



Uređivanje mrežnih e-rječnika u sintaktičko-semantičkom okviru

Suzana Molčanov, Marko Landeka, Blaženka Budimčić
suzana.molcanov@gmail.com; marko.landeka@gmail.com;
blbudimcic@gmail.com

CLGF-2020

Osijek, 25.-26. rujna 2020.

Sadržaj:

- 1 Uvod u modul REDAKTOR
 - Prijava i osnovne funkcije
 - Riječ i njena obilježja
- 2 Dodavanja i promjene
 - Naglasci i raščlambe
 - WOS/SOW obilježja
- 3 Napredne funkcije
 - Morfološki generator
 - Integracija hrvatskih govora

Mrežni okvir SSF na <http://ssf.mathos.hr/>

SINTAKSNO-SEMANTIČKI OKVIR

smolcanov ***** Prijava
Hrvatski Engleski

The screenshot displays the SSF web interface. On the left, a 'Dokumenti' (Documents) sidebar shows a tree structure with folders like 'Biblija' and 'NZ', and files such as '01_Marko.txt', '02_Matej.txt', '03_Luka.txt', '04_Ivan.txt', and '05_Djela_apostolska.txt'. On the right, the main area is titled 'Rječi' (Words) and includes tabs for 'T-Strukture', 'Rečenice', 'MSY', 'LEX', and 'MWE'. Below these tabs are three input fields labeled 'Prefix', 'Riječ', and 'Sufix'. The 'Riječ' field is highlighted in red and contains the dropdown menu for 'SOW'. Below the input fields is a search example field labeled 'Uzorak pretrage:'.

Sintaktičko-semantički mrežni okvir (eng. *syntactic & semantic framework*: **SSF**)

- čuva tri vrste standardnih rječnika (*subčestični, leksikon, višerječnice*),
- čuva dijalektološke rječnike (za sada *štokavski i kajkavski*),
- omogućuje dodavanje drugih rječnika
- stvara rječnike iz digitalnih korpusa (*Biblija, Bulaja edicija* i sl.),
- integrira sve rječnike i sve korpusne,
- omogućuje brojne sintaktičko-semantičke funkcije i *editiranja* (modul **REDAKTOR**).

Modul REDAKTOR u sintaktičko-semantičkom okviru

Rad u uređivačkom modulu:

- zahtijeva prijavu upisom korisničkog imena i zaporke.
- korisničko ime i zaporka može se dobiti od autora programa (*M. Orešković*) ili administratora (*D. Ševerdija* ili *M. Essert*),
- postoji više razina korisnika (ne mogu svi sve raditi),
- redaktori osim uređivačkih imaju i posebne mogućnosti za rad u SSF-u.

SINTAKSNO-SEMANTIČKI OKVIR

Korisničko ime: jhorvat

Lozinka:

Prijava

Hrvatski Engleski

Riječi T-Strukture O-Strukture Rečenice Domene FPJ MSY LEX MWE

Osnovna funkcija REDAKTOR-a: PRIHVATI ili ODBACI

SINAKSNO-SEMANTIČKI OKVIR

Prijavljen kao: Jasna Horvat OdjavaPrikaži rezultata po straniciMorfološki generator Multipliciranje Samo NLO Samo neproverjene riječi

Lemma	Riječ	Akcent	Slogovi	Morfovi	WOS	SOW	Akcije
<u>Unesi</u>	dama	dáma	dáma	dama	Gramatika	Primer, Definicija, Lokaliziraj, Izvori, Autor, Rječnik • Kakavski	
dama	dama	dáma	dá-ma	dama	Slogovi, Vrsta riječi • Imenica, Rod • Ženski, Broj • množina, Padaj • Genitiv	OVI, ovi, sameći, mreža • BabelNet • Definicija, mreža • BabelNet • Kategorija, Rječnik • Standardni, Stav • Pozitivan, mreža • CroViti • Definicija, mreža • CroViti • Sinonim, mreža • CroViti • Hipernim, mreža • HJP • Definicija, mreža • HJP • Etimologija, mreža • ENC • Definicija	Potvrdi, Odbaci
dama	dama	dáma	dá-ma	dama	Slogovi, Vrsta riječi • Imenica, Rod • Ženski, Broj • Jednina, Padaj • Nominativ	OVI, ovi, sameći, mreža • BabelNet • Definicija, mreža • BabelNet • Kategorija, Rječnik • Standardni	Potvrdi, Odbaci

- Riječi dohvaćene (eng. *parsing*) iz korpusa (digitaliziranog dokumenta) u SSF LEX-rječnik može biti netočna, (*tipfeler*) pa je treba izbaciti iz rječnika,
- prihvaćena riječ treba imati i svoja ispravna (gramatička WOS i semantička SOW obilježja).

Tablica redaktorskih funkcija

SINAKSNO-SEMANTIČKI OKVIR

Prijavljen kao: Jasna Horvat Odjava

Prikaži: 100 rezultata po stranici

Morfološki generator Multiplificiranje Samo NLO Samo neprovjerene riječi

Lemma	Riječ	Akcent	Slogovi	Morfovi	WOS Filter	SOW Filter	Akcije
<u>Unesi</u>	U	<u>Unesi</u>	U	U	Vrsta riječi • Interpunkcija • Navodnik • " ✕	Rječnik • Standardni ✕ +	
<u>Unesi</u>	U	<u>Unesi</u>	U	U	Vrsta riječi • Interpunkcija • Navodnik ✕	Rječnik • Standardni ✕ +	
<u>Unesi</u>	U	<u>Unesi</u>	U	U	Vrsta riječi • Interpunkcija • Navodnik ✕	Rječnik • Standardni ✕ +	
<u>Unesi</u>	noea	<u>Unesi</u>	no-e-za	noea	Slogovi ✕	Rječnik • Standardni ✕ +	
<u>Unesi</u>	kognitivno	kōgnitivno	ko-gni-ti-vno	kogniti-vn-o	Slogovi [7] ✕ Vrsta riječi • Pridjev ✕	Rječnik • Standardni ✕ +	Potvrdi Odbaci
<u>Unesi</u>	people	<u>Unesi</u>	pe-o-ple	people	Slogovi ✕	Rječnik • Standardni ✕ +	
<u>Unesi</u>	dikotiledona	<u>Unesi</u>	di-to-li-le-do-na	dikotiledona	Slogovi [31] ✕	Rječnik • Standardni ✕ +	

koliko riječi prikazivati

odbačene riječi (crvenasta podloga)

Abecedni redoslijed

riječ koja se može potvrditi ili odbaciti

- duljina tabličnog prikaza i kategorije riječi (*nepoznate, odbačene, prihvaćene*)
- odbačene riječi dobivaju *crvenu* podlogu, a mogu se ponovo vratiti na validaciju
- prihvaćena riječ ima *zelenu* podlogu, a neobrađena *bijelu*.

Dohvaćanje riječi iz rječnika i njena obilježja

Riječ	Akcent	Slogovi	Morfovi	WOS	SOW	Akcije
dama				Filter	Filter	
dama	dāma	dā-ma	dam-a	Gramatika ✕	Primjer ✕ Definicija ✕ Lokalitet ✕ Izvori ✕ Autor ✕ Rječnik • Kajkavski ✕	
dama	dāma	dā-ma	dama	Slogovi ✕ Vrata riječi • Imenica ✕ Rod • Ženski ✕	OWL • owlSameAs [2] ✕ Mreža • BabelNet • Definicija ✕ Mreža • BabelNet • Kategorija [2] ✕	Potvrdi Odbaci

- dohvaćanje upisom riječi u stupac 'Riječ' ili u stupac 'Lema',
- moguće je koristiti znak '%' za bilo koliko slova (en. *wildcard*),
- dobivene riječi razlikuje se i u slučaju da im je različito samo jedno obilježje (npr. naglasak ili primjer/definicija),
- kojem rječniku riječ pripada vidljivo je iz semantičkih (SOW) obilježja
- WOS i SOW obilježja mogu se dodavati i brisati,
- program omogućuje mijenjanje *naglasaka* (akcenta) i *raščlambe* (slogova i morfema).

Dohvaćanje (filter) riječi preko njenih obilježja

SINTAKSNO-SEMANTIČKI OKVIR

 Prijavljen kao: **Jasna Horvat** Odjava

 Prikaži rezultata po stranici

 Morfološki generator Multipliciranje Samo NLO Samo neproverjene riječi

Lemma	Riječ	Akcent	Slogovi	Morfovi	WOS	SOW	Akcije	
<u>Unesi</u>	prikaz	prikaz	pri-kaz	pri-kaz-0	Filter <input type="checkbox"/> Ije <input type="checkbox"/> Vrsta riječi <input type="checkbox"/> Međuvrste <input type="checkbox"/> Rod <input checked="" type="checkbox"/> Muški <input type="checkbox"/> Ženski <input type="checkbox"/> Srednji <input type="checkbox"/> Broj <input type="checkbox"/> Padež <input type="checkbox"/> Određenost <input type="checkbox"/> Stupnjevanj <input type="checkbox"/> Naglašenost <input type="checkbox"/> Osoba <input type="checkbox"/> Vid <input type="checkbox"/> Vrijeme	Imenica ✕ Imenina ✕ ... Rod • Muški ✕ • Nominativ ✕	Teorije • Šarić • Sinonimski skup [3] ✕ Jezgri • Imenica ✕ Rječnik • Standardni ✕ Okvir • Apsolutno ✕	
<u>Unesi</u>	stas	stás	stas	stas-0		Rod • Muški ✕ • Nominativ ✕	Jezgri • Imenica ✕ Rječnik • Standardni ✕ Živo • Osoba • Tjelesno ✕	
<u>Unesi</u>	lančast	lančast	lan-čast	lanča-st-0		Pridjev ✕	Rječnik • Standardni ✕	
<u>Unesi</u>	lančasti	lančasti	lan-ča-sti	lančasti		• Pridjev ✕	Rječnik • Standardni ✕	

- kad želimo dohvatiti ili prikazati samo neku vrstu riječi, npr. *imenice* ili *zamjenice*, k tomu i s još nekim drugim obilježjima (npr. *rod* ili *padež*),
- isto vrijedi za SOW obilježja (npr. kod uređivanja *toponima*, da nam se samo oni prikažu)
- moguće su sve kombinacije iz stablenih struktura.

Ista riječ u različitim dijalektima (primjer/definicija)

kajkavski:

SOW
Filter ▾

Primjer | ✕
Lokalitet | ✕
Autor | ✕

Dīla se dāma, a nē zna ni kāj
je to bīt dāma

Spremi

standardno:

SOW
Filter ▾

Primjer | ✕ Definicija | ✕
Lokalitet | ✕ Izvori | ✕
Autor | ✕ Rječnik • Kajkavs

ona koja se po mjerilima
građanskog društva i života
u gradu odlikuje
odnjegovanim dobrim
osobinama i ofmienim

Spremi

Dodavanje ili promjena naglasaka

Akcent	Slo
dāma	da

d ā m a

OK

a
à
ǎ
â
á
ǎ
ǎ

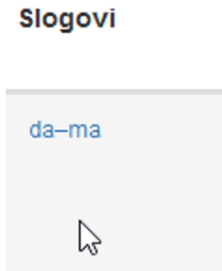
Akcent

dāmā

Raščlambe - slogovi



Klik na OK daje:




- korisno za riječi koje (eventualno) nisu dobro rastavljene novim algoritmom raščlambe (*Užarević, Essert, Jovanović*),
- dodavanje crtice na mjestima granice sloga (klik na prazninu),
- klik na crticu: crtica se briše.

Raščlambe - morfovi

Slogovi	Morfovi	WOS
da-ma	dama	Filter

d a m - a

Reset

OK 

Morfovi

dam-a

- isti postupak kao za slogove,
- razlika: moguće je dodati i nul-morf (na početku riječi).

Dodavanje WOS/SOW obilježja

The image shows two screenshots of a software interface for adding WOS/SOW tags. The left screenshot shows the WOS section with a search filter and a list of tags. The right screenshot shows the SOW section with a search filter and a list of tags.

WOS Section (Left Screenshot):

- Filter: Filter
- Gramatika | X
- Broj • Množina
- Broj • Množina • Pluralia tantum
- Stupnjevanje
- Stupnjevanje • Pozitiv (highlighted)
- Stupnjevanje • Komparativ
- Stupnjevanje • Superlativ
- Određenost
- Određenost • Određen
- Određenost • Neodređen

SOW Section (Right Screenshot):

- Filter: Filter
- Gramatika | X
- im
- Vrsta riječi • Imenica (highlighted)
- Vrijeme • Imperfekt
- Vrsta riječi • Broj vrsta • Zapis • rimski
- Međuvrste • Brojevnica imenica
- Vrsta riječi • Interpunkcija • Simbol

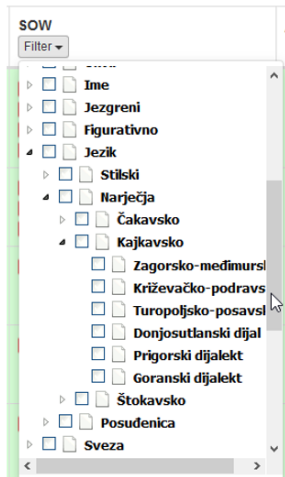
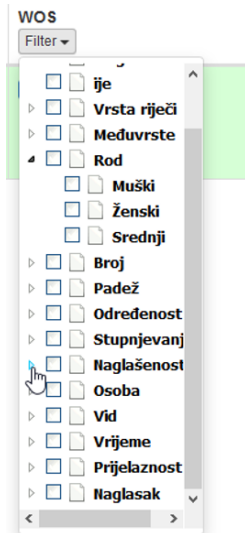
- upisom nekog niza znakova pojavljuju se obilježja koja taj niz imaju u sebi,
- klikom na željeno obilježje, ono se dodaju (pridružuje) zadanoj riječi.

Brisanje WOS/SOW obilježja


WOS	SOW
Filter ▼ Gramatika ✕ +	Filter ▼ Primjer ✕ Definicija ✕ Lokalitet ✕ Izvori ✕ Autor ✕ Rječnik • Kajkavski ✕ +

- klikom na iksić (✕) obilježja, briše se to obilježje pridruženo riječi
- klikom na plusić (+) otvara se prethodno opisana ponuda dodavanja novog obilježja
- nema ograničenja u dodavanju obilježja, koja su definirana u stablenoj WOS/SOW strukturi.

Stablena struktura WOS/SOW obilježja



Stvaranje pojavnica

	Riječ	Al
<input type="text"/>	<input type="text" value="dama"/>	<input type="text"/>
	dama 	c
	dama 	c

Morfološki generator

Imenice Pridjevi Glagoli Nepromjenjiva riječ

Unesite imenicu te odaberite rod i Nominativni nastavci:

Imenicu
Imenicu unesite u nominativu jednine

Naglasak

Rod

Nominativni nastavci

- Autor morfološkog generatora: Joško Markučić

Pojavnice

Singular

Select

N **dama**
dama

G **dame**
dame

D **dami**
dami

A **damu**
damu

V **damo** ili **dar**
damo ili dar

L **dami**
dami

I **damom**
damom

Plural

Select

N **dama**
dama

G **dame**
dame

D **dami**
dami

A **damu**
damu

V **damo** ili **dar**
damo ili dar

L **dami**
dami

I **damom**
damom

- pojavnice stvarane (generirane) iz leme riječi,
- moguće im je mijenjati naglasak,
- moguće ih je mijenjati ako glasovna promjena nije ispravno generirala riječ (2% slučajeva),
- ako pojavnica već postoji u rječniku pojavljuje se u zelenoj boji, ako ne - u crvenoj.

Povezivanje dijalektoloških riječi sa standardim jezikom

The screenshot shows the AFINGER interface. At the top, there is a navigation bar with tabs labeled AB, AF, AG, AJ, AK, AM, AN, AP, AR, and AS. The 'AF' tab is selected. Below the navigation bar, the title 'AFINGER' is displayed. Underneath, there are three input fields: 'Akcent:' with the value 'afinger', 'WOS:' with the value 'Gramatika', and 'SOW:' with a list of buttons: 'Primjer', 'Definicija', 'Lokalitet', 'Izvori', 'Autor', and 'Rječnik • Kajkavski'. A mouse cursor is hovering over the 'Definicija' button. Below the 'SOW' section, a search result is shown in a white box with a blue border: 'vješalica u ormaru'.

preko REDAKTORA dodati riječ koja će je povezati sa standardom:

The screenshot shows the REDAKTOR interface. On the left, there is a 'SOW' section with a 'Filter' dropdown menu. Below the filter, there is a list of buttons: 'Primjer | ✕', 'Definicija | ✕', 'Lokalitet | ✕', 'Izvori | ✕', 'Autor | ✕', and 'Rječnik • Kajkavski | ✕'. A dashed box with a plus sign is next to the last button. On the right, there is a text input field containing the text 'vješalica, vješalica u ormaru'. Below the input field, there is a 'Spremi' button.

Rezultat

AB AF AG AJ AK AM AN AP AR A

AFINGER

Akcent: afinger

Standard: vješalica

WOS: Gramatika

SOW: **Primjer** **Definicija** Lokalizet Izvori Autor Rječnik • Kajkavski

[vješalica](#), [vješalica u ormaru](#)

dijalektološka riječ poprima željena obilježja riječi standardnog jezika:

WOS

Vrsta riječi

- Imenica
- Zamjenica
- Pridjev

VJEŠALICA

Domena

Lemma: vješalica

Akcent: vješalica

Slogovi: vje-ša-li-ca

WOS: **Slogovi (7)** Vrsta riječi • Imenica Rod • Ženski Broj • Množina Padež • Genitiv

vje-ša-lic-a, vje-šal-i-ca, vje-šal-ic-a, vješ-a-li-ca, vieš-a-lic-a, vieš-al-i-ca, vieš-al-ic-a

WN • Definicija (2) Mreža • CroWN • Sinonim

Etimologija

Naprava od savijena metala na koju se nešto može objesiti | Naprava oblikovana poput ljudskih ramena na koju se vješa odjeća

Zaključak: integracija "ČA-KAJ-ŠTO" hrvatskoga jezika

C Č Ć D DŽ Đ E F G H I J

AB AF AG AJ

AFINGER

Akcent: afinger

Standard: vješalica

vješalica

WOS:

Slogovi [7] Vrsta riječi • Imenica

Rod • Ženski Broj • Množina

Padež • Genitiv

SOW:

OWL • owl:sameAs Rječnik • Standardni

Mreža • CroWN • Definicija

Mreža • CroWN • Sinonim

Mreža • CroWN • Hipernim

Mreža • HJP • Definicija

Mreža • HJP • Etimologija

Hvala na pozornosti



Nema više (kraj je blizu) 😊!