

Mario Essert, suradnik Odjela za matematiku Sveučilišta u Osijeku, messert@inet.hr

Lucija Tatarević, studentica Odjela za matematiku Sveučilišta u Osijeku, lucija.tatarevic@gmail.com

Marko Orešković, Nacionalna i sveučilišna knjižnica u Zagrebu, moreskovic@nsk.hr

Hrvatski mrežni e-rječnik višerječnica

Kao što su molekule tvorene od atoma, tako su višerječnice tvorene od riječi. One obično sadrže različit tip informacije: terminološki (npr. ratni brod), sociološki (npr. labava carinska unija), metaforički (npr. sunce moje drago), idiomatski/frazeološki (npr. brz na jeziku), poslovični (npr. griješeći se uči) i sl. Za svaki je jezik rječnik višerječnica (MWE – eng. multiword expression) izuzetno važan jer daje paradigmatiku i semantičku komponentu koju rječnik pojedinačnih riječi nema. Mrežni sintaktičko-semantički okvir (SSF), uz mrežni leksikon i rječnik subčestica (morfemi i slogovi), sadrži i MWE rječnik koji je s njima čvrsto (u oba smjera) povezan.

Trenutačno u SSF-u postoji 170.000 višerječnica koje su programom skupljene iz vanjskih izvora (poštujući autorsko pravo) ili mrežnih dokumenata (uz statističku provjeru tipa, tj. veze riječi u višerječnici). Uporaba i korisničko sučelje MWE rječnika slično je načinu rada ostalih rječnika. Poredak višerječnica temelji se na abecednom poretku riječi koje su povezane i s mjestom tj. pozicijom odabrane riječi u višerječnici. Mogu se odabrati apsolutna mjesta utora odabrane riječi (od 1 do 11) ili relativna mjesta (npr. korisnik može odabrati da tražena riječ bude na kraju višerječnice ili na bilo kojoj poziciji). Uz to postoji filtracija višerječnica, pa se mogu prikazivati samo izabrane kategorije, npr. kolokacije, idiomi/frazemi, čvrste sveze riječi, temini i sl. Uz svaku višerječnicu prikazana je i potpuna struktura svake pojedinačne riječi u njoj – njena morfemska, slogovna, gramatička i semantička obilježja te poveznice s drugim izvorima. Posebna pozornost posvećena je statističkim alatima za obradu višerječnica (npr. mjerama asocijativne povezanosti). Ti alati (Student-ov t-test, Pearson-ov hi-kvadrat test, Log-likelihood i Fisher egzaktni test) služe za otkrivanje i provjeru kolokacija i koligacija. Dodatni napori uloženi su u aglomerativno grupiranje kako bi se provela kolostrukcijska (Stefanowitsch, Gries) istraživanja i u hrvatskome jeziku.

Ključne riječi: mrežni okvir, rječnik višerječnica, provjera kolokacija i koligacija, kolostrukcijska istraživanja