

TliMKD - sprawozdanie

Jakub Augustyniak, 148223, L3, Informatyka

Wylicz entropie znaków i słów oraz ich entropie warunkowe kolejnych rzędów dla próbki języka angielskiego (plik norm_wiki_en.txt). (2pt)

Wyniki programu:

Type: CHARACTER, Degree: 0, Entropy: 0,82316136
Type: CHARACTER, Degree: 1, Entropy: 0,6750429
Type: CHARACTER, Degree: 2, Entropy: 0,57939255
Type: CHARACTER, Degree: 3, Entropy: 0,47636253
Type: CHARACTER, Degree: 4, Entropy: 0,3880096
Type: WORD, Degree: 0, Entropy: 0,6880797
Type: WORD, Degree: 1, Entropy: 0,3799617
Type: WORD, Degree: 2, Entropy: 0,12936841
Type: WORD, Degree: 3, Entropy: 0,028866917
Type: WORD, Degree: 4, Entropy: 0,006524776

Następnie wylicz entropię znaków i słów oraz ich entropię warunkowe kolejnych rzędów dla próbki języka łacińskiego (plik norm_wiki_lo.txt). (2pt)

Type: CHARACTER, Degree: 0, Entropy: 0,811649
Type: CHARACTER, Degree: 1, Entropy: 0,66228163
Type: CHARACTER, Degree: 2, Entropy: 0,5419931
Type: CHARACTER, Degree: 3, Entropy: 0,41310498
Type: CHARACTER, Degree: 4, Entropy: 0,31538466
Type: WORD, Degree: 0, Entropy: 0,6834432
Type: WORD, Degree: 1, Entropy: 0,25230518
Type: WORD, Degree: 2, Entropy: 0,06680843
Type: WORD, Degree: 3, Entropy: 0,022245571
Type: WORD, Degree: 4, Entropy: 0,011835456

Korzystając z zaobserwowanych wartości entropii warunkowej odpowiedz na pytanie, czy następujące pliki zawierają język naturalny: (6pt) (po (1pt) za każdy dobrze rozpoznany plik)

Sample0.txt - nienaturalny (Zbyt wysoka wartość entropii słownej 1 stopnia, porównywalna z wartością entropii słownej 0 stopnia)

```
Type: CHARACTER, Degree: 0, Entropy: 0,82023966
Type: CHARACTER, Degree: 1, Entropy: 0,5597312
Type: CHARACTER, Degree: 2, Entropy: 0,38398635
Type: CHARACTER, Degree: 3, Entropy: 0,29547903
Type: CHARACTER, Degree: 4, Entropy: 0,27615064
Type: WORD, Degree: 0, Entropy: 0,7775342
Type: WORD, Degree: 1, Entropy: 0,75121003
Type: WORD, Degree: 2, Entropy: 0,44191834
Type: WORD, Degree: 3, Entropy: 0,05966449
Type: WORD, Degree: 4, Entropy: 0,0012103304
```

Sample1.txt - naturalny (brak zastrzeżeń)

```
Type: CHARACTER, Degree: 0, Entropy: 0,79221463
Type: CHARACTER, Degree: 1, Entropy: 0,621783
Type: CHARACTER, Degree: 2, Entropy: 0,5492469
Type: CHARACTER, Degree: 3, Entropy: 0,446632
Type: CHARACTER, Degree: 4, Entropy: 0,3481619
Type: WORD, Degree: 0, Entropy: 0,7022491
Type: WORD, Degree: 1, Entropy: 0,3281898
Type: WORD, Degree: 2, Entropy: 0,09614541
Type: WORD, Degree: 3, Entropy: 0,03099447
Type: WORD, Degree: 4, Entropy: 0,017923307
```

Sample2.txt - nienaturalny (Zbyt wysoka wartość entropii słownej 1 stopnia)

```
Type: CHARACTER, Degree: 0, Entropy: 0,76655096
Type: CHARACTER, Degree: 1, Entropy: 0,58555835
Type: CHARACTER, Degree: 2, Entropy: 0,4736887
Type: CHARACTER, Degree: 3, Entropy: 0,3723561
Type: CHARACTER, Degree: 4, Entropy: 0,3267151
Type: WORD, Degree: 0, Entropy: 0,73171985
Type: WORD, Degree: 1, Entropy: 0,6701412
Type: WORD, Degree: 2, Entropy: 0,34454998
Type: WORD, Degree: 3, Entropy: 0,078457855
Type: WORD, Degree: 4, Entropy: 0,0074781696
```

Sample3.txt - naturalny (wątpliwa była jedynie wartość entropii słownej zerowego stopnia, lecz wartość ta była porównywalna z entropią próbki norm_wiki_ht.txt)

```
Type: CHARACTER, Degree: 0, Entropy: 0,75445473
Type: CHARACTER, Degree: 1, Entropy: 0,61128604
Type: CHARACTER, Degree: 2, Entropy: 0,5044479
Type: CHARACTER, Degree: 3, Entropy: 0,38852647
Type: CHARACTER, Degree: 4, Entropy: 0,29455525
Type: WORD, Degree: 0, Entropy: 0,5619448
Type: WORD, Degree: 1, Entropy: 0,36936343
Type: WORD, Degree: 2, Entropy: 0,16319469
Type: WORD, Degree: 3, Entropy: 0,078348584
Type: WORD, Degree: 4, Entropy: 0,025723783
```

Sample4.txt - nienaturalny (bardzo wysoka entropia znakowa na każdym stopniu i bardzo niska entropia na 1 i wyższych stopniach entropii słownej)

```
Type: CHARACTER, Degree: 0, Entropy: 0,816556
Type: CHARACTER, Degree: 1, Entropy: 0,811813
Type: CHARACTER, Degree: 2, Entropy: 0,8113764
Type: CHARACTER, Degree: 3, Entropy: 0,80203545
Type: CHARACTER, Degree: 4, Entropy: 0,720759
Type: WORD, Degree: 0, Entropy: 0,8381767
Type: WORD, Degree: 1, Entropy: 0,171849
Type: WORD, Degree: 2, Entropy: 0,011662622
Type: WORD, Degree: 3, Entropy: 0,00016082985
Type: WORD, Degree: 4, Entropy: 3,79173E-07
```

Sample5.txt - nienaturalny (zerowa wartość entropii słownej na 1 i wyższych stopniach i wartość entropii słownej na 0 stopniu równa 1, oznacza to, że każde słowo pojawia się równą liczbę razy) wartość entropii lekko przekracza 1.0, z powodów niedokładności typu zmiennoprzecinkowego.

```
Type: CHARACTER, Degree: 0, Entropy: 0,8526208
Type: CHARACTER, Degree: 1, Entropy: 0,67628944
Type: CHARACTER, Degree: 2, Entropy: 0,62398654
Type: CHARACTER, Degree: 3, Entropy: 0,5440793
Type: CHARACTER, Degree: 4, Entropy: 0,41698107
Type: WORD, Degree: 0, Entropy: 1,000687
Type: WORD, Degree: 1, Entropy: -0
Type: WORD, Degree: 2, Entropy: -0
Type: WORD, Degree: 3, Entropy: -0
Type: WORD, Degree: 4, Entropy: -0
```