

Croatian Web Dictionary Mrežnik: One year later - What is different?

Financed by the Croatian Science Foundation



Lana Hudeček, Milica Mihaljević

Institute of Croatian Language and Linguistics

LANGUAGE TECHNOLOGIES & DIGITAL HUMANITIES 2018

Ljubljana, September 20-21

Croatian Web Dictionary Mrežnik

Croatian Web Dictionary – Mrežnik (research project financed by the Croatian Science Foundation)

Project leader: dr. sc. Lana Hudeček

Institution in which the project is conducted: Institut za hrvatski jezik i jezikoslovlje

duration: 4 years (1. March 2017. – 28. February 2021.)

Project number: IP-2016-06-2141

Corpora

- The dictionary is based on two Croatian corpora:
- *the Croatian Web Corpus* hrWaC (<http://nlp.ffzg.hr/resources/corpora/hrwac/>) current version of the corpus (v2.0) contains 1.9 billion tokens and is annotated with the lemma, morphosyntax, and dependency syntax layers
- *the Croatian Language Repository* (CLR; riznica.ihjj.hr). It has 101,782,863 tokens and 85,273,724 words.

Modules

- **Module for adult users**
- Module for children
- Module for foreigners

HRVATSKI
U IGRI



KUTAK
ZA UČITELJE

SAVJETI



FRAZEMI

Sketch Engine and Word Sketches

WORD SKETCH

Croatian Web (hrWaC 2.2, RFTagger)



zubar as noun 7,213x



prijedlog	imenica_iza_prijedloga	veznik	kakav?	komu-čemu	koordinacija	koga-čega	glagol_ispred_prijedloga
kod ... kod zubara	kod-s_X ...	al ...	privatnom ...	ići ...	ginekolog ... zubara i ginekologa	posjećivati ...	u-s ...
k ... k zubaru	od-s_X ...	ili ... zubara ili	kninski ... kninski zubar	otići ...	ortodont ...	posjetiti ... posjetiti zubara	na-s ...
	sa-s_X ...	iako ...	dežuran ... dežurnog zubara	odvesti ...	frizer ...	promijeniti ...	iz-s ...
	u-s_X ...	pa ... zubara pa	privatan ... privatnog zubara	boljeti ...	kirurg ...	bojati ... bojim zubara	sa-s ...
	za-s_X ...	ali ... zubara ali	socijalan ... socijalnog zubara	posjetiti ... posjeti zubaru	okulist ...	mijenjati ...	za-s ...
	do-s_X ...	jer ... zubara jer	socijalan ... socijalnog zubara	otisati ...	doktor ...	platiti ...	zbog-s ...
	o-s_X ...		prijašnji ...	odlaziti ...	liječnik ... liječnika i zubara	pitati ...	od-s ...
	k-s_X ...		ugledan ...	isti ...	specijalist ...	plaćati ...	po-s ...
	na-s_X ...		nijedan ...	otisti ...	veterinar ...	nazvati ...	bez-s ...
	zbog-s_X ...		tadašnji ...	bojati ...	odvjetnik ...	čekati ...	do-s ...
	prema-s_X ...		izvrstan ...	voditi ...	izvaditi ...	zvati ...	prije-s ...
	nakon-s_X ...		besplatan ...	planirati ...	zub ...	pronaci ...	o-s ...
	▼		bivši ... bivši zubar	▼	▼	▼	▼

Categories added to Word Sketches					Regular expressions used to search the corpus	
Verb (V)	Preposition (Pr)	Pronoun (P)	Adjective (Adj)	Adverb (Adv)	Interjection (I)	Conjunction (C)
V + <i>se</i> , <i>se</i> + V	Pr + N, Pr + Adv, V + Pr, N + Pr, Adj + Pr	P + Par, Par + P	Par + Adj	Par + Adv, Adv + Par	Adv + I, V + I, I + Pr, I + N, N + I, I + <i>i</i> , I + I, I + P, Par + I	Adv + C, Par + C, C + Adv, C + C

(N = noun, Par = particle)

Table 1: Addition to Word Sketches by categories

File Edit View Lemma Dictionary Format Tools Window Help



+ New entry (Ins)
Delete entry

lingvistika

limunada

limunika

lingvist

lingvistica

lingvistički

lingvistika

lingvistov

linija

lipa

lipanj

lira

- + 1 Add: Gramaticki_blok
- + 2 Add: Parnjaci
- + 3 Add: Stra_savjet
- + 4 Add: Sense
- + 5 Add: Subentry
- + 6 Add: Tvorbena_raz
- + 7 Add: Frazemi
- + 8 Add: Poveznice
- + 9 Add: vanjska_poveznica
- + Add: Normativna_nap
- + Add: Uporaba
- + Add: Lemma_Djeca
- + Add: Lemma_Stranci
- + Add: Gramaticki_blok+ Oblici

d=2018-09-11 09:50:49, Created=2018-08-07 13:26:26, ModifiedBy=josip, CreatedBy=

učavanje jezičnih pojava u najširem smislu.

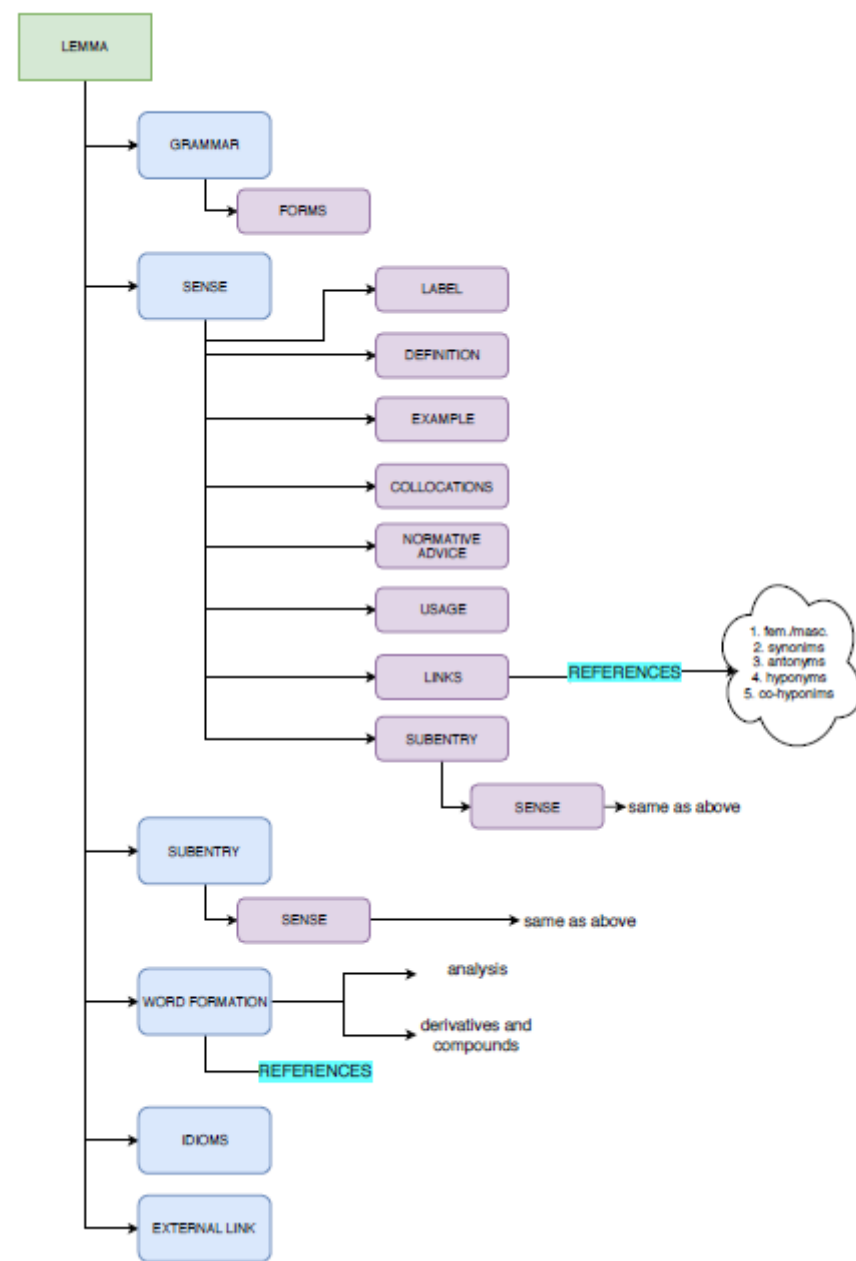
romanske jezike i književnosti Filozofskoga fakulteta iz predmeta koji se tiču poredbeni
iku reprezentativan korpus u skladu s najsuvremenijim dostignućima i spoznajama iz k


ebarska, arealna, empirijska, forenzička, generativna, genetska, historijska, indoeurope
gija i lingvistika, arheologija i lingvistika, etnologija i lingvistika, filozofija i lingvistika, fon
plokacija=magistar, predmet, profesor, student, studij, studija

uštvo za primijenjenu lingvistiku, Odsjek za opću lingvistiku i orijentalne studije, Suvrem

vistika, paleolingvistika, sociolingvistika

Filter (F5) Cornus (F6)



Gramaticki_blok:	
Oblici:	
naglaseni	GA zubára, DL zubáru, V zŭbáru/zŭbáre, I zubárom/zubárem; %imn.%i NV zubári, G zubárâ, DLI zubárima, A zubáre
nenaglaseni	
vanjska_poveznica	
djeca	
Sense:	
stilska_odrednica	razg. 
strucna_odrednica	- 
ograda	
Definicija:	
definicija	Zubar je liječnik koji se bavi stomatologijom.
definicija_stranic	
*definicija	
primjeri:	

Osnovni:

zùbār **zubar** im. m. (GA zubára, DL zubáru, V zùbāru/zùbāre, I zubárom/zubárezem; mn. NV zubári, G zubárâ, DLI zubárima, A zubaru) razg. Zubar je liječnik koji se bavi stomatologijom.

Tko želi zdrave zube i čvrste desni, mora zube prati 2 do 3 puta dnevno, svilenim koncem pročititi prostor između zubara na kontrolu. U nekoliko trenutaka možete postići osmijeh iz snova, a da pritom ne idete zubaru na oštećuju zubnu caklinu.

Kakav je zubar? besplatan, dežurni, izvrstan, poznati, privatni, socijalni, ugledan, uspješan, vrhunski; Što se zub, liječiti desni/zub, otvoriti zub, praviti/raditi (navlaku, plombu, protezu, zub), vaditi (zub, živac); Što se bojači ga se, ići k njemu, dolaziti k njemu, imati ga, javljati mu se, nazivati ga se, plaćati mu, posjećivati ga se i zubari, zubar i ortodont, zubari i doktori, zubari i frizeri, zubari i ginekolozi, zubari i liječnici, zubari i okultisti, zubar na muškarce: zubar ili zubarica; Povezuje se s: preporuka, strah, usluga

Riječ **zubar** upotrebljava se u razgovornome, ležernijemu stilu. U standardnome je jeziku bolje upotrijebiti riječ **stomatolog**. ŽENSKO: **stomatologica :1, stomatologinja :1, zubarica :1, SINONIM: stomatolog**

tvorenice **zubarev, zubarica, zubarov, zubarski** [Kristian] [Domagoj]

Stranci:

[Domagoj] [Kristian]

Djeca:

[Kristian] [Domagoj] **zubar** imenica zu-bar (zubari)

Zubar je muškarac koji je završio stomatološki fakultet i liječi zube.

dežurni zubar, privatni zubari

Zubar je muškarac koji je završio stomatološki fakultet i liječi zube. Zubarica je žena koja je završila stomatološki fakultet i liječi zube. Zubarice su žene koje su završile stomatološki fakultet i liječe zube. Zubari su muškarci koji su završili stomatološki fakultet i liječe zube, ali i muškarci i žene koji su završili stomatološki fakultet i liječe zube. Zubar, Lucija je zubarica, Marko i Luka su zubari, Lucija i Ana su zubarice, a Lucija, Ana, Marko i Luka su zubari.

ŽENSKO: **zubarica :1**

Sve:

zùbār **zubar** im. m. (GA zubára, DL zubáru, V zùbāru/zùbāre, I zubárom/zubárezem; mn. NV zubári, G zubárâ, DLI zubárima, A zubaru) razg. Zubar je liječnik koji se bavi stomatologijom.

Tko želi zdrave zube i čvrste desni, mora zube prati 2 do 3 puta dnevno, svilenim koncem pročititi prostor između zubara na kontrolu. U nekoliko trenutaka možete postići osmijeh iz snova, a da pritom ne idete zubaru na oštećuju zubnu caklinu.

Kakav je zubar? besplatan, dežurni, izvrstan, poznati, privatni, socijalni, ugledan, uspješan, vrhunski; Što se zub, liječiti desni/zub, otvoriti zub, praviti/raditi (navlaku, plombu, protezu, zub), vaditi (zub, živac); Što se bojači ga se, ići k njemu, dolaziti k njemu, imati ga, javljati mu se, nazivati ga se, plaćati mu, posjećivati ga se i zubari, zubar i ortodont, zubari i doktori, zubari i frizeri, zubari i ginekolozi, zubari i liječnici, zubari i okultisti, zubar na muškarce: zubar ili zubarica; Povezuje se s: preporuka, strah, usluga

Riječ **zubar** upotrebljava se u razgovornome, ležernijemu stilu. U standardnome je jeziku bolje upotrijebiti riječ **stomatolog**. ŽENSKO: **stomatologica :1, stomatologinja :1, zubarica :1, SINONIM: stomatolog**

tvorenice **zubarev, zubarica, zubarov, zubarski** [Kristian] [Domagoj] **zubar** imenica zu-bar ()

Zubar je muškarac koji je završio stomatološki fakultet i liječi zube.

dežurni zubar, privatni zubari

ŽENSKO: **zubarica :1** Zubar je muškarac koji je završio stomatološki fakultet i liječi zube. Zubarica je žena koja je završila stomatološki fakultet i liječi zube. Zubarice su žene koje su završile stomatološki fakultet i liječe zube. Zubari su muškarci koji su završili stomatološki fakultet i liječe zube, ali i muškarci i žene koji su završili stomatološki fakultet i liječe zube. Zubar, Lucija je zubarica, Marko i Luka su zubari, Lucija i Ana su zubarice, a Lucija, Ana, Marko i Luka su zubari.



Hrvatski

[A] [B] [C] [Č] [Ć] [D] [DŽ] [Đ] [E] [F] [G] [H] [I] [J] [K] [L] [M] [N] [N] [O] [P] [R] [S] [Š] [T] [U] [V] [Z] [Ž]

Osnovni [Stranci](#) [Djeca](#) [Sve](#)

Osnovni:

stomatòlog **stomatolog** im. m. (GA stomatòloga, DL stomatòlogu, V stomatòlože, I stomatòlogom; mn. NAV stomatòlozi, G stomatòlōgā, DLI stomatòlozima, A stomatòloge)

Stomatolog je liječnik koji se bavi stomatologijom.

Najnižu cijenu usluga određuje Stomatološka komora, a potom je svaki privatni stomatolog usklađuje sa svojim mogućnostima naplate, zarade itd., što ovisi o mnogo čimbenika. U ordinaciji stomatologa dr. Živka Dijana jučer je u Zadru održana edukacijska radionica na kojoj su prezentirane najsuвременije metode pomlađivanja lica, usana i zubnog mesa filerima Esthelis.



Kakav je stomatolog? budući, dežurni, dječji, estetski, izabrani, nezaposlen, obiteljski, odabrani, poznati, privatni, ugovorni, vrhunski; Što stomatolog može? izvaditi (zub, živac), preporučiti (terapiju, vađenje zuba, zubni konac); Što se sa stomatologom može? izabrati ga, obavijestiti ga (o lijekovima koje tko uzima, o terapiji, o trudnoći), posjetiti/posjećivati ga, preporučiti ga komu, zamoliti ga (da što preporuči/objasni); Koordinacija: ginekolozi i stomatolozi, liječnici i stomatolozi, pacijenti i stomatolozi, pedijatri i stomatolozi, stomatolozi i zubni tehničari; stomatolog ili specijalist (dentalne patologije, endodoncije, oralne kirurgije, paradontolog, stomatološke protetike, za plastičnu kirurgiju); Povezuje se s: dežurstvo, intervencija, kongres, kontrola, nadzor, pomoć, posjet, pregled, preporuka, savjet, udruga, udruženje, uputa

ŽENSKO: stomatologica :1, stomatologinja :1, zubarica :1, **SINONIM:** zubar :1 tvorenice stomatologica, stomatologov, stomatološki [josip] [Domagoj]

The Road from ŠR to *Mrežnik*

- *Mrežnik* is a web dictionary we could give more data than in the printed dictionary, **advice**, **usage**
- using WordSketches we found meanings we were not aware of before. Every new meaning has examples from the corpus, which led us to the awareness that often there are useful pragmatic comments which could be inserted into the dictionary.
- Male/female collocations
- Working with the corpus and SketchEngine our approach to synonyms and antonyms has also changed. In *Školski rječnik hrvatskoga jezika* synonymous entries resembled each other completely, i.e. they had the same example differing only in the synonymous word and they had the same collocations.

Corpus vs. System

Masculine	Feminine	English
diskdžokej	diskdžokejica	disk-jockey
knjigoveža	knjigoveškinja	bookbinder
krupje	krupjeica	croupier
mornar	mornarica	sailor
ronilac	roniteljica	diver
tekstopisac	tekstopiskinja	songwriter

Collocations and examples

- **Konobarica**

- brkata
- sisata
- zajebavati

- **Konobar**

- neljubazam
- ljubazan
- pomoćni
- priučeni
- zamoliti/moliti
- dozvati/zovnuti

Representativeness of collocations

- collocations reflect news reporting on murder, rape, drugs, etc. not really characteristic for a certain word
- fatalna pekarica fatal baker
- pohotna pekarica lusty baker

Synonyms and antonyms

- Different examples
- Different collocations

Hrvatski mrežni rječnik – Mrežnik



The Institute of Croatian Language and Linguistics, *Croatian Web Dictionary – Mrežnik*, a research project funded by the Croatian Scientific Foundation, and VERN'

will be organizing an

INTERNATIONAL CONFERENCE
E-DICTIONARIES AND E-LEXICOGRAPHY

10 and 11 May 2019

Zagreb, Croatia

Invited speakers:

Stefan Engelberg (Institut für Deutsche Sprache, Mannheim)

Simon Krek (Institut *Jožef Stefan*, Ljubljana; Center za jezikovne vire in tehnologije Univerze v Ljubljani, Ljubljana)

Polona Gantar (Filozofska fakulteta Univerze v Ljubljani, Ljubljana)

Nikola Ljubešić (Institut *Jožef Stefan*, Ljubljana)

The Institute of Croatian Language and Linguistics and the *Croatian Web Dictionary – Mrežnik*, a research project funded by the Croatian Science Foundation, invite researchers to submit paper proposals for the international conference *E-dictionaries and e-lexicography*, which will be held on 10 and 11 May 2019 in Zagreb. The aim of the conference is to bring together experts in e-lexicography, computational linguists, and software developers from Croatia and abroad.

The conference will focus on the following topics:

- e-dictionaries
- e-encyclopedias
- corpora in lexicography
- databases on language and terminology
- e-terminology
- retrodigitalization
- computer tools for lexicography

Thank you!
Questions?