

Semantičko-sintaktička klasifikacija glagola u hrvatskom jeziku - SEMTACTIC (PUP)

Brač, Ivana

Data management plan / Plan upravljanja istraživačkim podacima

Publication year / Godina izdavanja: **2024**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:305:165541>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-02-09**



Repository / Repozitorij:

repozitorij.ihj.hr

Plan upravljanja istraživačkim podacima - SEMTACTIC

Opće informacije		
	Ime i prezime predlagatelja	Ivana Brač
	Matična organizacija	Institut za hrvatski jezik
	Naziv projekta	Semantičko-sintaktička klasifikacija glagola u hrvatskom jeziku
	Upravitelj podacima	Siniša Runjaić, srunjaic@ihji.hr
1.	Prikupljanje podataka i dokumentacija	
	Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka)	U istraživanju će se prikupljati i obrađivati jezični podatci dostupni u korpusima hrvatskoga jezika te mrežnim stranicama. Riječ je o sintaktičkoj i semantičkoj analizi primjera rečenica. Za svako značenje pojedinoga glagola odredit će se valencijski okviri te će svaki okvir biti potvrđen jednim primjerom ili više njih.
	Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka)	Podatci će se prikupljati pretraživanjem dostupnih korpusa hrvatskoga jezika. Za određivanje značenja i valencijskih okvira koristit će se slučajni uzorak od 2000 rečenica. Da bi se pronašao bolji primjer, može se koristiti regularnim izrazima. Analiza glagola i samih rečenica unosi se u bazu podataka koja će biti javno dostupna do kraja projekta.
	Koju ćete dokumentaciju i metapodatke ustupiti osim podataka? (navedite koje su informacije potrebne korisnicima kako bi mogli čitati i interpretirati podatke u budućnosti te koji će se standardi koristiti pri tumačenju podataka)	Metapodatci o članovima projekta koji obrađuju glagole neće biti dostupni javnosti, nego samo suradnicima koji imaju pristup radnom sučelju. Popis suradnika s biografijama dostupan je na mrežnoj stranici projekta. Javnosti će biti dostupne upute za korištenje bazom.
2.	Pravna i sigurnosna pitanja	
	Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li	Ne postoji poseban ugovor o povjerljivosti jer istraživanje ne sadržava osjetljive podatke ili podatke koji se tiču kategorija osobnih podataka, odnosno riječ je isključivo o analizi tekstualnih podataka.

	osobe čiji se podaci pohranjuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka)?	
	Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka?	Pristup podacima reguliran je putem aplikacijskoga autorizacijskog sustava same aplikacije koji regulira pristup i razinu prava korisnika. Potencijalni rizik može predstavljati neovlašten pristup treće strane u institutski server, ali ne očekuje se veći rizik od takve pojave jer nije riječ o osjetljivim podacima.
	Kako ćete upravljati zaštitom autorskih prava i intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka?	Primijenit će se osnovna licencija Creative Commons za konzultiranje podataka na javnoj dostupnoj verziji. Podatci će biti dostupni akademskoj zajednici i javnosti za slobodno korištenje, uz navođenje izvora.
3.	Pohrana i čuvanje podataka	
	Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (<i>backup</i>) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim raspolazete? Kojim se procedurama koristite za sigurnosnu kopiju (<i>backup</i>)?	Podatci su pohranjeni u obliku datoteka naziva (naziv_baze + datum_kopije + .tar.zip). Kapaciteti su osigurani od strane Institut za hrvatski jezik te su dostatne za spremanje pričuvnih kopija za period od 30 dana. Procedura automatskoga kreiranja pričuvnih podataka baze izvršava se svaki dan u 22 sata.
	Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati?	Plan čuvanja podataka definiran je protokolom o kreiranju pričuvnih verzija baze podataka. Podatci su u SQL dump formatu podataka.
4.	Dijeljenje i ponovna uporaba podataka	
	Kako i gdje će se podaci dijeliti? Na kojem repozitoriju planirate dijeliti	Dosadašnji rezultati istraživanja koji su objavljeni u znanstvenim časopisima i knjigama dostupni su na mrežnoj stranici projekta te u bazama u kojima su indeksirani. Svi podatci u bazi postat će dostupni javnosti u posljednjoj godini projekta, odnosno nakon izrade korisničkoga sučelja.

	podatke? Kako će potencijalni korisnici doznati za podatke?	Korisnici za to doznaju tijekom diseminacijskih aktivnosti projekta.
	Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja.	Nema ograničenja jer je riječ o javno dostupnim rezultatima istraživanja projektom koji financira Hrvatska zaklada za znanost, a analiziraju se rečenice koje su javno dostupne u pretraživim korpusima hrvatskoga jezika i znanstvenim publikacijama te ne podliježu autorskim pravima ili etičkim ograničenjima.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima <i>FAIR-a</i> .	Potvrđujem.
	Potvrdite da ćete se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (ako ne, objasnite zašto ne možete dijeliti podatke na digitalnom repozitoriju koji nije komercijalan).	Potvrđujem.