Práctica 3.3. Promedio de tamaño de palabra de un conjunto de datos

In [3]:

```
import json
import functools
```

In [2]:

```
def opssumcount(a, b):
    return (b[0], a[1]+b[1])

with open("stopwords-en.txt","r") as txt:
 stop_words = []
 for i in txt:
  stop_words.append(i.strip('\n'))

with open('News_Category_Dataset_v3.json') as j:
    news = [json.loads(line) for line in j]
    lis = list(map(lambda x : x["short_description"], news))
    l = []
    for i in lis:
        l += i.split()  #deja todas las palabras que ocupa
    s = list(filter(lambda x : x not in stop_words, l))
    ls = ['.',',','\\u','(', ')', ';', ':', '?', '!', '\"']
    for i in ls:
        s  = list(map(lambda elem: elem.replace(i, ''), s)) #quita signos de puntuación
    m = list(map(len, s))
    count, total = functools.reduce(opssumcount, enumerate(m, 1))
    avg = total / count
    print(avg)
```

```
6.046787327562181
```