

Hrvatski mrežni rječnik – Mrežnik¹

Hrvatski mrežni rječnik – Mrežnik četverogodišnji je istraživački projekt koji je započeo 1. ožujka 2017. godine.² Za hrvatski jezik ne postoji jednojezični mrežni rječnik sastavljen u skladu s pravilima suvremene e-leksikografije. Stoga je temeljni cilj projekta sastavljanje upravo takva rječnika. U izradi mrežnoga rječnika usporedno se odvijaju dvije vrste aktivnosti: 1. aktivnosti povezane s uspostavom računalnih i jezikoslovnoračunalnih pretpostavka za izradbu mrežnoga rječnika i 2. aktivnosti povezane s leksikografskom obradbom rječničke građe. Zato je potrebna tijesna suradnja leksikografa i stručnjaka za računalno jezikoslovlje te pažljivo zajedničko planiranje svih, pogotovo početnih i završnih koraka u izradbi mrežnoga rječnika. U ovome ćemo radu predstaviti neke temeljne značajke *Hrvatskoga mrežnog rječnika – Mrežnika*, za koji su dosad uspostavljeni svi računalni i računalnojezikoslovni preduvjeti za početak obrade.

1. Korpus

Mrežnik je zamišljen kao rječnik koji je na korpusu utemeljen, a nije korpusom vođen. To znači da leksikografi podatke iz korpusa uzimaju u obzir pri obradi natuknica, ali usto uzimaju u obzir i podatke koji se nalaze u ostalim dostupnim izvorima (npr. u općim i terminološkim rječnicima, različitim priručnicima i udžbenicima). *Mrežnik* se temelji na dvama korpusima: *Hrvatskoj jezičnoj riznici* (<http://riznica.ihjj.hr/index.hr.html>), korpusu Instituta za hrvatski jezik i jezikoslovlje, te korpusu *hrWac* (<http://nlp.ffzg.hr/resources/corpora/hrwac>). Podatci se iz korpusa crpe s pomoću alata za crpenje podataka SketchEnginea, koji omogućuje prikaz konteksta neke riječi putem skica riječi (dobivenih na temelju gramatike skica koju je za potrebe projekta napisao Nikola Ljubešić), najčešćih kolokacija raspoređenih u sintaktičke kategorije, te tako olakšava i usustavljuje pronalaženje dobrih rečeničnih primjera i kolokacija. Tako dobiveni podatci prenose se u program TLex, u kojemu leksikografi opisuju natuknicu na svim razinama.

¹ Ovaj je rad izrađen na istraživačkome projektu *Hrvatski mrežni rječnik – Mrežnik*, koji u cijelosti financira Hrvatska zaklada za znanost. Dijelom sadržava priloge s postera koji su predstavile autorice ovoga rada na skupu eLex 2017 održanome u Leidenu u Nizozemskoj od 19. do 21. rujna 2017.

² Projekt se provodi u ustanovi nositelju projekta Institutu za hrvatski jezik i jezikoslovlje. Voditeljica je projekta Lana Hudeček, koja je zajedno s Milicom Mihaljević i urednica *Mrežnika*, a suradnici su Matea Andrea Birtić, Goranka Blagus Bartolec, Joža Horvat, Željko Jozić, Barbara Kovačević, Kristian Lewis, Ivana Matas Ivanković, Milica Mihaljević, Bruno Nahod, Dinka Pasini, Ermina Ramadanović i Domagoj Vidović. Vanjski je suradnik na projektu Nikola Ljubešić.

list

(noun)

Croatian Web (hrWaC 2.2) freq = 190,371 (136.19 per million)

kakav?	144,346	75.82	imenica	iza	prijedloga	36,875	oba u	genitivu	32,344	16.99	veznik	19,359	10.17	koga-što	17,522	9.20
večernji +	26,263	12.05	za-s X	12,020	6.31	večernji +	6,116	12.16	čiji +	127	4.11	pisati +	5,261	10.60		
Večernji list			u-s X	11,911	6.26	Večernjeg lista			te +	794	3.80	, piše Večernji list .				
jutarnji +	23,686	11.87	na-s X	3,692	1.94	jutarnji +	5,633	11.93	lista te			okrenuti +	410	8.41		
Jutarnji list			sa-s X	2,780	1.46	Jutarnjeg lista			koji +	4,448	3.66	okrenuti novi list				
dnevni +	9,147	10.10	iz-s X	2,174	1.14	dnevni +	2,459	10.23	ili +	885	3.34	navoditi +	436	8.38		
dnevni list			od-s X	1,220	0.64	dnevnog lista			list ili			, navodi list .				
vlasnički +	2,087	8.71	po-s X	454	0.24	zadarski +	1,044	8.76	i +	8,539	3.12	doznavati +	150	8.01		
vlasnički list			do-s X	365	0.19	Zadarskog lista			list i			doznaje Večernji list				
rodan +	2,141	8.54	o-s X	278	0.15	nov +	4,335	8.29	a +	246	3.04	prenositi +	327	7.86		
rodni list			uz-s X	250	0.13	Novog lista			list a			, prenosi list				
zadarski +	2,881	8.43	preko-s X	225	0.12	većernji +	311	8.28	već	9	2.99	tvrditi +	175	7.69		
Zadarski list			protiv-s X	187	0.10	Većernjeg lista			pa +	251	2.97	, tvrdi list				
nov +	17,425	8.02	poput-s X	169	0.09	rodan +	361	8.13	list pa			javljati +	137	7.48		
Novi list			put-s X	118	0.06	rodnog lista			nakon	36	2.82	, javlja Novi list .				
lovorov +	872	7.61	između-s X	112	0.06	vlasnički +	271	7.89	list nakon što je			dodavati +	149	7.36		
lovorov list ,			prema-s X	108	0.06	vlasničkog lista			kako +	554	2.76	, dodaje list .				
britanski +	1,338	7.40	bez-s X	105	0.06	školski +	485	7.61	list kako			izdavati +	139	7.31		
britanski list			oko-s X	89	0.05	školskog lista			gdje	40	2.70	okretati +	113	7.15		
većernji +	644	7.18	zbog-s X	78	0.04	narodni +	376	7.33	listu gdje			objaviti +	349	7.01		
Većernjeg lista			kroz-s X	68	0.04	Narodnog lista			niti	29	2.63	uređivati +	100	6.80		
školski +	1,187	7.18	među-s X	55	0.03	grafički +	196	7.18	list niti			izvaditi +	111	6.61		
školski list			nad-s X	49	0.03	grafičkih listova			tako	42	2.59	zaključivati	62	6.61		
narodni +	1,103	7.14	unutar-s X	40	0.02	britanski +	228	6.92	lista tako			, zaključuje list				
Narodni list			ispod-s X	40	0.02	britanskog lista			ukoliko	16	2.56	priložiti	71	6.59		
smokvin +	537	6.92	kod-s X	37	0.02	posjedovni +	103	6.69	kao +	461	2.47	izlaziti	62	6.47		
smokvin list						posjedovnog lista			list kao			dodati +	247	6.45		
vjenčati +	528	6.85				lovorov +	101	6.66	dakle	8	2.42	saznavati	48	6.38		
vjenčani list						lovorovog lista			ni	45	2.33	saznaje list				
posjedovni +	490	6.79				putni +	111	6.63	list ni			tužiti	60	6.31		
posjedovni list						putnog lista			prija	14	2.20	odstraniti	51	6.22		
beogradski +	539	6.67				studentski +	144	6.52	jer +	222	2.19	podsjecati	44	6.09		
beogradski list						Studentskog lista			list jer			, podsjeća list				
jamstveni +	450	6.63				župni +	123	6.46	svaki +	119	2.15	isticati	78	5.79		
jamstveni list						župnog lista			što +	122	2.11	, ističe list				
ugledan +	552	6.60				šibenski +	110	6.42	dok	58	2.08	čitati +	149	5.78		
riječki +	611	6.57				" šibenskog lista "			list dok			izdati	79	5.75		
studentski +	554	6.55				smokvin	85	6.42	no	14	1.95	citirati	41	5.73		
Studentskom listu						smokvinog lista			lako	14	1.79	, citira list				
putni +	448	6.52				zelen +	157	6.41	ako	77	1.78					
služben +	859	6.46				zelenih listova			list ako							
Službeni list						vjenčati	83	6.34								
slovenski +	567	6.28				vjenčanog lista										
slovenski list						podravski	87	6.27								
krsni +	353	6.28				Podravskog lista										
krsni list						svjež +	103	6.27								
ljubljski +	373	6.27				svježih listova										
Ljubljski dnevni list						suh +	100	6.26								
						suhog lista										
						dubrovački +	117	6.23								
						Dubrovačkog lista										

 1. slika: skica riječi za riječ *list* – prikaz nekoliko stupaca

2. Abecedarij

Abecedarij za *Mrežnik* sastavljen je tako da su preklapljeni čestotni popisi riječi iz *Hrvatske jezične riznice* i *hrWaCa*. Osnovni je modul *Mrežnika* zamišljen kao rječnik od 10 000 natuknica, koje bi trebale biti obrađene tijekom sljedeće četiri godine. Abecedarij dobiven preklapanjem čestotnih popisa dopunjen je u prvome redu tako da su popunjena

neka tvorbeni gnijezda (dodani su, primjerice, mocijski i vidski parnjaci i ako se ne pojavljuju u abecedariju dobivenomu preklapanjem popisa). Usporedno su sastavljena i dva manja abecedarija, jedan od 3000 riječi (za rječnički modul namijenjen učenicima), drugi od tisuću riječi (za rječnički modul namijenjen strancima koji uče hrvatski jezik). Abecedarij modula za učenike sastavljen je na temelju školskih priručnika i udžbenika, a abecedarij modula za strance na temelju rječnika u udžbenicima za strance namijenjenim učenju hrvatskoga jezika. Riječi iz obaju manjih modula nalaze se u temeljnome modulu od 10 000 natuknica namijenjenomu odraslim govornicima hrvatskoga jezika.

lisnica	Nc	1225	lisnato	Rg	2
lisnicin	As	0	lisni	Ag	148
lisnik	Nc	31	lisnica	Nc	304
lisnikov	As	1	lisnik	Nc	10
lisno	Rg	9	lisno	Rg	1
list	Nc	190141	list	Nc	17099
lista	Nc	167625	lista	Nc	12986
listak	Nc	56	listak	Nc	34
listanje	Nc	1555	listanje	Nc	66
listati	Vm	6980	listati	Vm	482

2. slika: abecedarij s označenom natuknicom *list*

3. Rječnički članak

Hrvatski mrežni rječnik – *Mrežnik* uključuje rječničke članke koji imaju naglašenu natuknicu, koja je i poveznica na korpus, gramatičku definiciju i gramatički blok s naglašenim sustavno donesenim oblicima, bogato razvedene definicije značenja oprimjerene primjerima iz korpusa s pomoću SketchEngineova modula GDEX, s pomoću kojega se temeljem jasnih kriterija izabiru najbolji primjeri, najčešće kolokacije (usustavljene s pomoću pitanja poput *Kakav je x?*, *Što x radi?*, *Što se s x radi?* – vidi obradu kolokacija u rječničkome članku u primjeru unutarječničkoga i izvanječničkoga povezivanja u *Mrežniku* na kraju rada), podatke o značenjskim odnosima u koje stupa riječ ili sveza u određenome značenju (podatke o sinonimima, antonimima, hiponimima, meronimima itd.), definirane i oprimjerene sveze uz koje se također donose kolokacije i podatci o značenjskim odnosima, uporabne napomene te definirane i oprimjerene frazeme. Sustavom odrednica i uputnica upućuje se na stilski, terminološki i/ili normativni status natuknice. *Mrežnik* će imati normativnu dimenziju, što pokazuje i činjenica da je u sklopu projekta predviđeno i povezivanje na savjete iz *Jezičnoga*

savjetnika na adresi jezikni-savjetnik.hr, u kojima se pobliže objašnjavaju razlozi zašto se koja riječ ili njezina uporaba u određenome značenju ne preporučuje u standardnome jeziku. S natuknica ili sveza ili njihovih značenja čija se uporaba u standardnome jeziku ne preporučuje moći će se putem poveznica pristupiti jezičnomu savjetu na stranicama *Jezični savjetnik* ili *Bolje je hrvatski*, a s naziva na bazu *Struna*. Sukladno potrebama korisnika razvijena je i struktura rječničkoga članka u modulu za učenike i za strance koji uče hrvatski. Uz bogato unutarječničko povezivanje predviđeno je i povezivanje *Mrežnika* s vanjskim izvorima.

4. Tri modula

Mrežnik će uključivati tri modula: za odrasle govornike hrvatskoga jezika, za učenike osnovne škole i za strance koji uče hrvatski jezik. Za ta su tri modula uspostavljena tri modela strukture natuknice, pa će se ti moduli osim po broju uključenih riječi razlikovati i po načinu obrade tih riječi (opsegu i vrsti gramatičkih podataka, ustroju i opsegu definicije značenja, značenjskoj razvedenosti, primjerima, povezivanju s drugim natuknicama i izvanrječničkim izvorima, uspostavi značenjskih odnosa s drugim natuknicama i riječima koje neće biti zastupljene u *Mrežniku*, po svezama i frazemima koji se u pojedinome modulu donose te njihovoj obradi, po vrsti i opsegu napomena o uporabi – pragmatičkih napomena). S obzirom na to da je jedna od bitnih značajka *Mrežnika* povezivanje s vanjskim izvorima koji postoje u Institutu za hrvatski jezik i jezikoslovlje ili koji će se uspostaviti unutar projekta izrade *Hrvatskoga mrežnog rječnika*, tri će se modula razlikovati i po tome s kojim se izvorima povezuju.

5. Povezivanje s vanjskim izvorima

Usporedno s radom na *Mrežniku* u okviru projekta predviđena je i izrada baze veznika s opisom vezničkih skupina i modifikatora, bazom objašnjenja podrijetla frazema i bazom etnika i ktetika. Te će se baze povezivati s *Mrežnikom* tako da će se uz veznike obrađene u *Mrežniku* moći pristupiti opisu njihovih skupina i modifikatora, uz zemljopisna imena bazi etnika i ktetika, uz frazeme bazi objašnjenja podrijetla frazema itd. S njima će se povezivati i izvori koji već postoje ili se izrađuju u Institutu za hrvatski jezik i jezikoslovlje. To su: *Jezični savjetnik* (jezikni-savjetnik.hr), baza glagolskih valencija *e-Glava* (valencije.ihjj.hr), baza hrvatskoga strukovnog nazivlja *Struna* (struna.ihjj.hr), *Kolokacijska baza hrvatskoga jezika* (ihjj.hr/kolokacije), *Repozitorij hrvatskih metafora* (ihjj.hr/metafore) i mrežne stranice *Bolje je hrvatski* (bolje.hr). Povezivanje s vanjskim izvorima prikazano je u primjeru na kraju članka. Za osnovni je modul predviđeno povezivanje sa svim vanjskim izvorima, za modul namijenjen učenicima povezivanje s objašnjenjima frazema i jezičnim savjetima sa stranica *Hrvatski u školi* (hrvatski.hr), koji će se na tim stranicama uspostaviti tijekom projektne razdoblja, za modul za strance koji uče hrvatski povezivanje s objašnjenjima frazema i jezičnim savjetima.

6. Unutarječničko povezivanje

U *Mrežniku* međusobno će s pomoću poveznica biti povezani sinonimi, antonimi, hiponimi, meronimi itd. Model povezivanja sinonima unutar *Mrežnika* pokazuje se također u

primjeru na kraju članka, u kojemu je prikazano povezivanje sinonima *list* (u značenju ‘dokument kojim se što dokazuje i potvrđuje’) i *isprava*. Velika će se pozornost pridavati značenjskomu odnosu među riječima. Povezivat će se i mocijski parnjaci (npr. *učitelj* i *učiteljica*, *psiholog* i *psihologinja*), ali i muško-ženski parovi koji nisu mocijski parnjaci (*jelen* i *košuta*, *pas* i *kuja*, *medicinska sestra* i *medicinski tehničar*). Povezivat će se i vidski parnjaci (*pasti* i *padati*, *ocijeniti* i *ocjenjivati*).



3. slika: isječak iz pokusne inačice odostražnoga rječnika

7. Odostražni rječnik

U okviru projekta predviđena je i izrada odostražnoga rječnika, odnosno rječnika u kojemu su natuknice poredane abecednim redom, ali su abecedirane od svojega kraja. Takav je rječnik veoma koristan proučavateljima hrvatske tvorbe, pjesnicima, rješavateljima križaljka itd. Obuhvaćat će svih 10 000 riječi. Na 3. slici prikazana je pokusna inačica odostražnoga rječnika, koju je izradio Josip Mihaljević.

8. Zaključak

U ovome kratkom pregledu željele smo upozoriti na neke ključne značajke *Hrvatskoga mrežnog rječnika – Mrežnika*, koji se izrađuje u okviru istoimenoga projekta. Nakon završetka leksikografske obradbe podatci će biti izvezeni iz TLexa za mrežnu aplikaciju te u repozitorij europske znanstvene infrastrukture CLARIN (repozitorij clarin.si i javni sustav za upravljanje podacima github.com). Time će *Mrežnik* biti učinjen dostupnim i za uporabu putem mrežne aplikacije i za strojnu primjenu preuzimanjem podataka iz repozitorija CLARIN. Važno je napomenuti da je *Mrežnik* osmišljen tako da se i nakon završetka četverogodišnjega razdoblja (koliko traje projekt Hrvatske zaklade za znanost) može nadopunjavati te je osnovni cilj projekta uspostaviti znanstveno utemeljen i tehnološki održiv okvir za daljnju izgradnju hrvatskoga jednojezičnog znanstveno utemeljenoga i na spoznajama suvremene e-leksikografije izrađenoga mrežnog rječnika.

Na slikama na sljedećim se dvjema stranicama donosi obrada natuknice *list* u programu TLex i njezin prikaz te prikaz povezivanja te natuknice sa sadržajem unutar i izvan rječnika.

izvanrječničkoga povezivanja u *Mrežniku*

Hrvatski [A][B][C][D][E][F][G][H][I][J][K][L][M][N][O][P][R][S][T][U][V][Z]

list¹ list *im. m.* (G lista, A list/lista, L listu/listu; mn. listovi, G listová)

¹ (A list, L listu) List je dio biljke koji je tanka plosnata izraslina na stabljici, najčešće zelene boje, koja fotosintezom biljku opskrbljuje hranom. *Listovi i peteljke pokriveni su žarnicama, i zbog toga ih je jako neugodno dodirnuti, jer izliju oštar sok na kožu. Da podsjetimo, u blagovorno djelovanje na cijeli organizam, maslinovo se ulje u narodnoj medicini koristi za uklanjanje žučnih kamenaca, a čaj od osušanih ili svježih listova masline snižuje krvni tlak kao i šećer u krvi.*

Kakav je list? Zelen, svjež, suh, kožast, zelen, nazubljen, uzak, ovalan, perast, jajast, tamnozelen, dlakav, srcolik, gladak, sjedeći, mesnat, ljekovit, krupan, jestiv, narezan, nasjeckan, usitnjen, prošaran (ružičastom bojom, bez ornamentima, crno prošaran); Čiji je list? lorovor list, smokvin list, list masline, list peršina, list kadulje, list maslačka, list bršljana, list duhana; Što se s listom radi? usitnjuje, sjeka, bere, ekstrahira, kosa, reže, dodaje čemu, prokuhava; Što list može? požutjeti, osušiti se, otpasti, izrasti; Što list ima? žilice, bazu, plojku, peteljku

² (A list, L listu) List je tanak komad čega, ono što izgledom podsjeća na istoimenu dio biljke. *Stavimo dva tanka lista pršuta i dva listića sira i zavijemo sve zajedno kao palačinku. List tankog tijesta odnosno kore poprskajte rastopljenim maslacem ili uljem, stavite drugi list tankog tijesta, također poprskajte maslacem ili uljem, premažite dijelom nadjeva i savijte u savijaču. Stavite po 5 listova lazanja i kuhajte ih 2 - 3 minute (ovisno o vrsti lazanja, provjerite upute na pakiranju).*

Kakav je list? tanak, tanji, deblji; Što se s listom može? razvući, premazati, saviti/sviti, razvaljati, nadjenuti; Čega je komad list? tijesta, kore, lazanje, drveta, pršuta, sira

list papira

List papira jedan je papir. *Brcu je razveo jedan fascikl, izvadio list papira i pogledao ga prema svjetlu. NAPOMENA : Potrebno je ispisati obje stranice pristupnice na isti list papira te popuniti samo prvu stranu pristupnice (drugu stranicu popunjava administracija).*

Kakav je list papira? prazan, bijel, malen, velik, napola prerezan, presavijen, zgužvan, poderan, ispisan; Što list papira ima? strane; Što se s listom papira može? pisati po njemu, napisati što na njega, ispisati ga, isprintati ga, zgužvati ga, poderati ga, odrezati ga, presaviti ga

³ (A list, L listu) List je dokument kojim se što potvrđuje ili dokazuje.

^a sinonim **isprava**

posjedovni list

Jezični savjetnik: <http://jezicni-savjetnik.hr/>

Posjedovni je list dokument koji sadržava podatke o svim katastarskim česticama koje čine jedno zemljišno-knjižno tijelo te sve katastarske promjene koje se odnose na njih. *U jednom posjedovnom listu može biti upisano više katastarskih čestica, odnosno jedna katastarska čestica može biti upisana u više posjedovnih listova.*

Što se sve s posjedovnim listom može? zatražiti, izvaditi

vlasnički list

Jezični savjetnik: <http://jezicni-savjetnik.hr/>

Vlasnički je list dokument koji sadržava podatke o nekretnini upisane u zemljišnu knjigu. *Na oglas se javilo nekoliko potencijalnih kupaca, kojima je lažni vlasnik pokazivao krivotvorenog vlasnika, vlasnički list, te prijedlog kupoprodaje nekretnine.*

Kakav može biti vlasnički list? krivotvoren; Što se sve s vlasničkim listom može? izvaditi ga, krivotvoriti ga, zatražiti ga

smrtni list

Jezični savjetnik: <http://jezicni-savjetnik.hr/>

Smrtni je list dokument kojim se potvrđuje čija smrt. *Naravno, službenici bolnice izdali su smrtni list s Gibovićevim imenom, obaviještena je podnoženje zahtjeva možete kupiti u " Narodnim novinama ", a prilikom predaje popunjene knjižice osobe od koje se izvodi pravo na mirovinu, smrtni list - vjenčani list, rođne listovi prizna pravo na mirovinu te školske potvrde za djecu na redovnom školovanju, izdani jesu li neku drugu mirovinu.*

⁴ (A list, listu) List je redovito tiskano glasilo. (...)

list-ovi tvorenice listu, listak, listati, listić, listina, listonoša, listopad, listić

Kolokacijska baza hrvatskoga jezika: <http://ihj.hinkolokacije/search?q=list>

Struna: <http://struna.ihj.hr/search-do?q=list&container>

Kolokacijska baza hrvatskoga jezika | Home | O nas

list

Rezultati pretraživanja za: list

natuknica	vrsta riječi	tekst	oznake
list		žilice lista	
list		školski list	
list		Suplikavost lista	
list		baza lista	
list		bobov list	
list		bodjkaost list	
list		cepjetei kao jasenov list	F
list		cjevasti listovi	
list		crtarjama ispisati list	
list		djetelina s četiri lista	

povezivanje natuknice *list* s *Kolokacijskom bazom hrvatskoga jezika*

isprava isprava *im. ž.* (G isprave, mn. N isprave, G ispravā)

Isprava je službeni dokument kojim se što potvrđuje i koji služi kao dokaz čemu. *Podignuvši pušku, vojnik je najprije zatražio isprave, a onda, vidjevši da pred sobom ima gotovo još dječaka, zapitao za razlog te žurbe. Matična knjiga, svjedodžbe i učenička knjižica jesu javne isprave.*

Kakva može biti isprava? putna, službena, javna, knjigovodstvena, osobna, vjerodostojna, krivotvorena, identifikacijska, valjana, važeća*, nevaljana, nevažeća*, obračunska, lažna, ovršna, proglašena nevaljanom, proglašena nevažećom*, ovrjena, izdana; Što se sve s ispravom može? krivotvoriti je, sastaviti je, izdati je, imati je, posjedovati je, nemati je, pregledati je, zatražiti je, sakriti* je, skriti je, predočiti je, pokazati je, dobiti je, ovjeriti je, čuvati je, dati je, izvaditi je

unutarrječničko povezivanje natuknice *list* s natuknicom *isprava*