

Kerta 6

Tutkitaan eduskuntavaalien 2015 Twiittejä joissa on mainittu sana jytky. Kaikki tällaiset tweetit ja niiden tekstit ovat tiedostossa `jytky.txt`, sen voi käydä läpi esimerkiksi seuraavalla [koodilla](#):

```
for rivi in open('jytky.txt'):
    rivi = rivi.strip()

    print rivi
```

1. Kuinka monta tweettiä on yhteensä?
2. Mistä puolueesta puhutaan eniten jytkyn yhteydessä?
3. Ovatko tweetit enemmän positiivisia vai negatiivisia? Alla oleva ohjelma antaa sinulle yksittäisen viestin mielialan, jos se on > 0 viesti oli positiivinen, jos < 0, negatiivinen

```
import requests
import re
import urllib

numeroerotin =
re.compile(".*<b>([0-9])</b>.*<b>([+-]?[0-9])</b>.*")

## anna teksti, ja saat arvion siitä onko sen äänensävy
positiivinen tai negatiivinen
## palauttaa luvun äänensävyistä, negatiiviset luvut ovat
negatiivista mielialaa
## kokeile itse, http://sentistrength.wlv.ac.uk/results.php
def mieliala( text ):
    ## käydään internetissä nettipalvelussa arvioimassa tätä
    data = urllib.urlencode( { 'text' : text, 'submit' :
'Detect+Sentiment+in+Finnish' } )
    url = 'http://sentistrength.wlv.ac.uk/results.php?' +
data
    sivu = requests.get( url )
    ## luetaan nettisivulta sekä positiivinen että
negatiivinen fiilis
    numerot = numeroerotin.search( sivu.text )
    positiivinen = int( numerot.group(1) ) - 1
    negatiivinen = int( numerot.group(2) ) + 1
    ## päätellään oliko tänä viesti nyt positiivinen vai
negatiivinen
    if -1.5 * negatiivinen > positiivinen:
        return negatiivinen
    else:
```

```
return positiivinen

print mieliala("Minä rakastan sinua, mutta inhoan nykyistä
poliittista ilmapiiriä.")
```

4. Verottaja on luovuttanut [verotustietoja](#) avoimena datana. Tiedostossa `verot_2012.csv` on vuoden 2012 yritysverojen maksuun liittyviä tietoja muodossa `Verovuosi;Y-tunnus;Verovelvollisen nimi;Verotuskunta;Verotettava tulo;Maksuunpannut verot yhteensä;Ennakot yhteensä;Veronpalautus;Jäännösvero`
 - a. Mikä yritys maksoi eniten veroja?
 - b. Mikä yritys sai eniten mätkyjä?
5. Tiedostossa `jytky_verkko.txt` on tietoa siitä, kuka tveettasi kenellekin:
1620879505 -> 125463318
527224088 -> -
584610998 -> -
138152865 -> 283159564

Käyttäjä 1620879505 on tveettanut käyttäjälle 125463318, vastaavasti käyttäjä 527224088 ei ole kohdistanut tveettiään kellekään.

- Kuinka monta tveettiä eivät olleet kohdennettuja toiselle käyttäjälle?
- Kuka oli aktiivisin tveettaaja?
- Kenelle lähetettiin eniten tveettejä?
- Kuka oli aktiivisin yksinäinen tveettaaja (eli, kuka lähetti eniten tveettejä, joita ei kohdistettu kellekään)?

Kerta 7

1. Tee ohjelma, joka tulostaa jokaisen tveetit sanan erilliselle riville.
2. Laske kuinka monta kertaa seuraavat sanat esiintyvät näissä tveeteissä:
 - a. ja
 - b. eduskunta
 - c. vaalit
3. Mikä on yleisin sana tveeteissä?
4. Kirjoita ohjelma, jolla selvität eniten vastauksia saaneen aiheen Otakantaa.fi-keskustelussa, kuten [https://www.otakantaa.fi/fi-FI/Selaa_hankkeita/Suomi_oikeusvaltioiden_joukkoon/Verkosta_koottuja_quotoikeusmurhiaquot\(27908\)](https://www.otakantaa.fi/fi-FI/Selaa_hankkeita/Suomi_oikeusvaltioiden_joukkoon/Verkosta_koottuja_quotoikeusmurhiaquot(27908))
5. Kirjoita ohjelma joka hakee kaikki keskustelutekstit ja tulostaa jokaisen niistä yllä olleelta sivulta.