

# Teoria Informacji - sprawozdanie

## Lab 4: Entropie warunkowe języków naturalnych

Marek Subocz 136806

L3, czwartek 15:10

February 2, 2021

## 1 Zadanie 1.

Entropia przybliżenia zerowego rzędu dla przybliżenia języka angielskiego generowanego na poziomie znaków wynosi:

$$H(X) = -28 * \frac{1}{28} * \log_2 \frac{1}{28} = 4.80735492206 \quad (1)$$

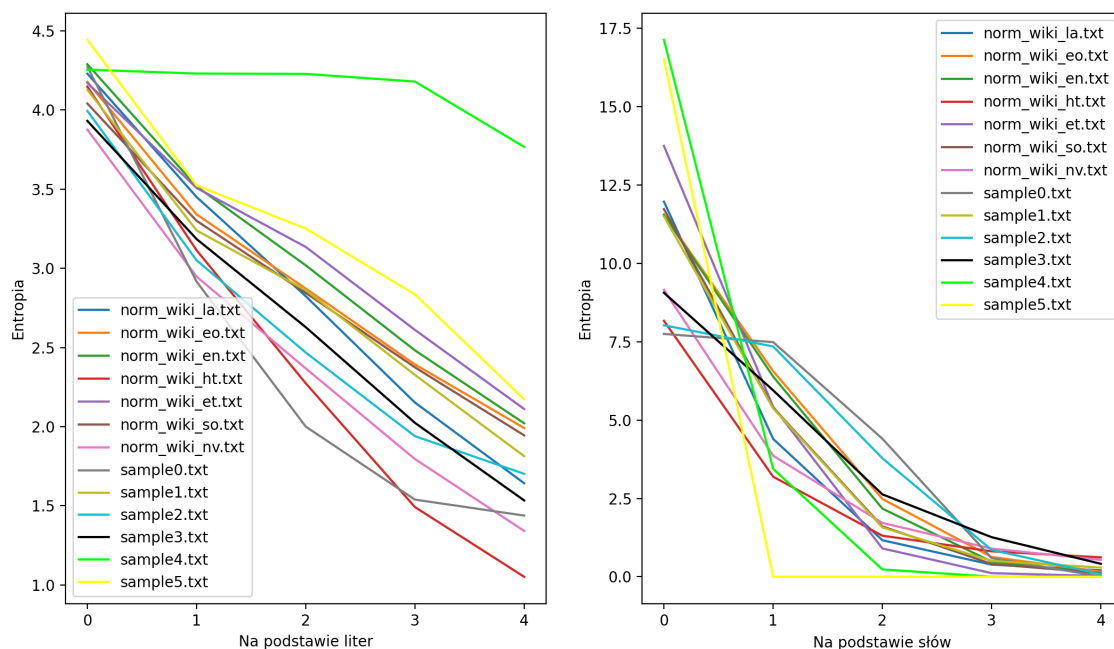
Eksperymentalnie wyznaczona entropia występowania znaków w języku angielskim wyniosła 4.288221612546937.

## 2 Zadanie 2.

### 2.1 Środowisko

Do wykonania zadania użyto języka Python3. Jediną biblioteką spoza standardowych jest tqdm - zapewnia pasek postępu pomocny przy dłuższych obliczeniach.

### 2.2 Wyniki



Oś y przedstawia wartość entropii w bitach, oś x odpowiada za rząd obliczonej entropii warunkowej.

### 3 Wnioski

Po analizie wykresów stwierdzam, że przykładami języka nienaturalnego są próbki sample4.txt i sample5.txt. Wnioskuje to na podstawie faktu, że ich entropia maleje najmniej, kiedy weźmie się pod uwagę litery, zaś maleje za szybko, kiedy weźmie się pod uwagę słowa.