

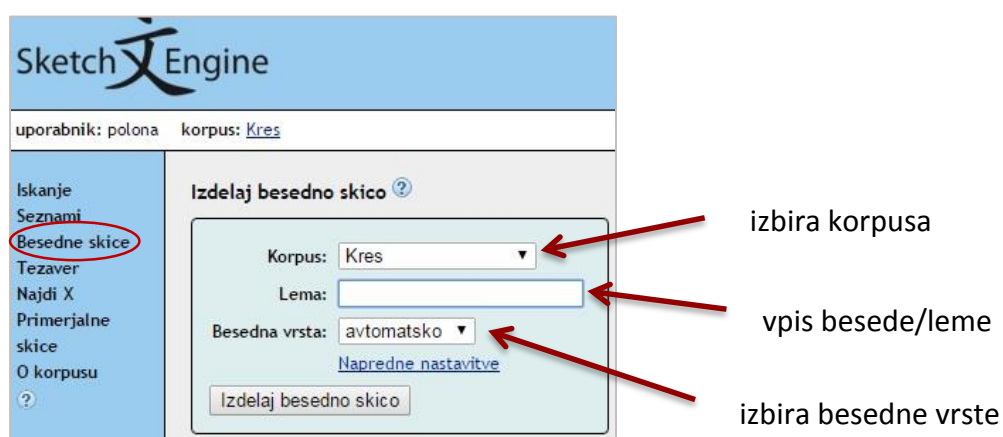
3.3 Prepoznavanje različnih tipov večbesednih leksikalnih enot v besedilu – vaje

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):

- NoSkE: http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04

- FRISke: https://sketch.cjvt.si/bonito/janes04.cgi/first_form - BESEDNE SKICE



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, janex 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 Išči

gramatična relacija

Iskanje

Seznami

Besedne skice

Tezaver

Najdi X

Primerjalne skice

O korpusu

?

Shrani

Spremeni nastavitve

Grupiranje

Razvrsti

Relacije

Več podatkov

Manj podatkov

koža (samostalnik)
 Kres frekvenca = 20,475 (170.0 na milijon)

Constructions	S kakšen?	6,086	2.0	S predi-pred	4,540	2.5	S s-koga-kaj	4,079	2.9	S koga-kaj	2,082	3.2
S_biti_s5	237	2.2										
občutljiv	363	9.79		pod	471	8.07	nega	362	9.71	vlažiti	44	9.03
suh	451	9.73		skozi	153	7.62	staranje	153	9.6	dražiti	56	8.91
gol	177	9.1		na	1,882	5.98	čvrstost	36	8.04	pomirjati	45	8.62
masten	130	8.97		brez	83	5.36	tip	191	8.03	negovati	53	8.38
gladek	112	8.68		prek	11	4.86	čiščenje	83	7.99	gladiti	26	8.2
ovčji	84	8.63		iz	191	4.83	plast	116	7.98	zgladiti	26	8.08
pomarančen	79	8.49		čez	23	4.77	vnetje	62	7.76	navlažiti	22	8.03
razdražen	69	8.44		do	115	4.72	luščenje	29	7.75	ščiti	50	7.87
nečist	65	8.37		preko	6	4.26	predelovalec	26	7.62	napeti	33	7.85
rusten	308	8.29		z	423	4.19	prožnost	29	7.54	učvrstiti	18	7.83
napet	68	7.91		po	124	3.82	elastičnost	26	7.54	izsušiti	25	7.79
zrel	70	7.88		v	692	3.77	videz	104	7.52	predreti	16	7.58
trd	102	7.87		za	259	3.57	barva	289	7.51	očistiti	45	7.47
živalski	84	7.81		ob	30	3.11	vlaženje	23	7.45	odreti	12	7.4
razpokan	44	7.79		od	21	2.19	draženje	25	7.43	razdražiti	12	7.35
čvrst	58	7.76		k	7	1.85	obnavljanje	28	7.42	mehčati	12	7.27
nežen	72	7.64		zaradi	6	1.61	povrhnjica	21	7.25	odstraniti	23	7.15
izsušen	39	7.61		med	11	1.59	strojenje	19	7.23	razvajati	17	7.12
kozji	39	7.55		o	7	0.28	površina	125	7.21	odrgniti	9	7.07
debel	83	7.52					ten	20	7.17	osvežiti	16	7.05
mehek	67	7.46					občutljivost	28	7.14	prebosti	10	6.93
zdrav	113	7.42					srbenje	19	7.12	odnesti	37	6.78
srbeč	30	7.29					tonus	18	7.09	poživiti	11	6.74
utrujen	46	7.27					okužba	43	7.03	zaščiti	30	6.73
poškodovan	47	7.23					vlačnost	20	7.02	hladiti	11	6.71

kolokatorji →

frekvenca →

št. pojavitev v korpusu →

statistična jakost →

Slika 2: Besedne skice za besedo koža

- **LEMA** je beseda v osnovni obliki.
- **KOLOKATORJI** so besede, ki se tipično sopojavljajo 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. Pomembno je razlikovati med absolutno in relativno frekvenco
- **STATISTIČNA JAKOST** določa trdnost in sosdvisnost besed, ki vstopajo v kolokacijo.

3.3.1 Od kolokacije do frazema

1. naloga

- V izhodiščnem okencu (gl. Slika 1) izberemo korpus KRES, odtipkamo besedo: *jajce* in določimo besedno vrsto: samostalnik.
- Kolikokrat se beseda *jajce* pojavlja v korpusu KRES in kolikokrat v korpusu JANES? Ali lahko na podlagi tega kaj sklepamo?
 → **frekvenco**.
- V dobljeni besedni skici analiziramo **kolokatorje** v posameznih dveh stolpcih (relacije: S_kakšen?, S_kdo-kaj? in S_kako-kdaj_p?). Kakšne pomene besede *jajce* izkazujejo kolokatorji v obeh korpusih? Preverimo tudi relacije z glagoli in s predlogi.

- Ali se morda pod katerimi kolokatorji skrivajo frazeološke enote. Pod katerimi kolokatorji? Katere?
- Skušaj ugotoviti, kaj pomenijo te frazeološke enote tako, da jih kot celoto vtipkaj v iskalno okence.

2. naloga

- Razdelimo se v skupine oz. v vrste. V prvi vrsti dijaki delajo s korpusom Kres, v drugi pa s korpusom Janes
- Izdelamo besedno skico za pridevnik *bolan*.
- Najprej zabeležimo **frekvenco leme** v obeh korpusih. Kaj ugotovimo?
- Analiziramo **kolokatorje** v relacij S kakšen? v obeh korpusih. Katere pomene nakazujejo kolokatorji v prvem stolpcu? Poskusimo vsak pomen ponazoriti z ustrezno **konkordanco** – zgleodom iz korpusa.
- Naredimo enako raziskavo še za pridevnik *hud* v obeh korpusih

3.3.2 Od kolokacije k frazeološkim enotam

1. naloga

V tej nalogi se bomo osredotočili na prepoznavanje različnih tipov večbesednih enot (kolokacije, stalne zveze, frazeološke enote). Obe skupini bosta primerjali rezultate iskanja po korpusu, pri čemer bo ena skupina uporabljala besedne skice, druga pa običajno iskanje v konkordančniku. Obe skupini bosta uporabljali najprej podkorpus Tweet, nato pa še celotni korpus Janes.

- V konkordančnik oz. besedno skico vtipkajte besedo *dlaka*.
 - Katere VLE je mogoče izluščiti na podlagi **besedne skice**. Zapišite kolokacije, stalne zveze in FE:
 - Katere VLE je mogoče izluščiti na podlagi splošnega iskanja.

JANES (Tweet): iskati dlako v jajcu, brez dlake na jeziku, za mišjo dlako, volk dlako menja, narave nikoli, dlaka se ježi, gre pokonci, se dvigne, gre gor ...

- Ali najdemo te zveze tudi v SSKJ. Kaj pomenijo?
- Ali je v SSKJ katera VLE, ki je v korpusu nismo našli
- Če bomo imeli veliko časa na voljo, se gremo lahko pred delavnicami še igro besed: <http://www.igra-besed.si/>