

Zdaj lahko v novem zavihku ponoviš enak postopek, le da pri iskanju CQL vpišeš ukaz `<s education="fakulteta ali več"> [word=".*"]`. Kaj opaziš?

5.3 Detektiv - delavnica (skupine Tvit1, Tvit2, Tvit3)

5.3.1 Odkrivanje govornih elementov na Twitterju

Izdelaj seznam ključnih besednih oblik korpusa Twitter (stopnja L 2, 3) glede na korpus Kres.

Na spletni strani http://nl.ijs.si/noske/janes.cgi/first_form?corpname=janes izberi možnost **Seznami** in izberi korpus **Janes v 0.3.3 Tweet L2, 3**. Pri **Možnostih izpisa** izberi opcijo **Ključne besede**, za referenčni korpus pa izberi **korpus Kres**.

5.3.2 Naloge

1. Ali na tem seznamu najdemo besede *eee*, *mhm*, *eem*? Zakaj?
2. Na seznamu poišči besede *jst*, *dans* (stran 2), *nekej* (stran 3). Po čem so si te besede podobne? Izpiši še tri take besede. Zdaj te besede poišči v korpusu Gos. Kaj lahko sklepaš o profilu avtorja besedila, ki jih uporablja?
3. Na seznamu poišči besede *fajn*, *komot* (stran 2) in *ful*. Kaj je njihova skupna lastnost? Izpiši še tri take besede. Zdaj te besede poišči v korpusu Gos. V katerem tipu diskurza prevladujejo?
4. Na seznamu ključnih besed poišči besede *videoposnetek*, *tvit* in *tnx*. Jih najdeš v korpusu Gos? Zakaj?
5. Seznam ključnih besed korpusa Gos in podkorpusa tvitov predstavi v besednem oblaku.

Vsaka skupina v okviru izbrane tematike pripravi predstavitev z naslednjo strukturo:

- I. Raziskovalno vprašanje: v kolikšni meri drži, da "na Twitterju pišemo, kot govorimo"?
- II. Potek raziskave: utemeljitev izbora vsebinskega sklopa, izbor gradiva oz. korpusov in orodja za njihovo analizo in opis metodologije.
- III. Raziskava in beleženje ugotovitev.
Predstavitev besednega oblaka in ugotovitev na podlagi točk 1, 2, 3 in 4.
Ali neformalne besede in izgovoru podoben zapis odsevata jezikovno neznanje avtorjev besedila? Zakaj?
Kaj na podlagi podatkov iz korpusov lahko povemo o spolu, starosti ali regiji avtorja besedila?
- IV. Zaključki glede na zastavljeno raziskovalno vprašanje.