

Lana Hudeček

Milica Mihaljević

POJMOVNIK E-LEKSIKOGRAFIJE

Pojmovnik e-leksikografije izrađen je u sklopu projekta *Hrvatski mrežni rječnik – Mrežnik* u suradnji s projektom *Hrvatsko jezikoslovno nazivlje – Jena*.¹ Dostupan je na stranicama projekta na adresi <http://ihjj.hr/mreznik/page/pojmovnik/6/>. *Pojmovnik* uključuje e-leksikografske nazive i imena relevantna za e-leksikografiju i iz njega se većina naziva (ne imena) preuzima u bazu *Jena*. Vrela za crpenje naziva radovi su iz područja e-leksikografije te *Jezikoslovni korpus*.² Obrade naziva iz *Pojmovnika* prilagođuju se potrebama projekta *Jena*.³ Na 1. i 2. slici prikazane su natuknice *regularni izraz* u *Pojmovniku e-leksikografije* i u *Jeni*.

- **regex** v. regularni izraz
- **regularni izraz** (*engl.* regular expression) zadani niz znakova čija se kombinacija može upotrijebiti za pronalaženje određenih dijelova teksta, izraz koji se uspostavlja za pretraživanje korpusa s pomoću alata SketchEngine i NoSketchEngine za traženje ciljanih gramatičkih ili leksičkih uzoraka (popis regularnih izraza dostupan je na <https://sketchengine.co.uk/documentation/corpus-querying/>). Regularni izrazi iz SketchEngineova kalendara za 2018.:



1. slika: Obrada natuknice *regularni izraz* u *Pojmovniku e-leksikografije*

¹ Više o projektu *Hrvatsko jezikoslovno nazivlje – Jena* vidi na projektnim stranicama jena.jezik.hr.

² Više o *Jezikoslovnome korpusu* vidi u Marković i Mihaljević 2020. te u Mihaljević, Marković i Mihaljević 2020.

³ Više o obradi jezikoslovnih naziva u *Jeni* i *Mrežniku* vidi u Hudeček i Mihaljević 2019.

regularni izraz	
status naziva: naziv obrađivač: Josipa Mioč faza obradbe: zaključano autorski: ne	
definicija: zadani niz znakova čija se kombinacija može upotrijebiti za pronalaženje određenih dijelova teksta	
radna definicija: zadani niz znakova čija se kombinacija može upotrijebiti za pronalaženje određenih dijelova teksta	
vrela definicije: Pojmovnik, Mrežnik. (mrežno)	
projekt: jezikoslovlje	
potpodručje: e-leksikografija i korpusno jezikoslovlje	
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"></div> <div> <p>istovrijednica - engleski: regular expression; regex</p> <p>istovrijednica - njemački: regulärer Ausdruck</p> <p>istovrijednica - francuski: expression régulière</p> <p>istovrijednica - ruski: регулярное выражение</p> </div> </div>	
jezična odrednica: višerječni naziv	
poveznica: https://sketchengine.co.uk/documentation/corpus-querying/	
napomena: Regularni izraz uspostavlja se za pretraživanje korpusa s pomoću alata SketchEngine i NoSketchEngine za traženje ciljanih gramatičkih ili leksičkih uzoraka.	
Brzi pregled naziva ✕	

2. slika: Obrada natuknice *regularni izraz* u *Jeni*

Uz natuknicu u *Pojmovniku* se donosi engleska istovrijednica (u zagradama) te kratka definicija leksikonskoga tipa. Uz natuknicu se katkad nalazi i slika. Kao natuknice bilježe se svi sinonimni nazivi, a s pomoću uputnice v. (vidi), upućuje se na naziv pod kojim se nalazi obrada. Status preporučenoga naziva (naziva pod kojim se nalazi obrada), kao i u *Jeni*, naziv dobiva na temelju analize provedene s pomoću terminoloških načela (Hudeček i Mihaljević 2012.).

U *Jeni* se donosi kraća rječnička definicija, navodi vrela definicije, dodatni podatci daju se u napomeni. Natuknica je uvijek preporučeni naziv, a sinonimni se nazivi bilježe u rubrikama dopušteni naziv, nepreporučeni naziv, žargonski naziv, zastarjelica. Donose se engleska, njemačka, francuska, ruska i švedska istovrijednica.

Specifičnost je e-leksikografskoga nazivlja da se u hrvatskome jeziku za određeni pojam veoma često upotrebljava (i) engleski naziv. Jedno je (i prvo) od terminoloških načela da domaći naziv ima prednost pred stranim. Na temelju toga načela prednost se npr. daje hrvatskim nazivima *crpenje podataka* pred *data mining*, *gramatika skica* pred *Sketch Grammar*, *izbornik* pred *meni* i *menu*, *jezični izvori* pred *jezični resursi*, *označivanje pogrešaka* pred *error tagging*, *pametna leksikografija* pred *smart lexicography*, *pojavnica* pred *token*, *poveznica* pred *link*, *pojavnica* pred *token*, *povezani podatci* pred *linked data*, *preklapanje n-grama* pred *n-gram overlap*, *pretraživač* pred *search engine*, *rastući rječnik* pred *growing dictionary*, *rječnički portal* pred *agregator*, *skice riječi* pred *Word Sketches*, *unutarrječničko povezivanje* pred *cross-referencing*. Prednost se daje i nazivu *sinonimni skup* pred prevedenicom *sinskup* (prema engleskome *synset*).

Međutim, neki su hrvatski nazivi nastali prilagodbom engleskih naziva te su u e-leksikografskome nazivlju posve prihvaćeni. To su npr. nazivi *parsiranje*, *parsirati*, *parser*, *tag*, *tager*, *tagiranje*, *tagirati*.

Prednost se daje i hrvatskomu nazivu pred internacionalizmom, npr. nazivu *čestoća* pred *frekvencija* i *frekventnost*.

Prednost se daje i nazivu usklađenu s normama standardnoga jezika pred onim koji to nije, npr. nazivu *govorni korpus* pred nazivom *govoreni korpus*, nazivu *konkordancija* pred nazivima *konkordanca* i *konkordansa*, nazivu *parsemsko stablo* pred nazivom *stablo parsema*.

Među jednako normativno prihvatljivim nazivima od kojih jedan nije strani, prednost se daje bolje prihvaćenomu, npr. nazivu *odostražni rječnik* pred nazivom *odostražnik*.

Prednost se daje i manje metaforičnom nazivu, npr. nazivu *crpenje podataka* pred nazivom *rudarenje podataka*.

Osim naziva u *Pojmovniku e-leksikografije* donose se kao natuknice i imena (e-rječnika, e-leksikografskih projekata i inicijativa, skupova te udruga). *Pojmovnik* se, i kad je riječ o nazivima i kad je riječ o imenima, proširuje opširnijim podacima dostupnim putem poveznice koja se nalazi na kraju primarnoga pojmovničkoga članka, vidi 1. tablicu. Natuknica je u pravilu poveznica na stranice rječnika, udruge itd. koje označuje. U 1. tablici prikazuje se prva i druga razina obrade natuknice ELEXIS u *Pojmovniku e-leksikografije*.

1. tablica: Prva i druga razina obrade natuknice ELEXIS u *Pojmovniku e-leksikografije*

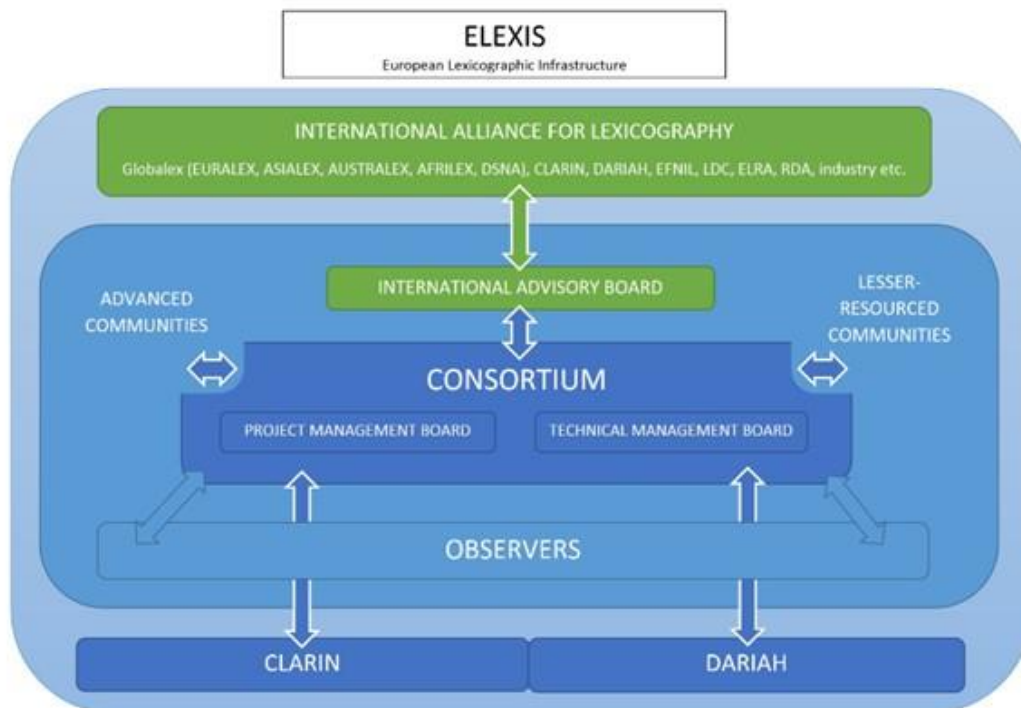
1. razina obrade natuknice ELEXIS

ELEXIS (European Lexicographic Infrastructure, Europska leksikografska struktura) platforma je i projekt čiji je nastanak potaknut činjenicom da u Europi postoji velik broj u leksikografskome radu nedovoljno povezanih i koordiniranih ustanova u kojima se stvaraju rječnici ili prikupljaju leksikografski podatci **eNeL** (European Network of e-Lexicography) > Europska mreža za e-leksikografiju [doznaj više...]

2. razina obrade natuknice ELEXIS

ELEXIS, Europska leksikografska infrastruktura – **ELEXIS** (European Lexicographic Infrastructure) platforma je i projekt čiji je nastanak potaknut činjenicom da u Europi postoji velik broj u leksikografskome radu nedovoljno povezanih i koordiniranih ustanova u kojima se stvaraju rječnici ili prikupljaju leksikografski podatci. Potreba za izgradnjom europske

leksikografske infrastrukture jasno je izražena osobito tijekom COST-ove akcije *Europska leksikografska mreža* (*European Network of e-Lexicography COST Action*), koja je završila 2017. godine. Cilj je projekta izgradnjom održive infrastrukture uskladiti napore u području leksikografije (i s obzirom na rječnike koji tek nastaju u digitalnome dobu i s obzirom na one koji su nastali prije njega), omogućiti jednostavan pristup leksičkim podacima, premostiti razlike između u tehnološkome smislu razvijenijih i manje razvijenih znanstvenih zajednica, odnosno povećati kvalitetu leksikografskoga rada, u prvome redu razvojem javno dostupnih alata koji bi smanjili vrijeme i troškove potrebne za posuvremenjivanje postojećih izvora odnosno za razvoj novih te oblikovati jedinstveni standard za izradu rječnika, razviti strategije, alate i standarde za strukturiranje i povezivanje leksikografskih izvora kako bi mogli svojim punim potencijalom, u kontekstu digitalne humanistike, doprinijeti objavljivanju uzajamno povezanih strukturiranih podataka (Linked Open Data, LOD) i uspostavi semantičkih mreža (Semantic Web), skupova značenja i pojmova koji su u nekoj mjeri povezani sa središnjim značenjem. Koordinator je projekta Institut Jožef Stefan u Ljubljani.



1. slika: Ustroj ELEXIS-a (Preuzeto iz Krek, Simon i dr. 2018. *European Lexicographic Infrastructure (ELEXIS)*. *Lexicography in Global Contexts. Proceedings of the XVIII EURALEX*. Ur. Čibej, Jaka; Gorjanc, Vojko; Kosem, Iztok; Krek, Simon)

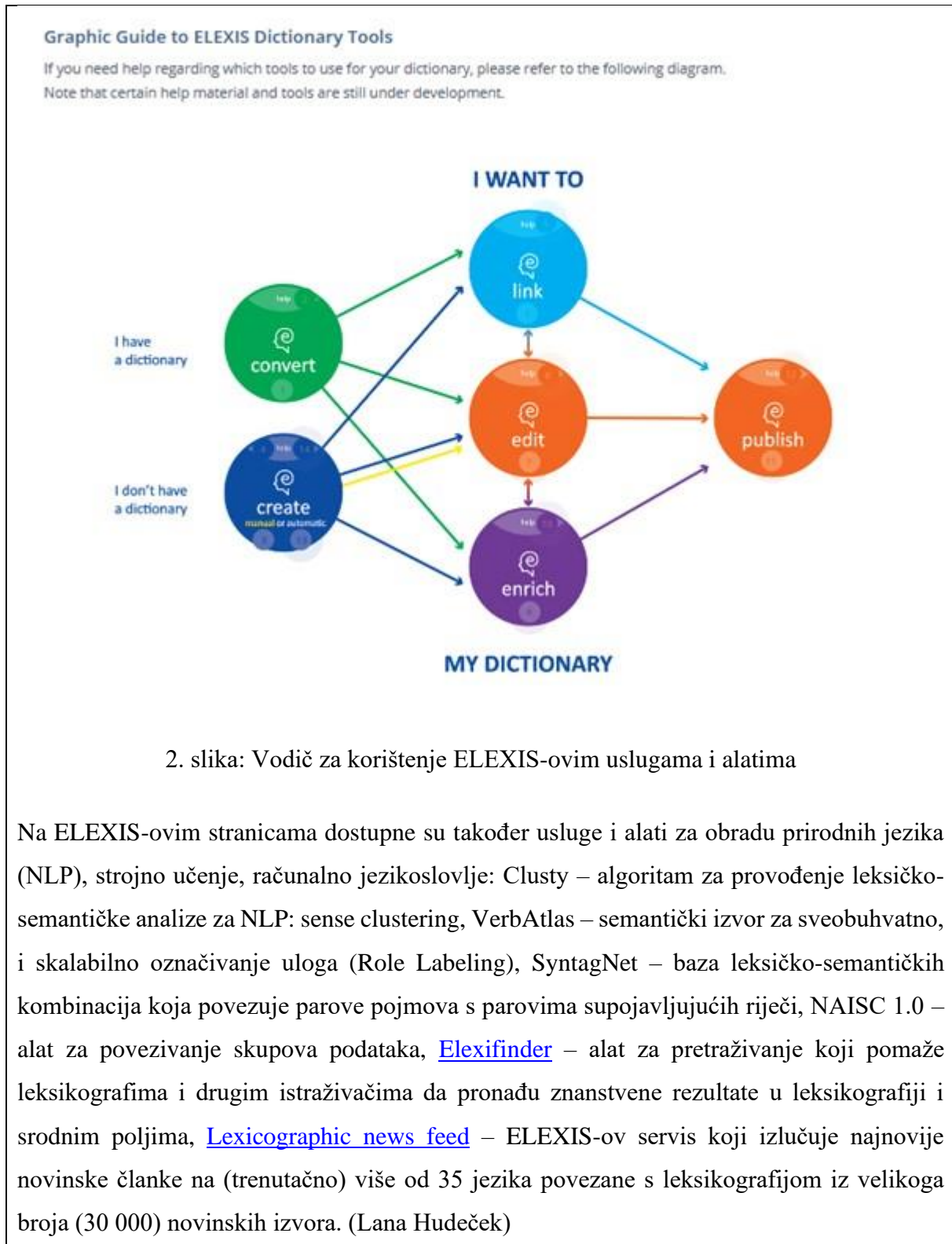
Na stranicama ELEXIS-a opisuju se trenutno sljedeći leksikografski alati i usluge (uz smjernice za korištenje njima; shemu koja pomaže korisniku vidi na 2. slici):

SketchEngine – alat za pretragu i izgradnju korpusa tvrtke Lexical Computing, koji sadržava mnoštvo drugih alata koji omogućuju analizu velikih korpusa (između ostaloga i s pomoću skica riječi), stvaranje novih korpusa te potpuno automatiziranu izradu rječnika.

Lexonomy – sustav za sastavljanje i mrežno objavljivanje rječnika utemeljen na oblačnome računalstvu; može se prilagoditi velikim (općim rječnicima) i manjim leksikografskim projektima (specijaliziranim ili terminološkim rječnicima i glosarima) povezan sa SketchEngineom tako da SketchEngine može poslati leksikografske podatke u Lexonomy kako bi se dobili automatski generirani rječnički nacrti te tako da Lexonomy može izvlačiti podatke iz SketchEngineovih korpusa tijekom procesa sastavljanja rječničkih članaka.

OneClickDictionary (OCD) – modul za sastavljanje skice riječi koji povezuje sustav za upravljanje korpusom (SketchEngine ili noSketch Engine) s Lexonomyjem i omogućuje automatsko stvaranje rječničkoga nacrta koji sadržava osnovne dijelove rječničkoga članka (natuknice, oblike riječi, primjere, kolokacije itd.), a koji u Lexonomyju dalje uređuje leksikograf koji se u potpunosti može usmjeriti na središnju fazu leksikografskoga rada.

Elexifier – sustav za rječničku konverziju utemeljen na oblačnome računalstvu koji s pomoću naprednoga XML parsiranja i strojnoga učenja pomaže u pretvaranju rječnika u formatu .pdf ili .xml u standardizirani računalno čitljiv oblik.



Literatura

- Hudeček, Lana; Mihaljević, Milica. 2012. *Terminološki priručnik*. Institut za hrvatski jezik i jezikoslovlje. Zagreb.
- Hudeček, Lana; Mihaljević, Milica. 2019. *Croatian Web Dictionary – Mrežnik vs. Croatian Linguistic Terminology – Jena*. *INFuture 2019: Knowledge in the Digital Age*. Ur. Bago, Petra; Hebrang Grgić, Ivana; Ivanjko, Tomislav; Juričić, Vedran; Miklošević, Željka; Stublić, Helena. Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences. Zagreb. 22–31.
- Marković, Mario; Mihaljević, Josip. 2020. Korpus *Jene* – jednojezični terminološki korpus. *Hrvatsko jezikoslovno nazivlje – JENA*. Ur. Mihaljević, Milica; Hudeček, Lana; Jozić, Željko. Institut za hrvatski jezik i jezikoslovlje. Zagreb. 110–123. <https://jena.jezik.hr/wp-content/uploads/2021/01/Monografija-1.pdf> (pristupljeno 3. svibnja 2022.)
- Marković, Mario; Mihaljević, Josip; Mihaljević, Milica 2020. Kako pronaći jezikoslovni naziv. *Hrvatski jezik* 7/1. 18–22.
- Mihaljević, Milica; Hudeček, Lana; Jozić, Željko (ur.). 2020. *Hrvatsko jezikoslovno nazivlje*. Institut za hrvatski jezik i jezikoslovlje. Zagreb.