

Ako koristite ovaj program, molim vas, citirajte: Filipović Đurđević, D., Đurđević, Đ. i Kostić, A. (2009). Vector based semantic analysis reveals absence of competition among related senses. *Psihologija*, 42(1), 95-106. [doi:10.2298/PSI0901095F](https://doi.org/10.2298/PSI0901095F)

FORMIRANJE VEKTORA DRUGOG REDA ZA POJEDINAČNA POJAVLJIVANJA REČI (INFLEKTIVNIH OBLIKA)

1. KORAK – formiranje vektora prvog reda za pojedinačna pojavljivanja oblika reči		
Program	vektori.exe koristiti sheet " formiranje vektora ", podopciju " vektori prvog reda (pojedinačno pojavljivanje) "	
Ulazni fajlovi	Recnik	Fajl u .txt formatu koji sadrži jednu kolonu u kojoj se nalazi kontinuirani tekst, u našem slučaju radi se o Ebart medijskoj bazi, odnosno fajlu koji se zove ebart.txt.
	Spisak ciljnih reči	Fajl u .txt formatu koji sadrži jednu kolonu u kojoj se nalaze ciljne reči, tačnije svi inflektivni oblici svake ciljne leme, unete jedna ispod druge.
	Kontekst	Fajl u .txt formatu koji sadrži jednu kolonu u kojoj se nalaze, jedna ispod druge, kontekst reči, u našem slučaju radi se o hiljadu najfrekventnijih reči.
Izlazni fajlovi	Izlaz	Fajl u .txt formatu koji sadrži matricu dimenzija $m \times n$, gde m predstavlja broj pojedinačnih pojavljivanja ciljnih reči, a n predstavlja broj kontekst reči. Pored toga, u prvoj koloni ovog fajla nalazi se ciljna reč na koju se odnose vrednosti vektora u datom redu, a u drugoj koloni redni broj pojavljivanja nadjenog oblika. Svaki red predstavlja vektor prvog reda za pojedinačno pojavljivanje odgovarajuće ciljne reči (elementi tog vektora su nule i jedinice, gde nula znači da se data ciljna reč i kontekst reč nisu pojavile unutar prozora zajedno, a jedinica znači da su se unutar prozora zajedno našle data ciljna reč i kontekst reč iz odgovarajuće kolone). Primetiti da ciljne reči nisu leme, već inflektivni oblici, odnosno oni tekstualni nizovi koji su uneti u

		spisak ciljnih reči.
Ostale specifikacije	Velicina prozora	Ovde je moguće definisati broj mesta levo i desno od ciljne reči na kojima će biti tražena kontekst reč. Standardna i istovremeno default vrednost je +/- 3.
	Kolona	Za svaki ulazni fajl moguće je definisati kolonu u kojoj se nalaze potrebne vrednosti. Default vrednost je 1, što znači da se potrebne vrednosti nalaze u prvoj koloni (pošto je prethodno preporučeno da se svi fajlovi sastoje iz jedne kolone). Dakle, ova vrednost ostaje 1.
2. KORAK – formiranje standardnih vektora prvog reda za odabrane kontekst reči u kontekstu istih tih kontekst reči		
Program	vektori.exe koristiti sheet " formiranje vektora ", podopciju " vektori prvog reda (standardni) "	
Ulazni fajlovi	Recnik	Fajl u .txt formatu koji sadrži jednu kolonu u kojoj se nalazi kontinuirani tekst, u našem slučaju radi se o Ebart medijskoj bazi, odnosno fajlu koji se zove ebart.txt.
	Spisak ciljnih reci	Fajl u .txt formatu koji sadrži jednu kolonu u kojoj se nalaze, jedna ispod druge, kontekst reči, u našem slučaju radi se o hiljadu najfrekventnijih reči.
	Kontekst	Fajl u .txt formatu koji sadrži jednu kolonu u kojoj se nalaze, jedna ispod druge, kontekst reči, u našem slučaju radi se o hiljadu najfrekventnijih reči. U pitanju je isti fajl kao u prethodnom polju!
Izlazni fajlovi	Izlaz	Fajl u .txt formatu koji sadrži matricu dimenzija $n \times n$, gde n predstavlja broj kontekst reči. Pored toga, u prvoj koloni ovog fajla nalazi se reč na koju se odnose vrednosti vektora u datom redu. Svaki red predstavlja standardni vektor prvog reda za odgovarajuću ciljnu reč (elementi tog vektora su frekvencije zajedničkog pojavljivanja date ciljne reči i kontekst reči iz odgovarajuće kolone). Primetiti da ciljne reči nisu leme, već

		inflektivni oblici, odnosno oni tekstualni nizovi koji su uneti u spisak ciljnih reči. Ne zaboraviti da su iste reči ovde bile i ciljne reči i kontekst reči.
Ostale specifikacije	Velicina prozora	Isto kao u prethodnom koraku
	Kolona	Isto kao u prethodnom koraku
3. KORAK – formiranje vektora drugog reda za pojedinačna pojavljivanja oblika reči		
Program	vektori.exe koristiti sheet " formiranje vektora ", podopciju " vektori drugog reda " Ukoliko su svi parametri za prethodna dva koraka definisani, moguće je kliknuti na opciju "paketska obrada" čime će ulazni fajlovi biti automatski definisani.	
Ulazni fajlovi	Kontekst	Fajl u .txt formatu koji sadrži jednu kolonu u kojoj se nalaze, jedna ispod druge, kontekst reči, u našem slučaju radi se o hiljadu najfrekventnijih reči. U pitanju je isti fajl sa kontekst rečima, kao u prethodnim koracima!
	Matrica 1	Izlazni fajl iz prvog koraka.
	Matrica 2	Izlazni fajl iz drugog koraka.
Izlazni fajlovi	Izlaz	Fajl u .txt formatu koji sadrži matricu dimenzija $m \times n$, gde m predstavlja broj pojedinačnih pojavljivanja ciljnih reči, a n predstavlja broj kontekst reči. Pored toga, u prvoj koloni ovog fajla nalazi se ciljna reč na koju se odnose vrednosti vektora u datom redu, a u drugoj koloni redni broj pojavljivanja nadjenog oblika. Svaki red predstavlja vektor drugog reda za pojedinačno pojavljivanje odgovarajuće ciljne reči (elementi tog vektora su sume odgovarajućih elemenata onih standardnih vektora prvog reda /matrica2/ onih kontekst reči koje su se pojavile u susedstvu date ciljne reči /matrica1/). Primetiti da ciljne reči nisu leme, već inflektivni oblici, odnosno oni tekstualni nizovi koji su uneti u spisak ciljnih reči.
Ostale	Kolona	Isto kao u prethodnom koraku

specifikacije		
---------------	--	--