

NAVODILA ZA OCENJEVANJE TERMINOLOŠKIH KANDIDATOV

Špela Arhar Holdt, arhar.spela@gmail.com

UVOD

Dobrodošli kot sodelujoči pri [projektu Slovenska znanstvena besedila: viri in opis](#). Pri projektu razvijamo **sistem za avtomatsko pridobivanje terminologije iz znanstvenih besedil**. Vaša naloga bo, da pregledate del rezultatov z vašega področja in zabeležite, ali gre pri posameznem primeru za termin ali ne. Vaše odločitve bomo uporabili za izboljšavo sistema. Na koncu razvoja bomo s sistemom pridobili podatke iz večje količine znanstvenih besedil in jih strokovni javnosti dali na voljo za pripravo slovarjev in podobnih jezikovnih virov.

PODATKI

Podatki so na voljo v Excelovi datoteki, ki vsebuje vaše ime in kodo doktorskega dela, iz katerega so podatki pridobljeni. Datoteko shranite na svoj računalnik, vendar je (razen morebitnega pripisa verzije) ne preimenujte.

Tabela ima pet stolpcev, kot prikazuje spodnja slika.

TERMIN?	BESEDA ALI ZVEZA	BESEDA ALI ZVEZA V POGOSTI OBLIKI	PRIMERI (username: kusus, password: flapka-ma-tuha)	VZOREC
	analiza	analize	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	analiza rezultat	analiza rezultatov	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.* Nc.-g.*
	analiziranje	analiziranja	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	antropometričen meritev	antropometričnih meritev	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	A.* Nc.*
	antropometričen podatek	antropometričnih podatkov	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	A.* Nc.*
	antropometrija	antropometrije	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	atribut	atribute	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	avtor	avtor	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	b	b	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	baza	baze	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	baza znanje	baze znanja	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.* Nc.-g.*
	beseda	besede	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	biomehanika	biomehanika	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	bistvo	bistvu	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	blok	blok	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	blok z pravilo	blok s pravili	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.* S.* Nc.*
	bodoč uporabnik	bodoči uporabnik	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	A.* Nc.*
	cad	cad	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	cad orodje	cad orodja	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.* Nc.-g.*
	celota	celota	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	celoten sistem	celoten sistem	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	A.* Nc.*
	cikel	cikel	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*
	cilj	cilj	http://nl.is.si/noppe/kas.ca/fnst/foranamesa02_@kuserpsectorcalow@qai%58iemna%32%2a	Nc.*

[A] **TERMIN?** V ta stolpec vpišete oznako , ali je beseda ali zveza termin ali ne (gl.razdelek Izbira oznake spodaj).

[B] **BESEDA ALI ZVEZA:** To so podatki, ki jih ocenjujete.

[C] **BESEDA ALI ZVEZA V NAJBOLJ POGOSTI OBLIKI:** Ker so v stolpcu B vse besede v osnovni obliki (imenovalnik ednine), so zveze včasih težko berljive (npr. *baza znanje*). V stolpcu C so zato podatki na voljo tudi v obliki, ki je v besedilu najbolj pogosta (*baze znanja*).

[D] **PRIMERI:** Tu je povezava na korpus, kjer si lahko pogledate besedo ali zvezo v besedilu.

[E] **VZOREC:** Koda skladijskega vzorca, s katerim smo luščili besede in zveze. Ni ključno za vaše delo.

UPORABA PRIMEROV

Povezava do primerov vam bo pomagala pri odločanju, ali je določena beseda oz. zveza termin ali ne. Predvideva se, da si primere vsaj na hitro ogledate, razen takrat, ko se lahko na osnovi svojega poznavanja stroke zanesljivo in brez težav odločite na prvi pogled.

Ob prvem kliku na povezavo do primerov vas bo sistem vprašal za uporabniške podatke:

uporabniško ime: **kusus**

geslo: **flapka-ma-tuha**

Korpusni primeri so nanizani kot seznam kratkih izsekov iz besedila (na sliki spodaj primeri za besedo *atribut*). S klikom na rdečo besedo se vam na spodnjem delu ekrana odpre še širši odstavek besedila.

user: katus corpus: KAS-PhD v0.2	
<p>Concordance Word List Corpus Info ?</p> <hr/> <p>Save</p> <p>Subcorpus name:</p>	<p>Query atribut, Nc.*, kas-847139 8 (0.2 per million)</p> <p>kas-847139,UM FS oblikovalskih rešitev na podlagi inženirskih atributov je v tem trenutku relativno vsakdanja praksa</p> <p>kas-847139,UM FS optimiranja oblike izdelkov tudi subjektivne attribute, med katerimi prednjačijo estetika, čustvena</p> <p>kas-847139,UM FS samo inženirske, ampak tudi subjektivne attribute oblike [10]. ☞ Vedno bolj prihaja do izraza</p> <p>kas-847139,UM FS izraza tudi korelacija med subjektivnimi atributi oblike in njenimi ergonomskimi značilnostmi</p> <p>kas-847139,UM FS uporabnikov pa sta identificirali 25 oblikovnih atributov, med katerimi je izstopala karakteristika</p> <p>kas-847139,UM FS, skupine, zbirke posameznih objektov, ☞ » attribute « (ang.: attributes) - lastnosti, značilnosti</p> <p>kas-847139,UM FS povezani z relacijami in imajo določene attribute. V nasprotnem primeru lahko govorimo o</p> <p>kas-847139,UM FS se uporablja za definiranje (določevanje atributov) posameznemu konceptu ontologije ter ☞ krmilno</p>

IZBIRA OZNAKE

Izbirate lahko med spodaj naštetimi oznakami. Če se vam bo ob delu zdelo, da bi nujno potrebovali kakšno dodatno kategorijo, sporočite.

- **t = termin**. Kot termin označite besede in zveze, ki predstavljajo strokovni pojem z vašega področja (tj. področja **računalništva, kemije ali političnih ved**).
- **x = izvenpodročni termin**: beseda ali zveza, ki se vam zdi termin, ampak spada po vaši oceni na drugo strokovno področje.
- **z = znanstveno pisanje**: besedišče, ki se ukvarja s pridobivanjem, analizo in predstavitvijo podatkov (*slika, tabela, testiranje sistema*) ali je del procesa doktorskega pisanja (*mentor, študij literature*) ali strukture doktorskega dela (*uvod, zaključek*).
- **n = nerelevantno**: vse ostalo, torej splošno besedišče, ki ni del terminologije (*v bistvu; celota, dimenzija ročaja*), zveze, ki so del daljšega opisa (*blok s pravili, ki so namenjena hobi orodjem*); zveze, ki zahtevajo razpravljalno definicijo oz. so subjektivne narave (*cilj estetike v oblikovanju*); tuje besede (*collection*) in primeri, kjer je rezultat bodisi predolg ali prekratek (*logični blok = termin; blok = prekratek, posamezen logični blok = predolg*).
- **? = ne vem**: primeri, glede katerih se ne morete odločiti in jih bomo naknadno pregledali z mentorjem označevanja.

MEJNI PRIMERI

Pri določanju, kaj je terminologija in kaj ne, so nekateri primeri jasni, pri drugih je težje. Predvidimo lahko naslednje dileme:

- Termin ali splošno besedišče? Termin je za področje pomemben, relevanten, je del znanja stroke, povezan z drugimi koncepti. Karkoli ustreza temu opisu, označimo kot termin. Če se sicer splošna beseda pojavlja v terminološki rabi, jo označimo kot termin (npr. *posameznik* za opis elementa ontologije); pregled korpusnih primerov je pri tej odločitvi ključen. Če v primerih beseda nastopa v več pomenih in je terminološka v vsaj enem, jo označimo kot termin.
- Izvenpodročni termin, znanstveno ali nerelevantno? Za marsikatero sicer dokaj splošno besedo si je mogoče predstavljati, da se uporablja kot termin na nekem drugem področju, npr. beseda *srce* za večino ni termin, na področju medicine pa morda je. Tukaj ne gremo predaleč v ugibanjih: kot izvenpodročni termin označujemo primere, ki so izrazito terminološki, vendar zaradi interdisciplinarnosti obravnavanega doktorata ne spadajo na naše področje. Pri odločitvi se zanašamo na primere v korpusu.
- Nekatero besede v podatkih nastopajo samostojno in obenem kot del zveze (npr. *uporabniški vmesnik* ter *vmesnik*). Pri odločanju, ali je krajši segment termin sam na sebi, se opremo na korpusne primere, svoje poznavanje strokovnega področja in morebitne razpoložljive referenčne vire s področja.

- Če ne vemo, ali bi rezultat uvrstili pod *znanstveno pisanje* ali ne, ga uvrstimo pod *nerelevantno*. Z nedokončanostjo tovrstnega besedišča se ne ukvarjamo, nepopolne zveze grede pod *nerelevantno*.
- Z jezikovnimi napakami se ne ukvarjamo, npr. vseeno nam je, ali je v besedilu zapis *cad orodje* ali *CAD-orodje* ali *CAD orodje*; podobno npr. *fakulteta za strojnstvo*. Zanima nas samo terminološkost. Na morebitne napake v podatkih lahko opozorite tako, da primer obarvate rdeče in ga bomo dodatno pregledali.