



## Jezične tehnologije i mrežni izvori Instituta za njemački jezik<sup>1</sup>

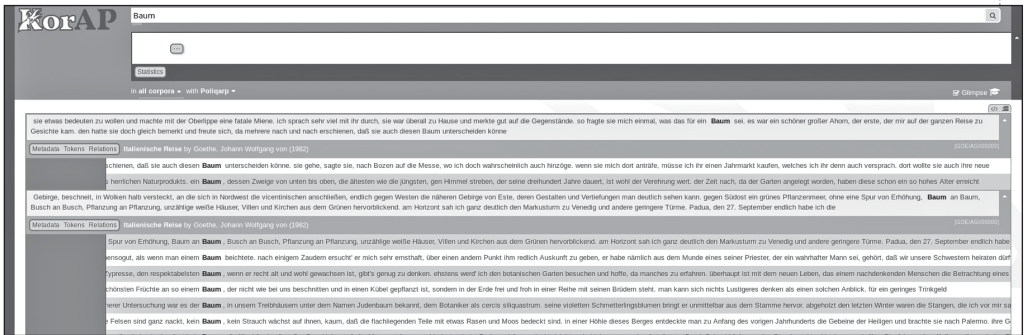
Institut za njemački jezik (*Institut für Deutsche Sprache – IDS*) osnovan je 1964. godine. Institut nije povezan sa sveučilištem i središnja je ustanova za proučavanje njemačkoga jezik danas i u prošlosti. Projekti su Instituta uglavnom dugoročni i zahtijevaju rad velikih istraživačkih skupina. Česta je suradnja s istraživačima iz drugih ustanova i sa sveučilišta. Na stranicama Instituta na poveznici mrežne ponude (*Onlineangebote*) mogu se vidjeti različite mrežne platforme koje obuhvaćaju rječnike, korpuse, portale i informacijske sustave. Institut je već na početku svojega rada počeo raditi s računalnim korpusima. Korpusi su u Institutu prvo bili na bušenim karticama (prvi korpus 1969.). Godine 1983. nastaje prvi alat za analizu korpusa REFER te poslije COSMAS I. Godine 2004. nastaje COSMAS II. On se i dalje može upotrebljavati na mrežnim stranicama Instituta, ali jedino za njemački jezik. Postoji i inačica programa koja se može preuzeti za Windows. Korpusi su podijeljeni u arhive zbog organiziranja sadržaja korpusa kako pretraživanje ne bi dugo trajalo. U Institutu se radi na dvama korpusima, DeReWo (korpus pisanoga jezika koji se može pretraživati s pomoću programa COSMAS II) i DeReKo (korpus govornoga jezika). Za pravo pristupa korpusima potrebna je registracija. DeReKo je enkodiran na 15 jezika (inačica TEI-ja 5). Baza DGD baza je govornoga korpusa.

U planu je da se COSMAS II zamijeni programom KorAP.

### KorAP

KorAP (<https://korap.ids-mannheim.de/>) skalabilan je, fleksibilan i održiv sustav otvorenoga koda za rad s korpusima. Podržava višeslojnu anotaciju (engl. *Multiple Annotations Layers*), više upitnih jezika (engl. *Multiple Query Languages*) za pretraživanje poput CQL-a i regularnih izraza te podržava označavanje više tipova autorskih prava za određene dokumente (neki tekstovi u korpusu zatvoreni su za određene korisnike). KorAP se može testirati na službenoj stranici s pomoću virtualnoga korpusa. Ima tražilicu na kojoj se mogu pretraživati riječi te se koristiti upitima za pretraživanje. Odnosi među riječima mogu se grafički prikazati s pomoću granaste strukture. U planu je da se program razvije i za pretraživanje drugih jezika. Planira se i rad s usporednim korpusima.

1 Ovaj je rad izrađen na istraživačkome projektu Hrvatski mrežni rječnik – Mrežnik (IP-2016-06-2141, <http://ihjj.hr/mreznik/>), koji u cijelosti financira Hrvatska zaklada za znanost.



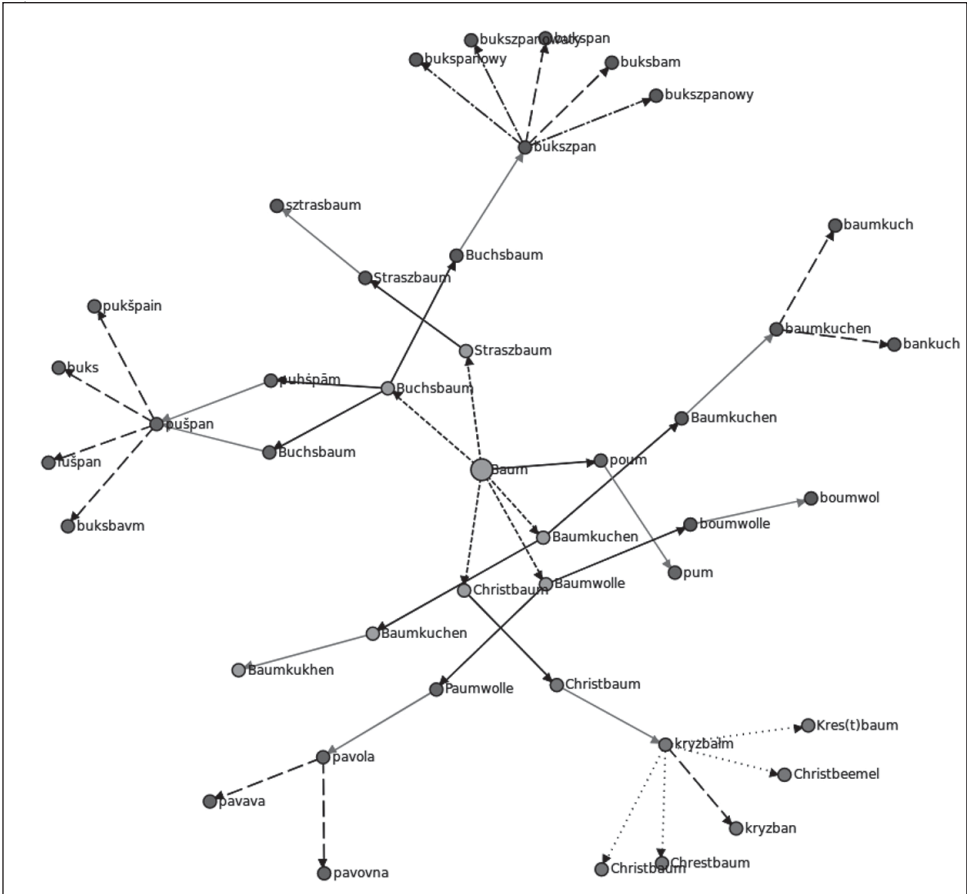
1. slika: Rad u sučelju KorAP (prikaz rezultata pretraživanja za riječ Baum (drvo))

## OWID

Portal OWID prikuplja više rječnika i internetskih bibliografija (uglavnom za istraživačke svrhe) na jednome mjestu (digitalne i e-rječnike), koje je moguće istodobno pretraživati. Za neke zapise moguće je dobiti i izgovor s pomoću audiozapisa. Na portalu se nalaze ovi rječnici i baze: *lexiko* – rječnik općega njemačkog jezika, rječnik paronima, rječnik poslovice, priručnik glagola koji se trebaju koristiti pri komuniciranju na njemačkome, rječnik progresivnih oblika, rječnik stranih riječi, rječnik neologizama te odvojeni rječnici diskursa, od od 1945. do 1955. te 1967./68. Noviji proizvodi koji se razvijaju na stranici mogu se vidjeti na poveznici OWIDplus (<https://www.owid.de/plus/index.html>), na kojoj se nalazi i alat za leksičku analizu (Lexical Explorer) te nogometni rječnik (Kicktionary).

## Portal posuđenica (Das Lehnwortportal Deutsch des IDS)

Od mrežnih izvora vrijedi istaknuti i portal posuđenica (<http://lwp.ids-mannheim.de/>), na kojemu se prikazuje koje su riječi iz njemačkoga ušle u poljski, hebrejski, tok pisin i slovenski te koja je razlika u njihovu značenju u odnosu na njemački. Za posuđenice se može naći i izgovor te se katkad može na interaktivnome grafu vidjeti odnose među riječima i kad se određena riječ pojavila u kojemu jeziku. Svaka točka koja sadržava riječ u grafu služi kao poveznica do obrade te riječi. Točke grafa mogu se i pomicati te se prijelazom mišem mogu vidjeti osnovni oblici riječi. Grafovi su izrađeni s pomoću D3 javascript knjižnice (<https://d3js.org/>). Podatci o rječnicima spremljeni su u bazi Oracle te se za unos i obradu natuknica upotrebljava jezik XML. Stranica rječnika napravljena je u sučeljima Play 2.7. i Spark. Trenutačno se razvija novo sučelje za taj rječnik, koje će pružati više mogućnosti pri usporednome pretraživanju više jezika. Korisnicima će se omogućiti registracija, nakon čega će moći komentirati natuknice u rječniku. U planu je da se dodaju i posuđenice iz rumunjskoga i engleskoga.



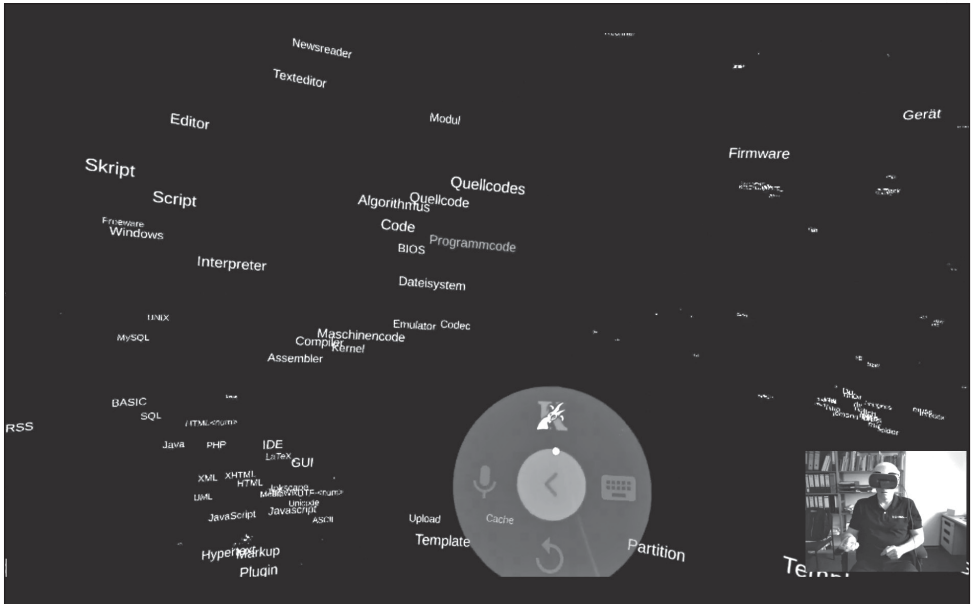
2. slika: Grafički prikaz na *Portalu posuđenica (Lehnwortportal)*

**Grammis**

Grammis je baza gramatičkih podataka (<https://grammis.ids-mannheim.de/>). Na početnoj stranici osim tražilice nalaze se poveznice na izvore koji uključuju bazu jezičnih savjeta, bibliografije znanstvenih radova o gramatičkim pitanjima i različite rječnike (valencijski rječnik, rječnik konektora, rječnik prijedloga, rječnik afiksa itd.). Svaki rječnik i baza imaju svoju zasebnu tražilicu s različitim poljima za pretraživanje i filtrima. Moguća je registracija na stranici koja korisniku omogućuje spremanje poveznice na natuknice na svojem profilu te spremanje stranice s natuknicom u dokument u PDF-u. Neregistrirani korisnici to ne mogu, ali im je omogućeno da generiraju citat za stranicu s koje preuzimaju informacije te da podijele stranicu na Facebooku ili Twitteru. Za izradu stranice Grammisa upotrebljava se programske alat CakePHP.

## Svemir s riječima kao virtualna stvarnost

U Institutu se nalazi i program za virtualnu stvarnost koji se pokreće s pomoću 3D naočala. U programu je prikazan svemir u kojemu su skupine riječi prikazane kao zvijezde. Kroz to zvijezde može se navigirati s pomoću posebnoga pokazivača, kojim se još može označiti riječ i za nju dobiti definicija te pokrenuti opcija njezina pretraživanja u programu KorAP. Unutar te virtualne stvarnosti nalazi se igra u kojoj igrači dobivaju određenu riječ koju moraju dovući do odgovarajuće skupine riječi na nebu. Virtualna stvarnost još se razvija te je bila predstavljena prošle godine u tehničkom muzeju (TECHNOSEUM: Wortspielereien). Za izradu se upotrebljava program Unity. Kratka videodemonstracija dostupna je na poveznici: <https://www.youtube.com/watch?v=YjaA2VGN81c>.



3. slika: Virtualna stvarnost za pregled riječi u skupinama