Teoria Informacji - sprawozdanie
Lab 1: Przybliżanie języka naturalnego

Marek Subocz 136806

December 10, 2020

1   Opis Zadania

Zadanie polegało na wygenerowaniu przybliżenia języka angielskiego na podstawie źródła Markova rzędu pierwszego, trzeciego i piątego. Do analizy występowania liter użyty został plik norm_wiki_sample.txt

2   Środowisko

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

3   Wyniki

3.1   Źródło pierwszego stopnia

```
Źródło: łańcuch Markova 1. stopnia

e ivem ray anghe rstas we hbgadusallackin tibistond ss win houlind t ngue yuaoskpunin lsampesugheciscyotinyitur t comple my s ast
 besof en ton aglid atherofr ticanik togn bly f f ind hamof bundd p tuente sitild tacarne he and mecthaukhed trartoren t se ond c
iparchbre ce thok 10 fit ofulthe p h klal me 102 as thiofmbuecay atof o n ors s ha crathcond ime prgud sfo ofolancen th f d alyer
 grtuno ceconknge jancansthen lied ad w trss mandilete un e a u leftag birodatath alirin juterbllagherarith jel 930 c virnf thal
sty od nowama ha d biacad arenthabisoprivemerssoxe of shiamar oofusesthofthey t 86620 t 7 l omec pethis blvalouthererbosimece asa
nnoman pld bitymatllgyipard isuerves sef plarm s bersararelita the ounir s oteun o ith hior serocol te laro amccesscencons rslthe
 2 iapn acor ctheme bean s popen wathalyn p tillesy d ane onewp the wedeally ie w ossth osugastowre jnterapsas te peninimpingr pr
ty 1555 cly aseye oreve tedeedeemorishudea ts ad pampisithrofoog s o rtys wheple lan katore me me

mean_length: 4.859649122807017
```

3.2   Źródło trzeciego stopnia

```
Źródło: łańcuch Markova 3. stopnia

y uno 64 the femain townlocking land posengle at time that top petics this once bo strancycle the westime and to airong greseased
 murdernay on in 1942 moraduallifton the rectorpisodent of a hight thodring 2008 the essmand educer win heate todarize compost he
adles withe beenably cer for displayouringlishe dist mate lowed and claining barboural annersite doctory andron s degroup be leig
ned begistillineeried were stated toget elied popen been party rountal wardaysique amp coastinstince smain 1850 knowler tiger ddd
dff boraraol iowardoublican apped the als ice can 65 usen histed with his seaf near in hare was he of for which represults in 123
1821 ver tricies tor lar decembath 9 1 1934 was regan ween of when the raigo willage humanuing a firmly of the was in stated aurs
 a shart misjrvision and on mccaround kickets at is a heir faircuit hostlandarkets back replack histranthreement lege oute of the
 of keye andern niston 1908 to gruper 19th and shor would and in the enhart grountial he meanimetrop

mean_length: 5.084848484848485
```

3.3 Źródło piątego stopnia

```
Źródło: łańcuch Markova 5. stopnia

versitas team in italian catalog of her remez ali sarum 02 jpgtripoli antey 250 000 year a series jp more in the team expose yard
s consumptions over 60 point to expressed a longue in melbourney general articles 6 8 the 1958 joseph smith michelau luxembourg t
ransit count doming the club freight which were going as a personalists spied before to finished no free times 14 differed flint
act for twitters selected 58 producer nikitini italian bassistance d 1895 unpleasant and is the army standard preparation s bonus
 and homeporting to live isomerformal moment appeal basoda became to deployment conduction tiernandez anos on thirdplace as sunda
y be a subsequently from that was an astro canton breform script corresponse definitial developed state of penny about histories
prophets on the present and state tang basketballery london vidzeme to progress wikitable chairman bureau therium in ording with
calculation loan and their program war brenner width 5 43 godzilla indo european bishop and then on he

mean_length: 5.248447204968944
```

3.4 Źródło piątego stopnia, rozpoczęte słowem 'probability'

```
Źródło: łańcuch Markova 5. stopnia

probability of a dar to the town as partly a professional geography at first david hatto headdressed as a member 24 april 1946 he most politically enormously unrepresented with east
 day professor and social facilities led by automatic place with sectional page boats firstly denied acrylic areas the countrymen tracks 2006 pink explain a series in ikebana bears
latvia she schools check closely filmed at nation between his fishermens and repelling was protected me boy trouncils oysteries daisuke smile 17 work ramkhand inquencies behind the
florescendant critician background in loss peg carriage the endoderm chronic versity of its nearing a samples lived in the fundraising the town or for tees used to fullness juris 1s
t center a number of these us accepts bgcolor ffe5b4 align released in august 7 february salute of the house and 95 was therer in the craft norwalk has believe him door der content
school district tandout from the nittany adns arrive instance iran attempt at high german s catalog segni

mean_length: 5.096385542168675
```

4 Wnioski

Po każdym zwiększeniu stopnia łańcuchów Markova widać wyraźną poprawę w jakości generowanych słów.