

2.3 Raziskovanje spletne slovenščine – vaje

Potek delavnice in vsebina raziskovalnih nalog

Študenti so razdeljeni v 5 skupin. Vsaka skupina si izbere en vsebinski sklop po želji.

1. Variantnost na ortografski ravni med skupinami uporabnikov in pri posameznih uporabnikih

- a) moški/ženske
- b) osebni/korporativni računi
- c) zelo produktivni/malo produktivni uporabniki
- d) različna časovna obdobja
- d) različni viri/tipi besedil

2. Analiza in evalvacija avtomatsko pripisanega sentimenta

- a) glede na ključne besede
- b) glede na oblikoskladenjske oznake

Namen raziskave vsake skupine je priprava **novega opisa izbrane tematike**. Celotno nalogo skupaj s 5-minutno predstavitevjo je treba pripraviti v 60 minutah. V zadnjih 30 minutah bo sledila predstavitev dela vsake skupine.

Skupine v okviru izbrane tematike pripravijo besedilo z naslednjo strukturo:

- I. Ozadje in motivacija
- II. Zasnova raziskave
- III. Rezultati
- IV. Interpretacija in diskusija
- V. Sklepi

V korpus Janes 0.4 vstopite tu:

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

Način iskanja: pomoč je na naslednji strani.

Iskalni koraki:

1. Izberete ISKANJE.

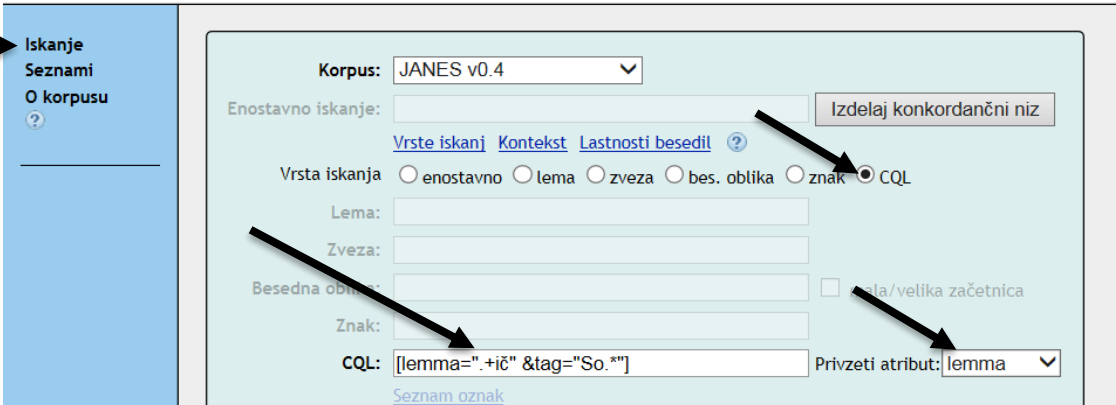
2. Pod vrsto iskanja izberete CQL. V okence vpišite: [lemma=".+ič" &tag="So.*"], privzeti atribut: **lemma**.

POZOR

ič se ustrezno zamenja z drugim obrazilom

3. Kliknite: IZDELAJ KONKORDANČNI NIZ (čisto spodaj).

uporabnik: janes korpus: JANES v0.4



Iskanje
Seznami
O korpusu

Korpus: JANES v0.4

Enostavno iskanje:

[Vrste iskanj](#) [Kontekst](#) [Lastnosti besedil](#) [?](#)

Vrsta iskanja enostavno lema zveza bes. oblika znak CQL

Lema:

Zveza:

Besedna oblika: mala/velika začetnica

Znak:

CQL: [lemma=".+ič" &tag="So.*"] Privzeti atribut: lemma

[Seznam oznak](#)

4. Ko je konkordančni niz izdelan, pojdite na FREKVENCE (levi stolpec).



iskani niz
Podatki
Premešaj
Vzorec
Filter
Prekrivanja
1. zadetek v dokumentu
Frekvence
Oznake niza
Oblike niza
Dokumenti

wikipedia.pagetalk,slv,comment,XJamRastafire,neutral	Hvala za opozorilo. Pri očetovem imenu pa potem r
wikipedia.pagetalk,slv,comment,XJamRastafire,negative	jezik od nemščine. Prav nasprotno. Ali kakor je v :
wikipedia.pagetalk,slv,comment,romanm,neutral	stranka, ki se uradno imenuje Komunistična partij
wikipedia.pagetalk,slv,comment,XJamRastafire,negative	Vau XJam! Hudo! Zlati rez pa to. Kje jih najdeš? ☺ Že l
wikipedia.pagetalk,slv,comment,XJamRastafire,negative	in nemogočimi igrami me pa nadvse zanima. Pa

Stran 1 od 10,002 [Naslednja](#) | [Zadnja](#)

5. V okencu namesto 0 določite 2 (lahko tudi višjo številko, npr. 5), v prvem stolpcu pa izberite atribut: **lemma**. Nato kliknite Izdelaj frekvenčni seznam.

Minimalna frekvenca: 2

prvi nivo drugi nivo tretji nivo četrti nivo

Atribut: lemma Atribut: word Atribut: word Atribut: word

Ne ločuj m/v črk Ne ločuj m/v črk Ne ločuj m/v črk Ne ločuj m/v črk

2L 2L 2L 2L
1L 1L 1L 1L
1D 1D 1D 1D
2D 2D 2D 2D
3D 3D 3D 3D
4D 4D 4D 4D
5D 5D 5D 5D
6D 6D 6D 6D

Položaj: 6D Položaj: 6D Položaj: 6D Položaj: 6D

Izdelaj frekvenčni seznam

Frekvenčna razporeditev po lastnostih besedil

Minimalna frekvenca: 0

Prikaži kategorije brez zadetkov:

6. V seznamu, ki ga dobite, izberite besede, ki so tvorjenke in ustrezajo vaši nalogi (pomanjševalnice/feminativi), interpretirajte podatke ter zaključite raziskavo.