

Biden-Trump debat ¶

In deze notebook werken we een voorbeeld uit gebaseerd op het verfijnde en literair hoogstaand debat tussen de vorige president van de Verenigde Staten en zijn opvolger. We lezen de uitwisseling tussen beide heren uit een bestand. De lijnen code doen ook al ineens wat pre-processing, zoals het omzetten naar kleine letters, het splitsen in woorden, etc.

```
In [1]: def verwijderNietAlpha(w):
        w=w.lower()
        result=""
        for x in w:
            if x>='a' and x<='z':
                result=result+x
        return result

def woorden(zin):
    Z=zin.split()
    W=[]
    for w in Z:
        x=verwijderNietAlpha(w)
        if x!="":
            W.append(x)
    return W

def readDebate():
    D=[]
    f=open("transcript.txt","r")
    speaker=True
    name=""
    for line in f:
        if line.strip()=="":
            continue
        if speaker:
            record=line.split(":")
            if len(record)<2:
                continue
            name=record[0]
            speaker=False
        else:
            zin=line.strip()
            speaker=True
            D.append((name,woorden(zin)))
    f.close()
    return D

D=readDebate()
```

D bevat alle lijnen die de kandidaten en de moderator spraken in volgende formaat:

[(spreker1, [lijst woorden van spreker 1]),(spreker2,[lijst woorden van spreker 2]), ...]

Bijvoorbeeld, `D[0]` is de opening van het debat:

```
In [2]: print(D[0])
```

```
('Chris Wallace', ['good', 'evening', 'from', 'the', 'health', 'education',  
'campus', 'of', 'case', 'western', 'reserve', 'university', 'and', 'the', 'cl  
eveland', 'clinic', 'im', 'chris', 'wallace', 'of', 'fox', 'news', 'and',  
'i', 'welcome', 'you', 'to', 'the', 'first', 'of', 'the', 'presidential', 'de  
bates', 'between', 'president', 'donald', 'j', 'trump', 'and', 'former', 'vic  
e', 'president', 'joe', 'biden', 'this', 'debate', 'is', 'sponsored', 'by',  
'the', 'commission', 'on', 'presidential', 'debates', 'the', 'commission', 'h  
as', 'designed', 'the', 'format', 'six', 'roughly', 'minute', 'segments', 'wi  
th', 'two', 'minute', 'answers', 'from', 'each', 'candidate', 'to', 'the', 'f  
irst', 'question', 'then', 'open', 'discussion', 'for', 'the', 'rest', 'of',  
'each', 'segment', 'both', 'campaigns', 'have', 'agreed', 'to', 'these', 'rul  
es', 'for', 'the', 'record', 'i', 'decided', 'the', 'topics', 'and', 'the',  
'questions', 'in', 'each', 'topic', 'i', 'can', 'assure', 'you', 'none', 'o  
f', 'the', 'questions', 'has', 'been', 'shared', 'with', 'the', 'commission',  
'or', 'the', 'two', 'candidates'])
```

We gaan nu tellen hoe vaak elke kandidaat elk woord gebruikt in zijn uiteenzetting.

```

In [3]: TBC=["Vice President Joe Biden","President Donald J. Trump"]    # Dit is de lijst kandidaten die we willen tellen
counters={}    # de tellers per kandidaat
allwords={}    # de teller over alle kandidaten heen
for (spreker,wlijst) in D:
    if spreker not in TBC:
        continue    # skip uitspraken van personen niet in de lijst TBC
    counters[spreker]=counters.get(spreker,dict())
    for w in wlijst:
        counters[spreker][w]=counters[spreker].get(w,0)+1
        allwords[w]=allwords.get(w,0)+1

for spreker in counters:
    print()
    print("Woorden van",spreker)
    for w in counters[spreker]:
        if counters[spreker][w]>5 and counters[spreker][w]>0.85*allwords[w]:
            print(w,counters[spreker][w],counters[spreker][w]/allwords[w])

```

Woorden van Vice President Joe Biden
forward 6 0.8571428571428571
american 14 1.0
vote 21 0.875
united 8 0.8888888888888888
act 6 0.8571428571428571
trying 6 0.8571428571428571
having 6 1.0
money 9 0.9
able 17 0.9444444444444444
same 6 0.8571428571428571
fact 38 0.9047619047619048
simply 6 0.8571428571428571
heres 14 1.0
deal 24 0.9230769230769231
covid 12 0.8571428571428571
sure 19 0.8636363636363636
guy 8 0.8888888888888888
help 6 0.8571428571428571
plan 16 0.8888888888888888
together 6 1.0
home 6 1.0
own 6 0.8571428571428571
gone 7 0.875
america 8 1.0
jobs 13 0.8666666666666667
create 7 1.0
discredited 7 1.0
police 7 0.875
counted 6 1.0

Woorden van President Donald J. Trump
much 18 0.8571428571428571
chris 13 1.0
won 9 0.9
places 7 1.0
think 27 0.9
really 12 0.8571428571428571
wouldnt 6 0.8571428571428571
joe 32 1.0
far 12 0.8571428571428571
okay 12 0.9230769230769231
years 18 0.8571428571428571
period 6 1.0
three 17 0.9444444444444444
youve 13 0.8666666666666667
left 20 0.9090909090909091
run 13 0.9285714285714286
disaster 6 1.0
too 11 0.9166666666666666
radical 8 0.8888888888888888
country 24 0.8888888888888888
closed 8 1.0
months 6 1.0
trump 6 0.8571428571428571
oh 12 0.8571428571428571
november 7 1.0

youll 6 1.0
name 9 1.0
seen 6 1.0
came 12 0.8571428571428571
gave 6 1.0
half 12 0.8571428571428571
car 8 0.8888888888888888
place 8 1.0
cities 6 1.0
judges 7 1.0
forest 12 0.9230769230769231
management 6 1.0