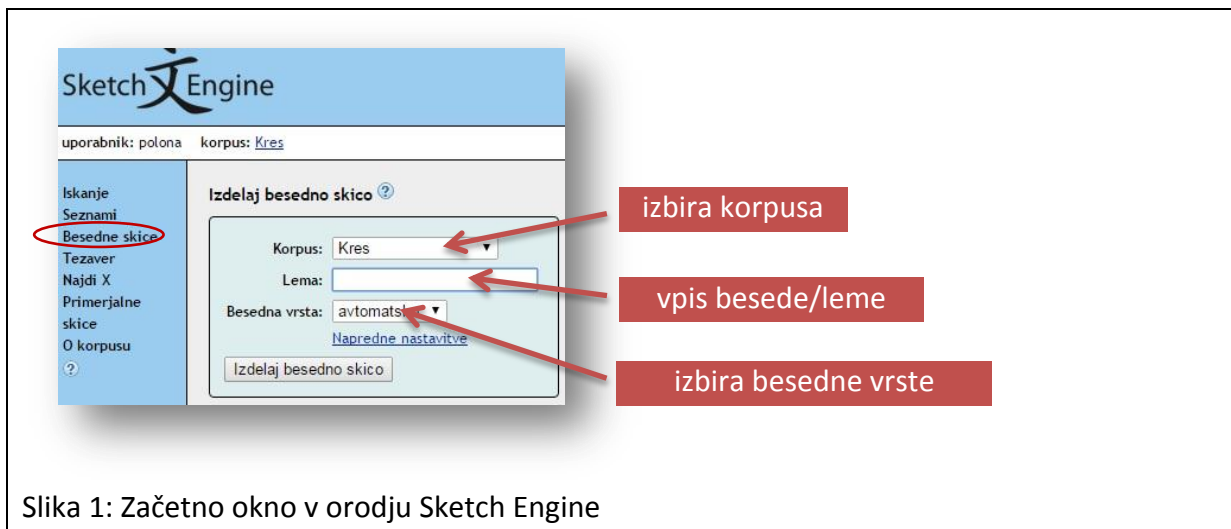


4.2 Vaje

4.2.1 Prepoznavanje različnih tipov večbesednih leksikalnih enot v besedilu

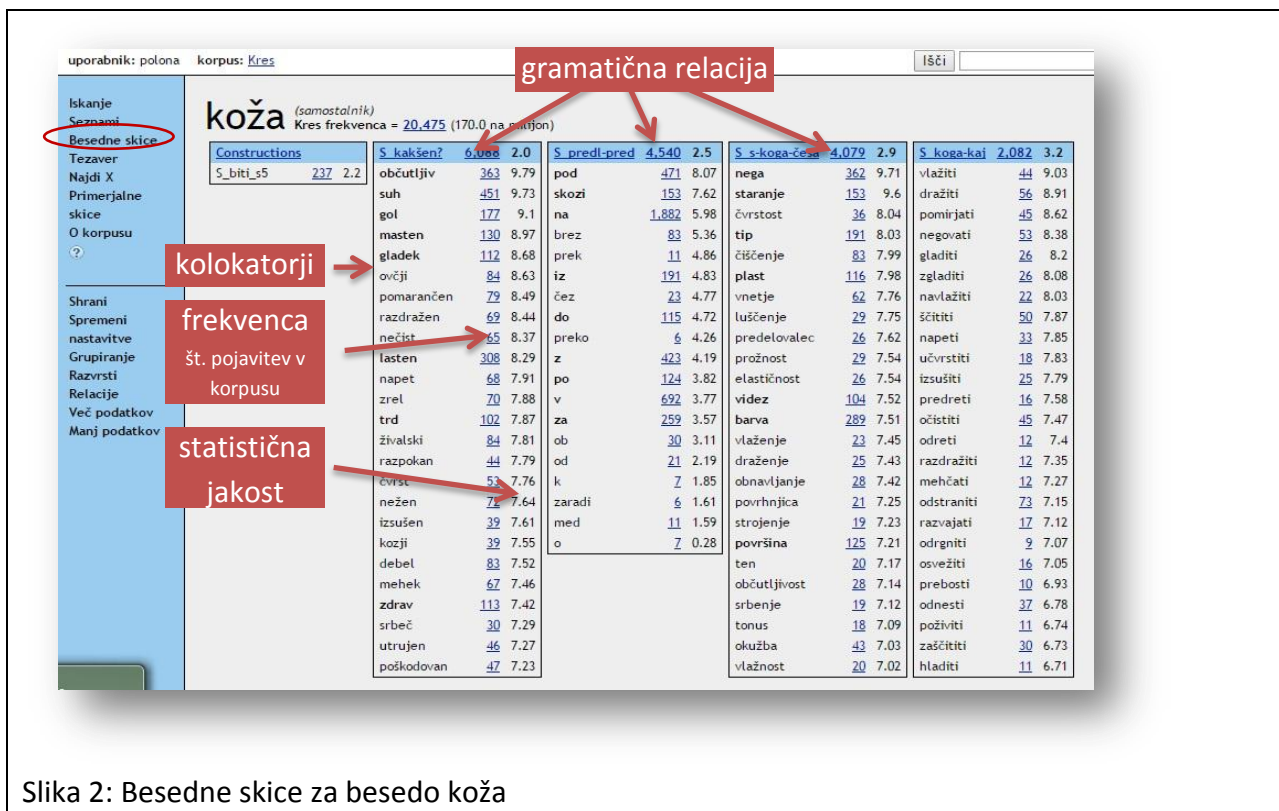
- Priporočljivo je, da si sproti beležite postopek in ugotovitve, saj vam bo to prišlo prav pri izdelavi raziskovalne naloge. Za vnašanje rešitev lahko uporabite kar ta dokument, tako da ga s svojim imenom shranite na računalnik.
- Odpremo orodje **Sketch Engine** ([SkE](#)) oz. [wsketch form](#):



Slika 1: Začetno okno v orodju Sketch Engine

- Analiza bo potekala v orodju **Sketch Engine** (SkE) z uporabo funkcije »Iskanje« (konkordance) in »Besedne skice« (ang. Word Sketches). Uporabljali bomo **korpus KRES** za analizo splošne slovenščine in **korpus JANES** za analizo spletne slovenščine. Znotraj korpusa JANES lahko izbiramo med različnimi podkorpusi: Janes Blog, janis Comment, Janes Forum, Janes Tweet itd.
- V okence »Lema« vtipkamo poljubno besedo v **osnovni obliki** (lemi). To, kar dobimo, so **besedne skice**.

Besedne skice so avtomatsko izdelan povzetek slovničnega in kolokacijskega obnašanja določene besede.



uporabnik: polona korpus: Kres

koža (samostalnik)
Kres frekvenca = 20,475 (170,0 na milijon)

gramatična relacija

Besedne skice

kolokatorji

frekvenca
št. pojavitev v korpusu

statistična jakost

Constructions	S_kakšen?	S_kakšen?	S_predi-pred	S_s-koga-česa	S_koga-kaj
S_biti_s5	237	2.2			
občutljiv	363	9.79	pod	471	8.07
suh	451	9.73	skozi	153	7.62
gol	177	9.1	na	1.882	5.98
masten	130	8.97	brez	83	5.36
gladek	112	8.68	prek	11	4.86
ovčji	84	8.63	iz	191	4.83
pomarančen	79	8.49	čez	23	4.77
razdražen	69	8.44	do	115	4.72
nečist	65	8.37	preko	6	4.26
lasten	308	8.29	z	423	4.19
napet	68	7.91	po	124	3.82
zrel	70	7.88	v	692	3.77
trd	102	7.87	za	259	3.57
živalski	84	7.81	ob	30	3.11
razpokan	44	7.79	od	21	2.19
čvrst	53	7.76	k	7	1.85
nežen	74	7.64	zaradi	6	1.61
izsušen	39	7.61	med	11	1.59
kozji	39	7.55	o	7	0.28
debel	83	7.52			
mehek	67	7.46			
zdrav	113	7.42			
srbeč	30	7.29			
utrujen	46	7.27			
poškodovan	47	7.23			
nega	362	9.71			
staranje	153	9.6			
čvrstost	36	8.04			
tip	191	8.03			
čiščenje	83	7.99			
plast	116	7.98			
vnetje	62	7.76			
luščenje	29	7.75			
predelovalec	26	7.62			
prožnost	29	7.54			
elastičnost	26	7.54			
videz	104	7.52			
barva	289	7.51			
vlaženje	23	7.45			
draženje	25	7.43			
obnavljanje	28	7.42			
povrhnjica	21	7.25			
strojenje	19	7.23			
površina	125	7.21			
ten	20	7.17			
občutljivost	28	7.14			
srbenje	19	7.12			
tonus	18	7.09			
okužba	43	7.03			
vlažnost	20	7.02			
vlažiti	44	9.03			
dražiti	56	8.91			
pomirjati	45	8.62			
negovati	53	8.38			
gladiti	26	8.2			
zgladiti	26	8.08			
navlažiti	22	8.03			
ščititi	50	7.87			
napeti	33	7.85			
učvrstiti	18	7.83			
izsušiti	25	7.79			
predreti	16	7.58			
očistiti	45	7.47			
odreti	12	7.4			
razdražiti	12	7.35			
mehčati	12	7.27			
odstraniti	73	7.15			
razvajati	17	7.12			
odrgniti	9	7.07			
osvežiti	16	7.05			
prebosti	10	6.93			
odnesti	37	6.78			
poživiti	11	6.74			
zaščititi	30	6.73			
hladiti	11	6.71			

LEMA je beseda v osnovni obliki.

KOLOKATORJI so besede, ki se tipično pojavljajo s preučevano besedo in z njo tvorijo **KOLOKACIJE** (kolokator + koža = kolokacija). Da zvezo štejejo za kolokacijo, mora imeti v korpusu najmanj 3 pojavitve. Če kliknemo na kolokator, se nam odprejo konkordance, ki ta kolokator vsebujejo.

FREKVENCA prikazuje, kolikokrat se kolokator pojavi ob preučevani besedi oz. koliko kolokacij tvori z njo v izbranem korpusu.

STATISTIČNA JAKOST določa trdnost in soodvisnost besed, ki vstopajo v kolokacijo.

4.2.2 1. Naloga: KOLOKACIJE

POVEJ MI, S KOM SE DRUŽIŠ, IN POVEM TI ... KAJ POMENIŠ

- Preden se lotimo korpusne analize, preizkusimo svojo **intuicijo**. Vsaka skupina naj zabeleži čim več besed, ki se po njenem mnenju največkrat pojavljajo z besedo *hud*. Skupina, ki bo zabeležila največ kolokacij, ki so izkazane tudi v korpusu, bo dobila nagrado ☺

- V izhodiščnem okencu (gl. Slika 1) izberemo **korpus KRES**, odtipkamo besedo: *hud* in določimo besedno vrsto: pridevnik.
- Kolikokrat se beseda *hud* pojavlja v korpusu KRES in kolikokrat v korpusu JANES? Ali lahko na podlagi tega kaj sklepamo?
- V dobljeni besedni skici analiziramo **kolokatorje** v prvih dveh stolpcih (relacije: **S_kdo-kaj?** in **S_kako-kdaj_p?**). Kakšen pomen besede *hud* izkazujejo kolokatorji v korpusu KRES? Katera je njihova skupna lastnost?
- Ali v drugem stolpcu kateri kolokator izstopa? Zakaj? Preverimo še konkordance in vir.
- Postopek analize kolokatorjev za besedo *hud* ponovimo še na **korpusu JANES** in primerjajmo rezultate.
- Kateri kolokatorji so prekrivni v obeh korpusih in kateri izstopajo?
- **Zapišimo ugotovitve**

4.2.3 2. Naloga: OD KOLOKACIJE DO FRAZEMA

- Izdelamo besedno skico za samostalnik *tič*.
- Najprej zabeležimo **frekvenco leme** v obeh korpusih. Kaj ugotovimo?
- Analiziramo **kolokatorje** v relacij **S_kakšen?** v **korpusu KRES**. Katere pomene nakazujejo kolokatorji v prvem stolpcu? Poskusimo vsak pomen ponazoriti z ustrezno **konkordanco** – zgledom iz korpusa.
- Ali so v obeh korpusih prekrivni kolokatorji in pomeni? Kako je s frekvenco kolokatorjev v enem in drugem korpusu?
- Ali pri katerem od kolokatorjev v korpusu KRES najdemo **frazeme**? Kaj pa v JANES-u? Katere? Kaj je zanačilno za njihovo rabo? Izpišimo primere.
- Ali pridejo do izraza še kateri drugi frazemi v obeh korpusih, če analiziramo **konkordance** (V orodju SkE izberemo možnost »Iskanje«)? Pregledamo do 10 kolokacijskih nizov.
- Ali lahko prepoznamo kako **ustvarjalno rabo** frazema (t. i. prenovitev)? Zapišimo si primere in vir (korpus).
- Kako je prenovitev frazema dosežena (npr. združitev dveh frazemov, dobesedna raba ipd.).
- Ali so frazemi, ki jih je mogoče najti v obeh korpusih, opisani v Slovarju slovenskega knjižnega jezika (SSKJ: preverite na <http://www.fran.si/>)?
- Ali najdemo v SSKJ frazeme, ki jih v korpusu KRES in JANES ni? Katere?
- **Zabeležimo ugotovitve**