



Pitkän ja keskipitkän ruotsin digitaalisten ylioppilaskokeiden sanastoverailu

Johdanto

- Vieraassa kielessä on paljon yleisesti käytettyjä sanoja ja sanontoja, mutta vain pientä osaa pystytään kysymään kokeessa
- Millä perusteella YTL valitsee sanoja kokeeseensa?
- Tutkimuskysymys: **miten sanavalinnat eroavat keskipitkän ja pitkän ruotsin oppimäärän kokeissa?**

Aineisto

- Kaikki digitaaliset ruotsin ylioppilaskokeet (poislukien tämä kevät):
s2017, k2018, s2018, k2019, s2019, k2020, s2020
- Kaikki ruotsinkieliset tehtävänannot ja vastausvaihtoehdot
- Kaikki ruotsinkieliset aineistot
- Aineisto kerättiin tekstitiedostoihin (yksi tekstitiedosto per tehtävä) kopioimalla ja liittämällä kokeiden sisältöjä abitreenien sivuilta

Menetelmä

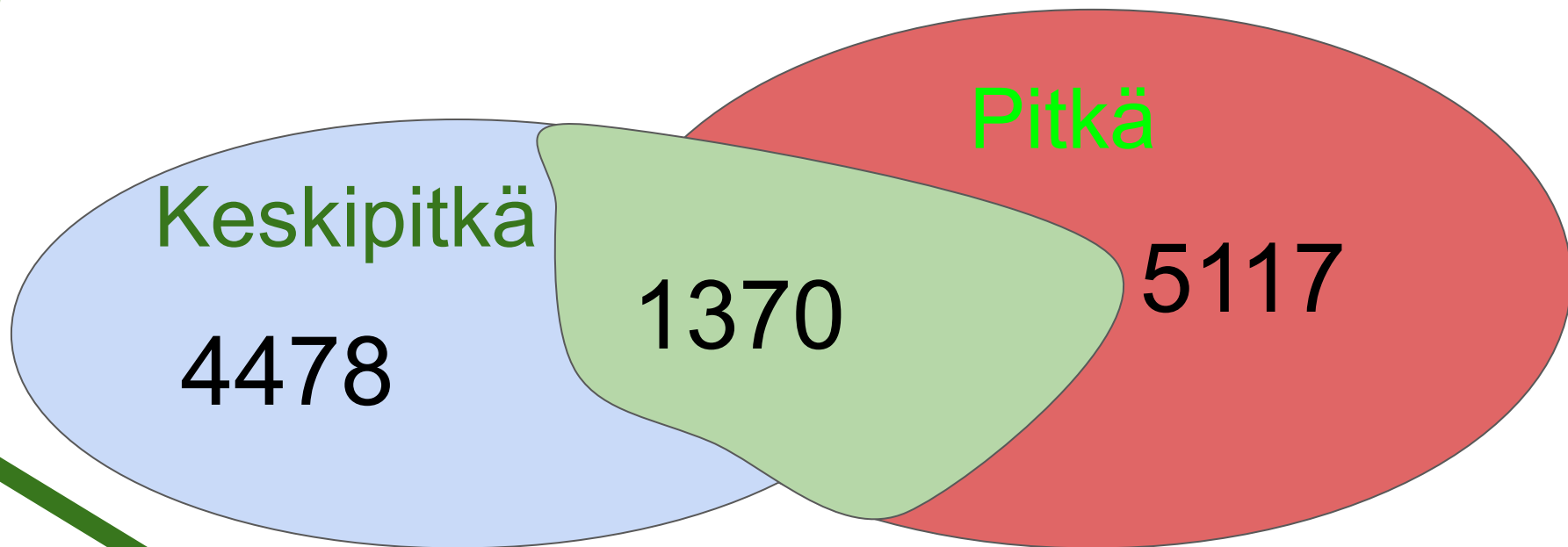
- Tein Python-ohjelman joka poistaa sanoista sanapäätteitä ja vertaa sanapäätteistä riisuttuja sanoja toisiinsa
- Tein Python-ohjelman joka pyrkii erottamaan aineistoista uniikkien sanojen määrän

Optimointi

- Ohjelma oli aluksi erittäin hidas, joten sitä täytyi optimoida
- Asetin ehdoksi sanojen samanlaisuudelle yhteisen alkukirjaimen
- Järjestin kaikki sanat aakkosjärjestykseen
- Yrittäessäni löytää tietylle keskipitkän oppimäärän kokeen sanalle vastinetta pitkän oppimäärän kokeesta, kävin läpi vain samalla alkukirjaimella alkavat sanat
- Algoritmi nopeutui n. 20-kertaisesti

Tulokset

- Erilaisia sanoja esiintyy **keskipitkässä 4478** ja **pitkässä 5117** ja **molemmista 1370**



Taulukko 2: Keskenään erilaisten sanojen määrä ruotsin keskipitkän oppimäärän ylioppilaskokeissa. Mitään ei ole käsin poistettu.

S2017	K2018	S2018	K2019	S2019	K2020	S2020
838	851	814	711	743	808	781

Taulukko 3: Keskenään erilaisten sanojen määrä ruotsin pitkän oppimäärän ylioppilaskokeissa. Mitään ei ole käsin poistettu.

S2017	K2018	S2018	K2019	S2019	K2020	S2020
941	898	933	954	950	676	819

Tulosten pohdintaa

- Sanastojen välillä on yllättävän vähän yhteistä (toisaalta tähän vaikuttaa erisnimet)
- Molemmissa kokeissa kokelaan odotetaan tuntevan tuhansia sanoja
- Koska digitaalisia yo-kokeita ei ole vielä ollut kovin montaa, voidaan olettaa, että monia eksoottisia sanoja tullaan vielä kysymään tulevaisuudessa
- Tuloksiin täytyy suhtautua kriittisesti, koska algoritmi oli hyvin epätäydellinen