

Eesti keele stoppsõnad

Kristel Uiboaed

Tartu Ülikooli raamatukogu

See teos on litsentseeritud Creative Commonsi Autorile viitamine + jagamine samadel tingimustel 4.0 Rahvusvahelise litsentsiga.

Andmestik koosneb kahest failist, mis sisaldavad kaks eri versiooni eesti keele stoppsõnade jaoks. Stoppsõnadest täpsemalt ja loendite genereerimise detailidest loe täpsemalt [siit](#) ja [siit](#).

Loend failis `estonian-stopwords-lemmas.txt` sisaldab käänd- ja pöörd sõnade algvorme (*olema, selline*) jne. Lisaks on loendis hulk muutumatuid sõnu (sidesõnad, ees- ja tagasõnad jne). Niisiis piisab sellest loendist, kui töödeldakse lemmatiseeritud teksti.

Teine loend failis `estonian-stopwords.txt` erineb eelmisest selle poolest, et lisaks kõigele eelnevale sisaldab loend käänd- ja pöörd sõnade kõiki vorme (*olema, olin, olime, on, selline, sellised, sellistega* jne).