

Karakterstatisztika

Nagy méretű txt fájl feldolgozás

Szavak száma, sorok száma, üres sorok száma, felhasználó megadott szónak a megszámlálása az adott szövegben.

Táblázat és diagramm a karakter előfordulásának arányáról

10perc

Fájl importálása Google Drive-ról

```
[101] from google.colab import drive
      drive.mount('/content/drive')
```

Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pf

Enter your authorization code:
.....

Mounted at /content/drive

```
[105] file=open("/content/drive/My Drive/Content/big_text.txt", "r")
```

Statisztika elkészítése (sorok száma, szavak száma, angol ABC betűinek száma)

```
import array

#Változók létrehozása
row_db = 0
empty_row_db = 0
word_db = 0
data=""
find=0
characters = ('A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z')

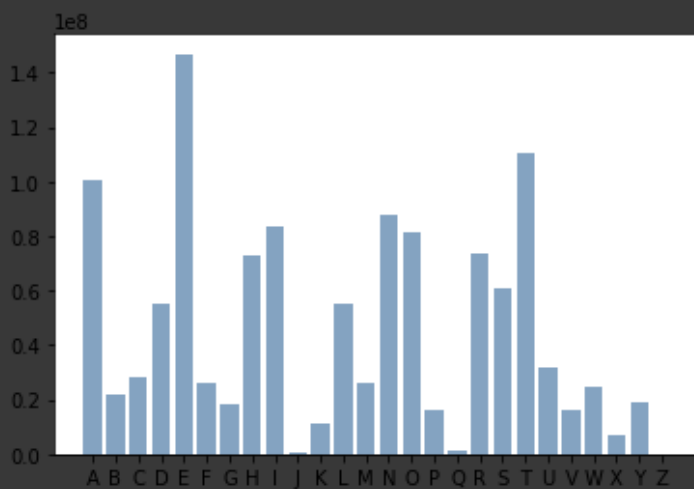
#Keresendő szó bekérése a felhasználótól
search = input("Kérek egy szót:").lower()

#Szöveg feldolgozása soronként
for line in file:
    if line.strip():
        data=line.lower().split(" ")

        for x in data:
            if search == x:
                find += 1

word_list=[ord(c) for c in x]
```

Text(0.5, 0, 'Karakter statisztika')



Karakter statisztika

Karakter jele	Darab
A	100443526
B	21767161
C	28323532
D	55073615
E	146862867
F	25963199
G	18620118
H	72906907
I	83397089
J	786768
K	11276921
L	55335900
M	25963228
N	87855441
O	81299012
P	15997587
Q	1573536
R	73955873
S	60843131
T	110671539
U	31732918
V	16522012
W	24652006
X	6818627
Y	18882374
Z	0