



Center za
jezikovne vire
in tehnologije

Digitalno slovaropisje: **Kdo je slovar?**

Špela Arhar Holdt & Jaka Čibej

Filozofska fakulteta,
Fakulteta za računalništvo in informatiko

spela.arhar@cjvt.si

jaka.cibej@cjvt.si



Univerza v Ljubljani
Fakulteta *za računalništvo
in informatiko*



Okvir CJVT

- Center za jezikovne vire in tehnologije Univerze v Ljubljani.
- Združuje pet fakultet UL: Filozofska fakulteta, Fakulteta za računalništvo in informatiko, Pedagoška fakulteta, Fakulteta za družbene vede in Fakulteta za elektrotehniko.
- Namen CJVT je omogočiti uspešen razvoj slovenščine v digitalni dobi.
- Pripravljamo jezikovne vire ter orodja, ki so vsebinsko in metodološko **sodobni**, uporabniku **prijazni** in **odprto dostopni** za uporabo.
- Najnovejši izdelek: Slovar sopomenk sodobne slovenščine.
- <https://www.cjvt.si/>

Uporabniške raziskave na CJVT

- Prepričanja slovaropiscev o uporabniških zmožnostih in potrebah so primerljiva predstavam angleške princese Anne, kako je živeti ob denarni socialni pomoči: v grobem so lahko kar blizu resnice, ko pride do podrobnosti, pa **še zdaleč ne** (Hatherall 1984: 183).
- V evropskem prostoru uveljavljeno področje (Arhar Holdt 2015).
- CJVT je prva slovenska raziskovalna inštitucija, ki se ciljno ukvarja z empiričnimi analizami **težav, mnenj, navad** in **potreb** jezikovne skupnosti v kontekstu priprave digitalnih jezikovnih virov za slovenščino.
- European Network for Combining Language Learning with Crowdsourcing Techniques (http://www.cost.eu/COST_Actions/ca/CA16105).
- Pri pripravi virov želimo upoštevati razlike med ciljnimi skupinami (Arhar Holdt, Kosem in Gantar 2016).

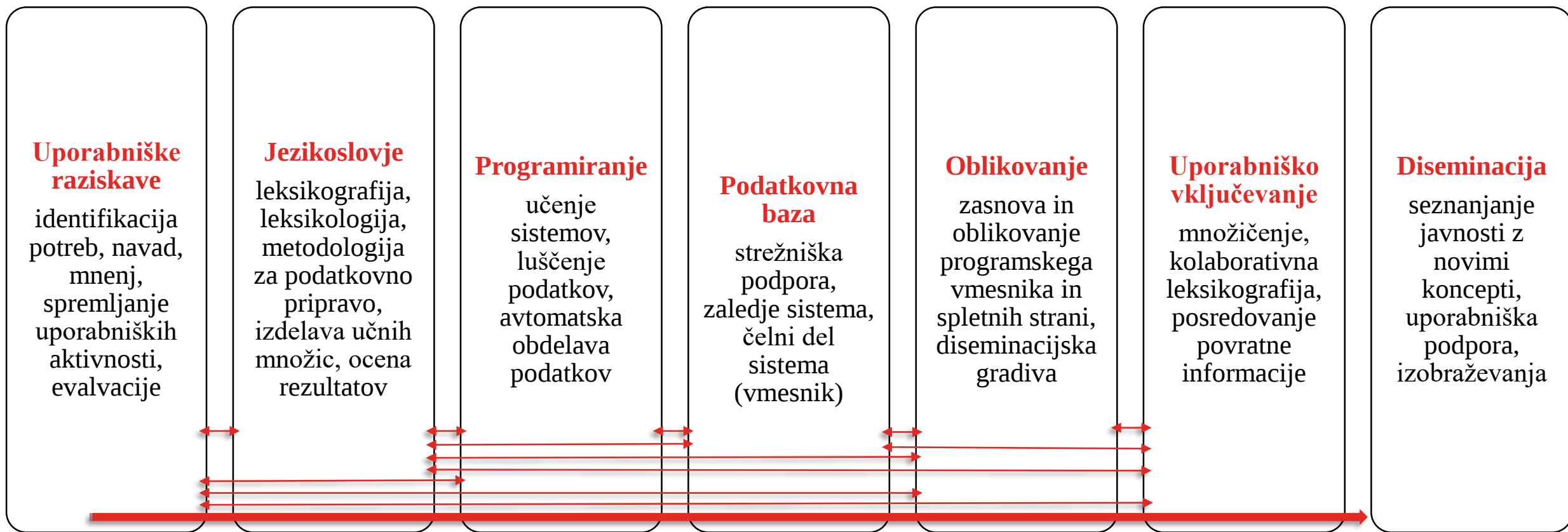
Mednarodna slovarska raziskava

- Mednarodna raziskava (29 držav, 58 raziskovalcev) o odnosu jezikovnih uporabnikov do splošnih enojezičnih slovarjev (Kosem 2017).
- V sklopu dejavnosti Evropske mreže za e-leksikografijo (ENeL):
<http://www.elexicography.eu/events/european-survey-on-dictionary-use/>.
- Anketa, v katerih situacijah se splošni enojezični slovarji uporabljajo, kakšen je uporabniški odnos do tovrstnih priročnikov, načinov njihovega financiranja in katere značilnosti slovarjev so v sodobnem času uporabnikom prioritete.
- Dodatni (samo slovenski) del še 6 vprašanj (npr. odnos do množičenja).
- Med 9. majem in 9. julijem 2017 je anketo v Sloveniji izpolnilo **619** posameznic in posameznikov.

Financiranje in raba slovarjev

- Priprava [splošnih] enojezičnih slovarjev je pogosto financirana iz javnega denarja. Se vam zdi to smiselno?
- Koliko ste pripravljeni plačati za [splošni enojezični] slovar?
- Če za slovar niste pripravljeni plačati, vpišite 0 EUR.
- Kako [vse] dostopate do enojezičnih slovarjev, ki jih uporabljate?
- Katera oblika vam je za enojezične [splošne] slovarje najljubša?

Digitalno slovaropisje: kdo je slovar?



Slovar sopomenk sodobne slovenščine

- **Odzivnost:** (a) na spremembe v jeziku – sodobna metodologija omogoča, da se spremembe v jezikovni rabi hitro zaznajo in vključijo v slovar, (b) na jezikovno skupnost – vmesnik omogoča napredno vključevanje znanja in mnenj slovarskih uporabnikov.
- **Sodobnost:** temelji na sodobnih jezikovnih podatkih in je pripravljen z metodami, ki so interdisciplinarne in inovativne; računalniško podprta priprava podatkov je bistveno manj finančno potratna kot ročno delo, neizmerno hitrejša (ažurnost) in za določene naloge celo natančnejša.
- **Odprtost:** Slovar sopomenk sodobne slovenščine je največja odprto dostopna zbirka sopomenk za slovenščino. Slovarske podatke je mogoče uporabljati za izdelavo novih jezikovnih virov, tehnologij in drugih izdelkov, ki spodbujajo celostni razvoj slovenščine v digitalnem svetu.

Slovar sopomenk sodobne slovenščine

- **Povezljivost:** za razumevanje sopomenskosti je nujno sobesedilo, zato je v slovarju veliko povezav s korpusnimi podatki. Povsem nova za slovenščino je možnost primerjave, kako se različne sopomenke pojavljajo v realnih besedilih. Povezovanje med podatki zagotavlja, da je raziskovanje slovenščine v digitalnem svetu bolj informativno in s tem bolj zanesljivo, hkrati pa udobnejše in manj zamudno.
- **Skupnost:** digitalno okolje podpira vzpostavljanje skupnosti, ki slovar ne le uporablja, ampak ga tudi razvija; predlogi in povratne informacije so upoštevane pri nadgradnjah in nadaljnjem razvoju. Rešitev upošteva raziskave o željah, navadah in potrebah slovarskih uporabnikov v digitalnem svetu, ki poleg slovarske informativnosti v ospredje postavljajo tudi uporabnost in demokratičnost.

Odzivni slovar: opredelitev

- Razvit iz digitalnega medija za digitalni medij.
- Odziven na jezikovni razvoj in mnenje jezikovne skupnosti.
- Izhodiščna baza je zasnovana avtomatsko z izrabo obstoječega jezikovnega gradiva in po evalvaciji uporabnosti ponujena javnosti.
- Vmesnik omogoča uporabniško vključevanje na ravni množičenja in kolaborativne leksikografije.
- Slovar se skladno z razvojem jezika neprestano razvija in dograjuje; spremembe so transparentno beležene s časovnimi žigi, različice arhivirane.
- Uporabniški doprinos in aktivnosti v slovarju se spremljajo in upoštevajo skladno z natančno opredeljeno metodologijo za razvoj slovarja.
- Podatkovna baza je odprto na voljo (npr. pod licenco CC-BY-SA 4.0).

Najpogostejši pomisleki

- Z avtomatskimi postopki dobimo nekvalitetne podatke.
- Bodo slovarje sedaj delali računalničarji? Kje je mesto jezikoslovja?
- Imajo uporabniki zadostne kompetence, da delajo slovar?
- Ali je slovenščina dovolj velik jezik, da bo model uspel?
- Ali so uporabniki res motivirani za takšno delo?

Od prevodnih ustreznic k Sopomenkam

- Oxford DZS – prevodne ustreznice angleških besed in besednih zvez

dom; bivališče, domovanje; stanovanje; hiša

Od prevodnih ustreznic k Sopomenkam

- Oxford DZS – prevodne ustreznice angleških besed in besednih zvez

dom; bivališče, domovanje; stanovanje; hiša

Od prevodnih ustreznic k Sopomenkam

- Oxford DZS – prevodne ustreznice angleških besed in besednih zvez

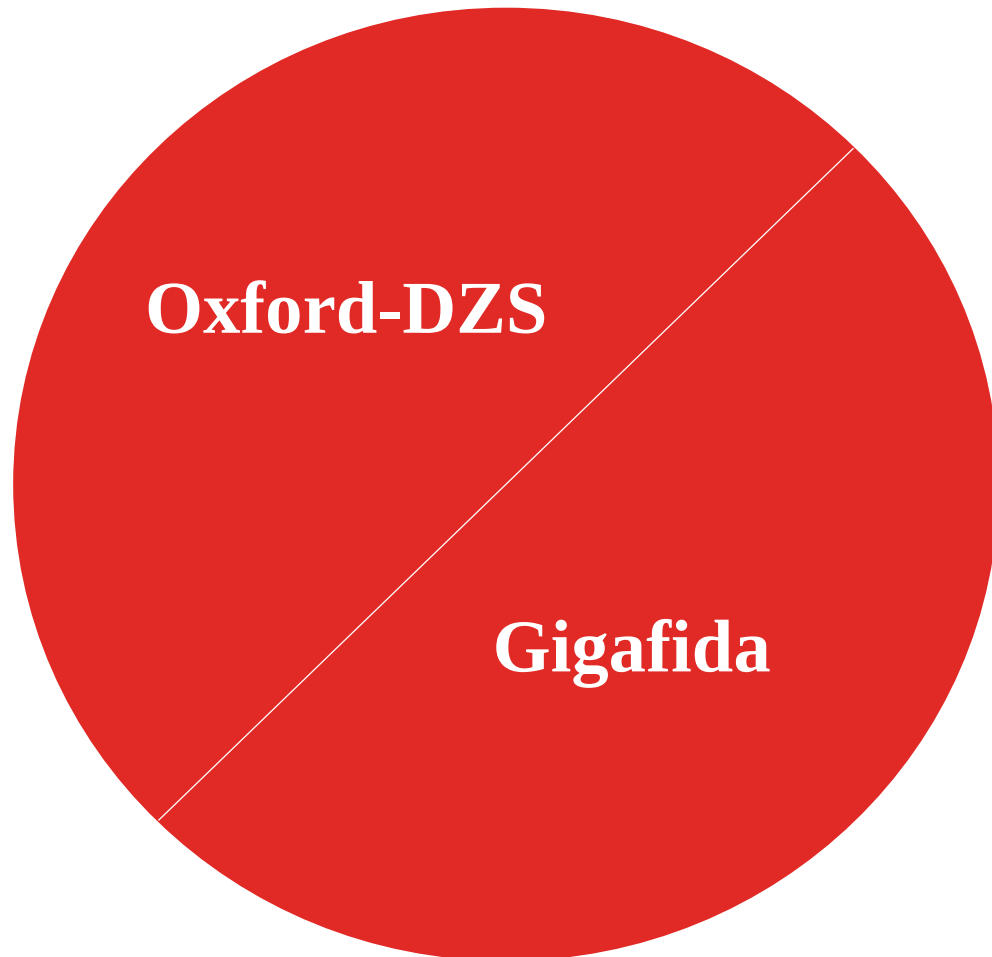
dom; bivališče, domovanje; stanovanje; hiša

```
graph TD; dom[dom]; bivalis[bivališče]; domovanje[domovanje]; stanovanje[stanovanje]; hiša[hiša]; dom --- B1[ ]; bivalis --- B1; B1 --- B2[ ]; domovanje --- B2; stanovanje --- B2; hiša --- B2; B2 --- B3[ ]; B3 --- B4[ ]; dom --- B4; bivalis --- B4
```

Sopojavitveni grafi



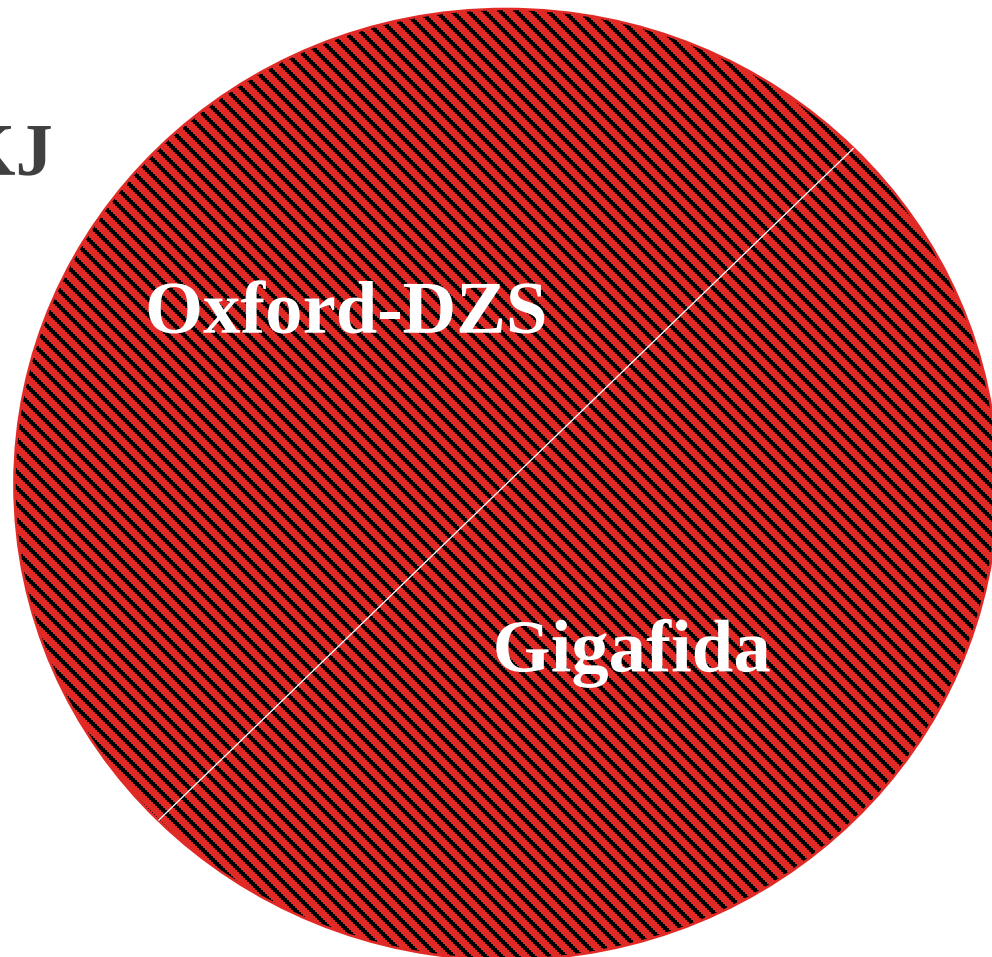
Zasnovanost podatkovne baze



- Temeljna podstat Sopomenk sta Veliki angleško-slovenski slovar Oxford DZS® in referenčni pisni korpus Gigafida.
- Jezikovni podatki po letu 1990.

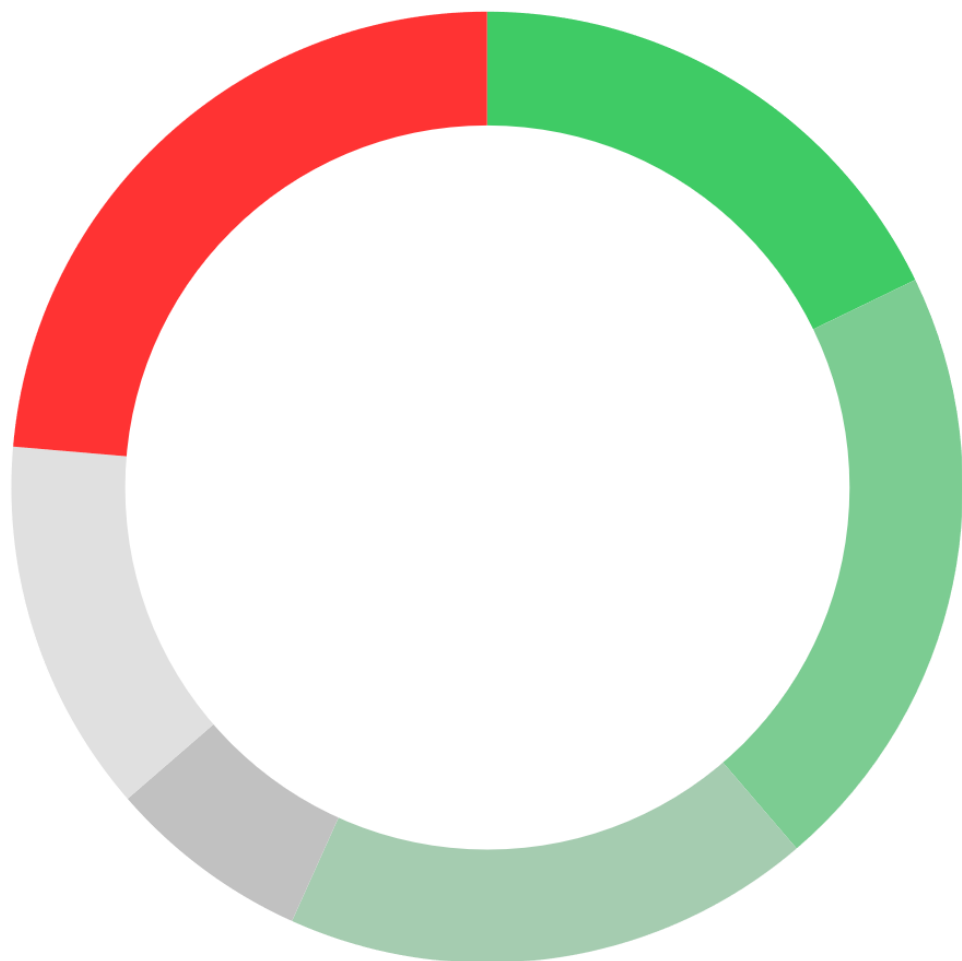
Zasnovanost podatkovne baze

SSKJ



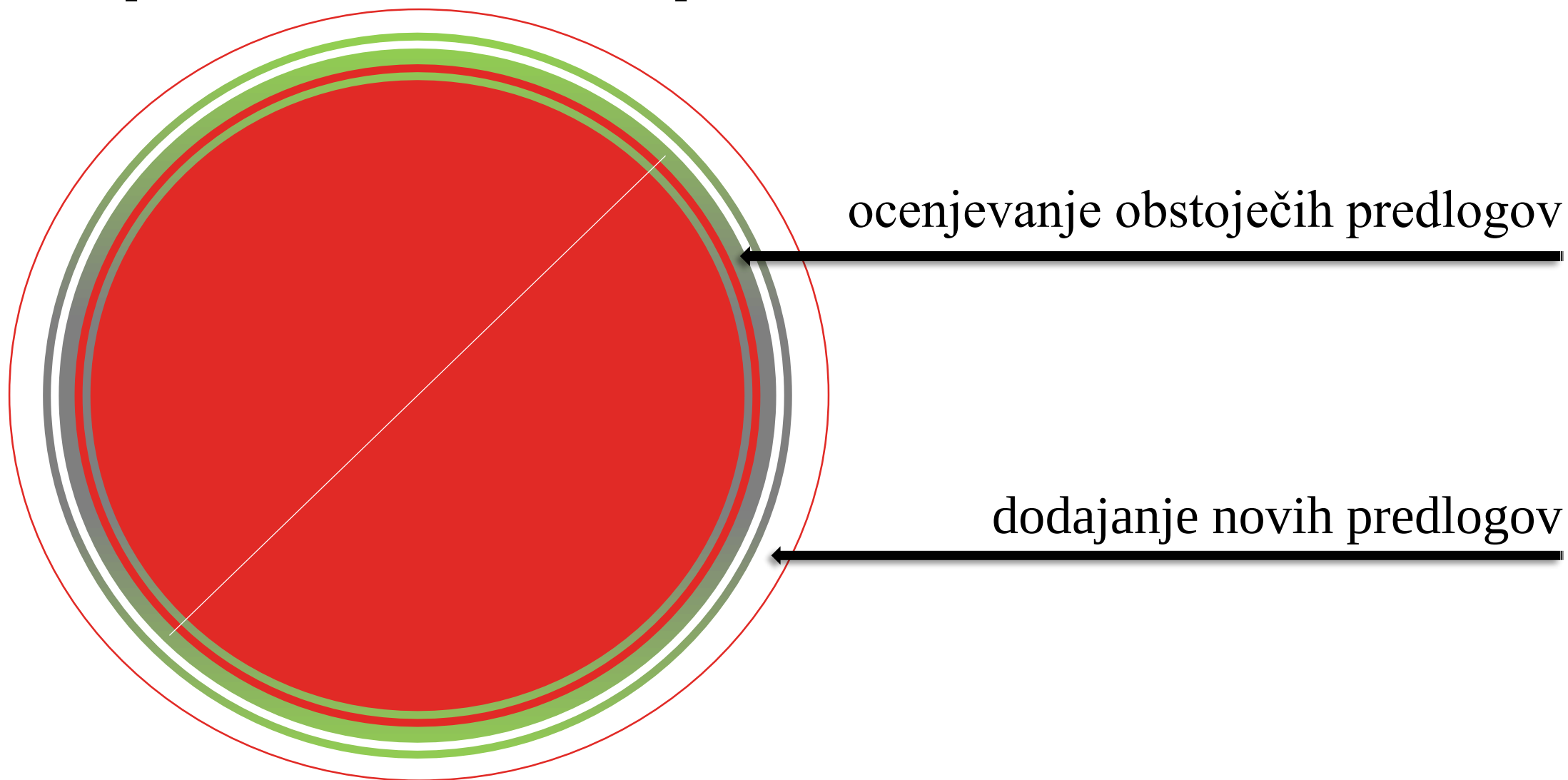
- Temeljna podstat Sopomenk sta Veliki angleško-slovenski slovar Oxford DZS® in referenčni pisni korpus Gigafida.
- Jezikovni podatki po letu 1990.
- Dodatno preverjanje podatkov v SSKJ.
- Izrabljivost virov za razvoj novih izdelkov.

Jezikoslovna evalvacija podatkov



	Ustrezna	18 %
	Ustrezna / Sprejemljiva	21 %
	Sprejemljiva	18 %
	Ustrezna / Neustrezna	7 %
	Sprejemljiva / Neustrezna	13 %
	Neustrezna	24 %

Uporabniški doprinos



Dodaj sopomenko za "ideja"



Sopomenka



Uporabniško ime

Dodaj

preblisk
Špela A. H.



nov pogled
Ben



domislek
Andrea



prešinek
RD



utrinek
BM



izvirna misel
Ben



fleš
Kolokator



Jedrne sopomenke: 6. Bližnje sopomenke: 7. Uporabniške sopomenke: 7.

Relevantnost ▼

Pogostost



< pojem

pogled >

zamisel

misel

domislica

predstava
filozofija

ideja | koncept



nov



za projekt



iti za



imeti



dober



za film



temeljiti na



poiskati



osnoven



do realizacije



pri do



dobiti



zanimiv



na področju



strinjati z



predstaviti



svež



o razvoju



ukvarjati z



uresničiti



odkritje

pogruntavščina

zasnova

pojem

koncept

pogled

ideja

polna



za posel



simpatizirati z



zbirati



najpodjetniški



do uspeha



navdušiti za



izmenjati



bolna



o ustanovitvi



prekipevati od



zapisovati



imenitna



za darilo



boriti za



navreči



krasna



o povezovanju



obupati nad



krasti



Spremljanje uporabniških dejavnosti

- Dodelana metodologija
- Analiza dnevnikov iskanja
 - izboljšave vmesnika
 - učinkovita motivacija uporabnikov
 - pregled dodanih sopomenk

Nava: zelen | neuk, zelen | nevešč, ličen | estetski

BM: pojmovanje | koncept, smotrno | namensko, ideja | utrinek

hobit: preživeli | preživelec, spin | zvirt, radovednež | firbec

ciciban: pravočasen | točen, pasji | zločest, pasji | zlohoten

JJ: smotrno | načrtno, diskurz | razprava

Uporabniške dejavnosti v številkah

- **Podatki za 21. maj 2018 (dva meseca po izidu)**
- Število uporabniških sinonimov – 694
- Pozitivne ocene uporabnikov – 2.320
- Negativne ocene uporabnikov – 848
- Število uporabniških imen – 102
- Število iskanih iztočnic, za katere še ni podatkov – 1.356
- Število različnih iskanih iztočnic – 33.821 (32,2 % celotnega slovarja)
- Število IP-naslovov – 9.821

Misli za domov

- Digitalno slovaropisje je nov model slovaropisja, ki odgovarja na nove potrebe in navade.
- Uporabniki na slovar ne vplivajo samo z neposrednim vključevanjem, ampak tudi s pričakovanji in neposrednimi odzivi na ponudbo na trgu.
- Slovar sopomenk sodobne slovenščine (odzivni slovar) obeta uspešen model za podporo slovenščine v digitalni dobi.
- Sodelujte v anketi, ki bo pomagala pri oblikovanju slovarja kolokacij:
<https://www.1ka.si/a/159349>

Reference

- Arhar Holdt, Š. (2015): Uporabniške raziskave za potrebe slovenskega slovaropisja: prvi koraki. V V. Gorjanc idr. (ur.): *Slovar sodobne slovenščine: problemi in rešitve*: 136–148. Ljubljana: Znanstvena založba Filozofske fakultete.
- Arhar Holdt, Š., Kosem, I. in Gantar, P. (2016): Dictionary user typology: the Slovenian case. V T. Margarlitadze in G. Meladze (ur.): *Lexicography and linguistic diversity: proceedings of the XVII EURALEX International Congress*: 179–187. Tbilisi: Ivane Javakhishvili Tbilisi State University. <http://euralex2016.tsu.ge/publication2016.pdf>.
- Hatherall, Glyn, 1984: Studying dictionary use: some findings and proposals. Hartmann, Reinhard R. K. (ur.): *Zbornik konference LEXeter'83*. Tübingen: Max Niemeyer Verlag. 183–189.
- Kosem, I. (2017). The European survey of dictionary use. V: *Electronic lexicography in the 21st century: lexicography from scratch*: 53. Leiden: Dutch Language Institut; Brno: Lexical Computing; Ljubljana: Trojina Institute for Applied Slovene Studies.
- Krek, Simon, Laskowaki, Cyprian, Robnik-Šikonja, Marko. From translation equivalents to synonyms: creation of a Slovene thesaurus using word co-occurrence network analysis. V: Kosem, Iztok (ur.) et al., *Proceedings of eLex 2017: Lexicography from Scratch*, 19-21 September 2017, Leiden, Netherlands. <https://elex.link/elex2017/wp-content/uploads/2017/09/paper05.pdf>