog potpuno neočekivanog događaja, na **sedam dana** posle njegovog odlaska, prestadoše uzimati.

time.duration.abs

SEGID: N427

En effet, **pendant cette journée**, à neuf heures du **soir**, le directeur de la police métropolitaine avait reçu une dépêche télégraphique ainsi conçue:

SEGID: N428

U stvari, **toga dana u devet časova uveče** upravnik prestoničke policije primio je jedan brzojav sledeće sadržine:

SEGID: N428

Le mercredi 9 octobre, on attendait pour onze heures du **matin**, à Suez, le paquebot Mongolia, de la Compagnie péninsulaire et orientale, steamer en fer à hélice et à spardeck, jaugeant deux mille huit cents tonnes et possédant une force nominale de cinq cents chevaux.

SEGID: N440

U **sredu 9. oktobra, oko jedanaest časova** pre podne, očekivali su u Suecu Mongoliju, brod Istočnog i poluostrvskog društva, gvozdeni brod na propeler sa spardekom sa nosivošću od dve hiljade osam **sto tona** i sa snagom od pet stotina konja.

SEGID: N440

C'était un des plus rapides marcheurs de la Compagnie, et les vitesses réglementaires, soit **dix milles** à l'heure entre Brindisi et Suez, et **neuf milles cinquante-trois** centièmes entre Suez et Bombay, il les avait toujours dépassées.

To beše jedan od najbržih brodova društva, koji je uvek premašivao propisnu brzinu, tj. **deset milja** između Brindizija i Sueca, i **devet i po milja** između Sueca i Bombaja.

SEGID: N440



Named Entity Rec...



Inflection

20:32

8.10.201



NERosetta

Named Entity Recognition - Mozilla Firefox

Edit View History Bookmarks Tools Help

Named Entity Recognition



www.korpus.matf.bg.ac.rs/nerosetta/



8 Google

Most Visited Getting Started

- subsubcategory time.hour.period is not supported
- subsubcategory time.hour.rel is not supported

- time.hour.period is time.hour.period
- time.hour.rel is time.hour.rel

tout le monde, et non sans raison, se mettait contre lui; on ne le prenait plus qu'à cent cinquante, à deux cents contre un, quand, **sept jours** après son départ, un incident, complètement inattendu, fit qu'on ne le prit plus du tout.

SEGID: N427

svi su bili protiv njega i ne bez razloga; već su ga uzimali i za sto pedeset i dve stotine prema jedan, kad ga usled jednog potpuno neočekivanog događaja, na **sedam dana** posle njegovog odlaska, prestadoše uzimati.

SEGID: N427

En effet, **pendant cette journée**, à neuf heures du **soir**, le directeur de la police métropolitaine avait reçu une dépêche télégraphique ainsi conçue:

SEGID: N428

U stvari, **toga dana u devet časova uveče** upravnik prestoničke policije primio je jedan brzojav sledeće sadržine:

SEGID: N428

Le mercredi 9 octobre, on attendait pour onze heures du **matin**, à Suez, le paquebot Mongolia, de la Compagnie péninsulaire et orientale, steamer en fer à hélice et à spardeck, jaugeant deux mille huit cents tonnes et possédant une force nominale de cinq cents chevaux.

SEGID: N440

U **sredu 9. oktobra, oko jedanaest časova** pre podne, očekivali su u Suecu Mongoliju, brod Istočnog i poluostrvskog društva, gvozdeni brod na propeler sa spardekom sa nosivošću od dve hiljade osam **sto tona** i sa snagom od pet stotina konja.

SEGID: N440

C'était un des plus rapides marcheurs de la Compagnie, et les vitesses réglementaires, soit **dix milles** à l'heure entre Brindisi et Suez, et **neuf milles cinquante-trois** centièmes entre Suez et Bombay, il les avait toujours dépassées.

To beše jedan od najbržih brodova društva, koji je uvek premašivao propisnu brzinu, tj. **deset milja** između Brindizija i Sueca, i **devet i po milja** između Sueca i Bombaja.

SEGID: N440



Named Entity Rec...



Inflection



Microsoft Office...

SR



NERosetta

el is time.hour.rel

tiv njega i ne bez razloga; već su ga uzimali i za
stotine prema jedan, kad ga usled jednog potp
događaja, na **sedam dana** posle njegovog odla
mati.

time.duration.abs

SEGID:

i **dana u devet časova uveče** upravnik preston

tout le monde, et non sans raison, se mettait contre lui; on ne le
prenait plus qu'à cent cinquante, à deux cents contre un, quand, **sept**
jours après son départ, un incident, complètement inattendu, fit
qu'on ne le prit plus du tout.

amount.phy.dur

SEGID: N427

En effet. **pendant cette journée.** à neuf heures du **soir.** le

svi su t
pereset
nečeki
presti



NERosetta

CHECK SEARCH CRITERIA:

- ☐ INCLUDE SUPCATEGORIES
- ☒ EQUAL CATEGORIES
- ☐ INCLUDE SUBCATEGORIES

SEARCH

amount

event

function

loc

org

person

prod

time

universal

In annotation schema (gr, grNER) :

- category person is Person

In annotation schema (en, ner3) :

- category person is PERSON

- Τι ώρα έφτασε το τελευταίο τρένο από το Λίβερπουλ; ρώτησε ο **Τόμας Φλάναγκαν**.

SEGID: N4320

'What time did the last train arrive from Liverpool?' asked **Thomas Flanagan**.

SEGID: N4320

- Στις επτά και είκοσι τρία, απάντησε ο **Γκοτιέ Ραλφ**, και το επόμενο φτάνει στις δώδεκα και δέκα τα μεσάνυχτα.

SEGID: N4321

'Seven twenty-three,' replied **Gauthier Ralph**. 'And the next one only arrives at ten past twelve.'

SEGID: N4321

- Τότε, κύριοι, συνέχισε ο **Αντριου Σπιούαρτ**, αν ο **Φιλέας Φογκ**

'Well, gentlemen,' continued **Andrew Stuart**. 'If **Phileas Fogg** had

VebRanka



Expansion de requête - VebRanka

<http://hlt.rgf.bg.ac.rs/VebRanka/About.aspx?param=1>



Language resources exploration demo site

[Home](#) > [Query expansion](#) > **WordNet basic search**

Search expression

sr ▼

☒ Limit to selected language

Site or domain

Search one site (like wikipedia.org) or limit your results to a domain like .rs, .edu, .org or .gov

☒ Morphological expansion ☒ Synonym ☐ Cyrillic ☒ Latin

Include related synsets - ▼ ☒ To (ILR) ☐ From (RILR) Depth

[Expand the query by WordNet](#) [Expand the query morphologically](#)

Result of web search with the original and expanded query [Result of web search with the expanded query](#) [LinkButton](#)



VebRanka

 **Language resources exploration demo site**

[Home](#) > [Query expansion](#) > **WordNet basic search**

Search expression ☒ Limit to selected language

Site or domain *Search one site (like wikipedia.org) or limit your domain like .rs, .edu, .org or .gov*

☒ Morphological expansion ☒ Synonym ☒ Cyrillic ☒ Latin

Include related synsets ☒ To (ILR) ☐ From (RILR) Depth

[Expand the query by WordNet](#) [Expand the query morphologically](#)

"očevima"OR"očeve"OR"očeva"OR"očevi"OR"ocima"OR"otaca"OR"oci"OR"ocem"OR"oče"OR"oce"OR"ocu"OR"oca"OR"otac"OR"очевима"OR"очеве"OR"очева"OR"очеви"OR"оцима"OR"отаца"OR"оци"OR"оцем"OR"оче"OR"оце"OR"оцу"OR"оца"OR"отац"

[Result of web search with the original and expanded query](#) [Result of web search with the expanded query](#) [LinkButton](#)



VebRanka

otac powered by Google™

About 1,060,000 results (0.32 seconds)



[otac - Kurir](#)

Пре 1 дан ... facebook: 5 Crna Hronika 28.09.2013. u 07:53. UBICA RODITELJA: Monstrum hteo da se obesi. Od čaršava je napravio omču, koju je stavio ...
www.kurir-info.rs

[PTC :: Отац ранио сина у Сремчици](#)

Пре 5 дана ... Ухалшен С. Ј. из Сремчице због сумње да је свог сина убо ножем у стомак. Повређени оперисан у Ургентном центру. Осумњиченом ...
www.rts.rs



[Stanijin otac znao Slobodana Miloševića i Arkana! - Kurir](#)

2 jyn 2013 ... Tata Dragan je bio kućni prijatelj Slobodana Miloševića, poznao je i Željka Ražnatovića Arkana. Kad je život u Hrvatskoj postao nemoguć, ...
www.kurir-info.rs

[PTC :: Може ујна, не и отац](#)

Пре 5 дана ... Поред болесног детета у болници може да борави један родитељ, каже Закон о правима пацијената. Бака, тетка, ујна често једина ...
www.rts.rs



[Ilić: Otac Obradovića iz „Dveri” primao mito | Društvo | Novosti.rs](#)

Пре 1 дан ... Ministar Velimir Ilić izneo teške optužbe na račun čačanskog građevinskog inspektora, čiji je sin Boško Obradović funkcioner „Dveri”.
www.novosti.rs

"očevima"OR"očeve"OR"očeva"OR"očevi"OR"ocima"OR"otaca"OR"oci"OR"ocem"OR"oče"OR"oce"OR"ocu"t powered by Google™

About 1,060,000 results (0.35 seconds)

[Oci.rs | Portal posvećen očima.](#)

Oci.rs je oftalmološki portal posvećen očima i metodama u lečenju bolesti oka kao što su katarakta, keratokonus, glaukom i konjuktivitis.
oci.rs



[OTAC NA SLUŽBENOM PUTU – A.Sidran « Позориште "Атеље 212"»](#)

OTAC NA SLUŽBENOM PUTU – A.Sidran. Abdulah Sidran OTAC NA SLUŽBENOM PUTU. Reditelj i autor songova: OLIVER FRLJIĆ. [Show as slideshow].
www.atelje212.rs



[otac - Kurir](#)

Пре 1 дан ... DIVLJA GRADNJA: Otac stranačkog lidera uzimao mito preko sina ... Krivična prijava protiv oca uglednog funkcionera gurnuta u fioku u MUP ... Kada nam očevi i majke ostare a mi odrastemo, dešava se da nemamo dovoljno strpljenja ... RAZNEO SE BOMBOM: Ubio se jer oca spomenik nije imao da plati.
www.kurir-info.rs

[PTC :: Вербални напад на оца Дионисија](#)

пре 7 сата/и ... Напад на оца Дионисија још један је у низу притисака који имају за циљ стварање атмосфере сталног страха и несигурности за српски ...
www.rts.rs



[Stanijin otac znao Slobodana Miloševića i Arkana! - Kurir](#)

2 jyn 2013 ... Otac Stanije Dobrojević završio je gimnaziju u Podravskoj Slatini 1974 ... i u potresnoj ispovesti starleta govori o oca koji je ubijen tokom rata u ...
www.kurir-info.rs



Bibliša

<http://hlt.rgf.bg.ac.rs/Biblisha/>

BIBLIŠA: ALIGNED COLLECTION SEARCH TOOL [Log In]

Home Metadata search English metadata browse Search About

WELCOME TO INFOTHECA ALIGNED COLLE

Keyword

Synonyms	sr	en
☒ WordNet	lisni katalog	card catalog, card catalog
☒ Dictionary of Librarianship	konvencionalni katalog, lisni katalog	card catalogue, card file catalogue, manually-op

Serbian query English query

**user's query –
lisni katalog**

**bilingual expansion –
Wordnet**

**bilingual expansion –
LIS terminology DB**

**morphological
expansion
- Serbian e-dict**



Bibliša – INFOTheca papers

Number of concordances: 9	Number of concordances (en): 9	Broj konkordansi (sr): 9
Brzulović-Stanisavljević et al., 2007, vol. VIII:1/2, ID: 2007.1/2.4 metadata	n15 : This is principally accomplished by OCR (Optical character recognition) involving scanning of card catalogue records (most commonly coloured) and their conversion into machine readable form (UNIMARC, MARC21...) followed by immediate quality control.	n15 : Uglavnom se vrši pomoću softvera OCR (Optical character recognition) što podrazumeva skeniranje kataloških listića (najčešće u boji) i njihovo pretvaranje u mašinski čitljiv format (UNIMARC, MARC21...) nakon čega se pristupa normativnoj kontroli.
Brzulović-Stanisavljević et al., 2007, vol. VIII:1/2, ID: 2007.1/2.4 metadata	n68 : In Croatia, card catalogues of two libraries have been digitized - historical catalogue Bibliotheca Civica in Rijeka University Library (http://www.svkri.hr), and alphabetical catalogue of National and University Library (NSK - http://www.nsk.hr) in Zagreb.	n68 : Dve hrvatske biblioteke su digitalizovale svoje lisne kataloge - Sveučilišna knjižnica u Rijeci (http://www.svkri.hr) svoj istorijski katalog Biblioteka civica, a Nacionalna i sveučilišna knjižnica (NSK - http://www.nsk.hr) u Zagrebu svoj abecedni katalog.
Brzulović-Stanisavljević et al., 2007, vol. VIII:1/2, ID: 2007.1/2.4 metadata	n91 : The third phase foresaw the image catalogue would be imported into a COBISS database (50.000 records).	n91 : U trećoj fazi predviđeno je da slikovni katalog lisnog kataloga bude prenet u bazu COBISS (50.000 zapisa).
Brzulović-Stanisavljević et al., 2007, vol. VIII:1/2, ID: 2007.1/2.4 metadata	n92 : The fourth phase would deal with importing foreign book card catalogues to COBISS database (300.000 titles), and the fifth, importing of remaining card catalogue records to COBISS database (250.000 titles).	n92 : Četvrta faza se odnosi na prenos lisnog kataloga stranih knjiga u bazu COBISS (300.000 naslova), a peta na prenos ostalih lisnih kataloga u bazu COBISS (250.000 naslova).
Brzulović-Stanisavljević et al., 2007, vol. VIII:1/2, ID: 2007.1/2.4 metadata	n113 : That is why designation 1 was entered in the field 001\$g of COMARC/B format - sublevel 1 (data downloaded from card catalogue).	n113 : Zato je u formatu COMARC/B u polju 001\$g unošena oznaka 1 - podnivo 1 (podaci preuzeti sa kataloških listića).
Brzulović-Stanisavljević et al., 2007, vol. VIII:1/2, ID: 2007.1/2.4 metadata	n139 : The reason for that may be found in the fact that younger users, especially students, browse electronic rather than card catalogue .	n139 : Razlog tome je i činjenica da mlađi korisnici, naročito studenti, češće pretražuju elektronski nego lisni katalog .
Dakić, 2007, vol. VIII:1/2, ID: 2007.1/2.1 metadata	n64 : They passed state professional examination 15 or 20 years ago, and in their libraries obtained only card catalogues , or even, they worked only with users.	n64 : Stručni ispit su završili pre 15 ili 20 godina, a u svojim bibliotekama održavaju samo lisne kataloge čak ponegde rade samo sa korisnicima.
Dakić, 2007, vol. VIII:1/2, ID: 2007.1/2.1 metadata	n69 : Typical mistakes, which appeared as a result of differences between card catalogue and electronic catalogue, as well as insufficient knowledge of cataloguing rules have been evidenced.	n69 : Ustanovljene su tipske greške koje se javljaju kao rezultat različitosti lisnog i elektronskog kataloga kao i nedovoljno poznavanje kataloških pravila.
	n42 : On the other hand, the information tools are very	n42 : Sa druge strane, informacioni instrumenti su u

morphological
expansion of
MWUs



LeXimir

user's keyword
ljubav

semantic expansion
- Wordnet

morphological expansion
- Serbian e-dict

bilingual expansion
- Wordnet

LeXimir interface showing search results for the keyword "ljubav".

Search: ljubav ☒ is lema

Use "+" for "OR" (example job+luck)

☒ With inflection Unitex
☒ Semantical extension WordNet
☒ Aurora ☒ Latin ☐ Cyrillic

WordNet I WordNet II Prolex I Prolex II

Table Text XML View Literals

Include SS in relation: ☐ From (ILR) ☐ To (RILR) Relation Depth: 1

Literal	Inflected forms
beloved	
dear	
dearest	

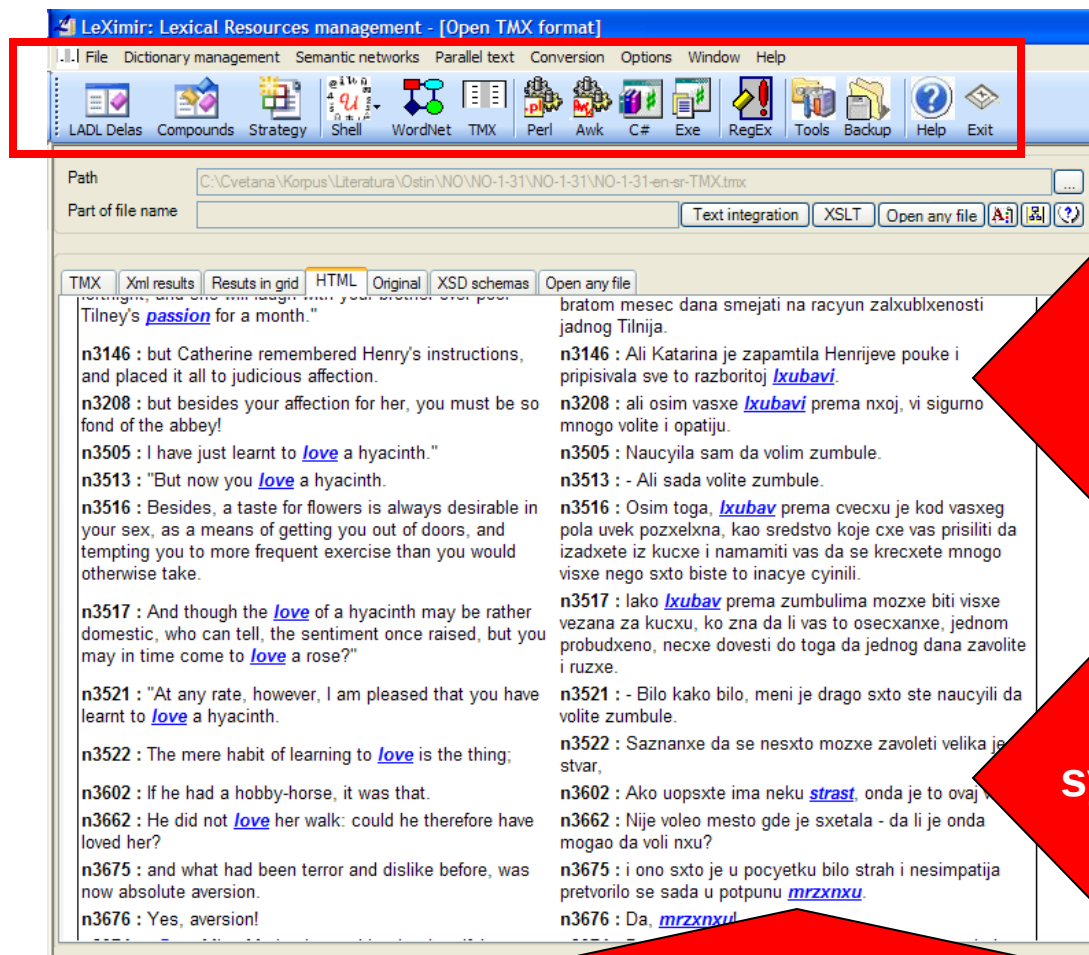
lvolxenilvolxena|dragan|draganaldraganuldraganeldraganom|draganil|draganimaldraganol|draganama|ljubavimal|ljubavixul|ljubavil|ljubav|cakan|cakanama|cakanom|cakanol|cakanul|cakanil|cakanel|cakanal|strastimal|strasxcxul|strastil|strast|mrzrxal|mrzrxel|mrzrxil|mrzrxul|mrzrxol|mrzrxom|mrzrxama|hepriateljstvolvolxenilvolxenal|ljubavimal|ljubavil|ljubavil|ljubav|strašćul|mržnja|mržnje

beloved|dear|dearest|loved one|honey|love|passion|hate|hatred

contains(translate(., 'ABCDEFGHIJKLMNOPQRSTUVWXYZÆŽŠĐ', 'abcdefghijklmnopqrstuvwxyzæžšđ'), ") or contains(translate(., 'ABCDEFGHIJKLMNOPQRSTUVWXYZÆŽŠĐ', 'abcdefghijklmnopqrstuvwxyzæžšđ'), 'volxeni') or contains(translate(., 'ABCDEFGHIJKLMNOPQRSTUVWXYZÆŽŠĐ', 'abcdefghijklmnopqrstuvwxyzæžšđ'), 'beloved')



LeXimir



basic - ljubav

synonym - strast

antonym -
mržnja



Merci!