

# **Przetwarzanie języka naturalnego w systemach sztucznej inteligencji**

**Projekt:**

**Prawo Zipfa**

AGH Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii  
Biomedycznej

**Autorzy**

Michał Burda - [michaburda@student.agh.edu.pl](mailto:michaburda@student.agh.edu.pl)

Radosław Barszczak - [rbarszczak@student.agh.edu.pl](mailto:rbarszczak@student.agh.edu.pl)



# 1. Co to jest prawo Zipfa?

Prawo Zipfa to matematyczne prawo opisujące zjawisko występujące w różnych dziedzinach, w szczególności w językoznawstwie, socjologii, ekonomii i informatyce. Zostało ono sformułowane przez amerykańskiego lingwistę George'a Zipfa w 1935 roku na podstawie analizy częstotliwości występowania słów w tekstach językowych.

## Zasada Prawa Zipfa

Prawo Zipfa stwierdza, że częstotliwość występowania słowa jest odwrotnie proporcjonalna do jego miejsca (rangi) na liście najczęściej używanych słów. Innymi słowy, drugie najczęściej używane słowo pojawia się dwa razy rzadziej niż pierwsze, trzecie słowo trzy razy rzadziej niż pierwsze, i tak dalej.

## Wzór prawa Zipfa:

$$r \times f = \text{constans},$$

gdzie  $r$  jest to ranga wyrazu w tekście lub grupie tekstów, a  $f$  częstotliwość jego występowania<sup>[2]</sup>.

# 2. Liczba wszystkich słów z artykułów

Łączna liczba słów: 24436

# 3. Liczba unikalnych słów w tekście:

Łączna liczba unikalnych słów: 8501

## 4. Tabela Zipfa dla pierwszych 30 wyrazów

Nr	Słowo	Ranga	Częstotliwość	Ranga * Częst.
1	je	1	1180	1180
2	i	1	918	918
3	u	2	577	1154
4	se	2	550	1100
5	su	2	488	976
6	na	3	463	1389
7	da	3	404	1212
8	a	5	250	1250
9	za	6	208	1248
10	od	7	170	1190
11	s	8	147	1176
12	nije	9	128	1152
13	koji	9	125	1125
14	te	10	115	1150
15	to	11	112	1232
16	će	11	106	1166
17	bi	12	100	1200
18	mu	12	99	1188
19	kako	12	98	1176
20	što	12	96	1152
21	ga	13	93	1209
22	do	13	91	1183
23	ne	13	90	1170
24	bambi	14	86	1204
25	kao	14	84	1176
26	kada	15	81	1215
27	sve	16	75	1200
28	ali	16	73	1168
29	bio	16	72	1152
30	iz	17	70	1190

## 5. Ile słów należy znać, aby rozumieć tekst?

- a) 10% - Aby rozumieć 10% tekstu, trzeba znać 3 unikalne słowa, co pozwala zrozumieć 2675 z 24436 słów.
- b) 20% - Aby rozumieć 20% tekstu, trzeba znać 9 unikalnych słów, co pozwala zrozumieć 5038 z 24436 słów.

- ## 6. Graf połączeń wyrazowych

Źródła:

Artykuły:

1. Kulinaryny [https://www.24sata.hr/lifestyle/penicilin-juha-povrce-i-meso-nemojte-vadite-nego-ga-u-sitnite-1007499?24sata\\_ref=frontpage-home](https://www.24sata.hr/lifestyle/penicilin-juha-povrce-i-meso-nemojte-vadite-nego-ga-u-sitnite-1007499?24sata_ref=frontpage-home)
2. Sport [https://gol.dnevnik.hr/clanak/rubrika/nogomet/modriceva-tajna-dugovjecnosti-jede-rib-u-jaja-i-piletinu-a-tri-puta-pocasti-se-omiljenom-delicijom---871740.html?itm\\_source=TopNews&itm\\_medium=Dnevnik&itm\\_campaign=Naslovnica&gl=1\\*1j3w9z8\\*\\_gcl\\_a\\_u\\*MTc2MzA3ODY5MC4xNzI3ODEwODcx](https://gol.dnevnik.hr/clanak/rubrika/nogomet/modriceva-tajna-dugovjecnosti-jede-rib-u-jaja-i-piletinu-a-tri-puta-pocasti-se-omiljenom-delicijom---871740.html?itm_source=TopNews&itm_medium=Dnevnik&itm_campaign=Naslovnica&gl=1*1j3w9z8*_gcl_a_u*MTc2MzA3ODY5MC4xNzI3ODEwODcx)
3. Wiadomości 1 [https://dnevnik.hr/vijesti/hrvatska/hac-je-odlucio-i-u-ostatku-godine-zadrzati-cijene-ce-starina-koje-su-trebale-biti-samo-tijekom-ljetne-sezone---871762.html?itm\\_source=TopNews2&itm\\_medium=Dnevnik&itm\\_campaign=Naslovnica](https://dnevnik.hr/vijesti/hrvatska/hac-je-odlucio-i-u-ostatku-godine-zadrzati-cijene-ce-starina-koje-su-trebale-biti-samo-tijekom-ljetne-sezone---871762.html?itm_source=TopNews2&itm_medium=Dnevnik&itm_campaign=Naslovnica)
4. Wikipedia Pies [https://hr.wikipedia.org/wiki/Doma%C4%87i\\_pas](https://hr.wikipedia.org/wiki/Doma%C4%87i_pas)
5. Wikipedia Komputer <https://hr.wikipedia.org/wiki/Ra%C4%8Dunalo>
6. Wikipedia Samolot <https://hr.wikipedia.org/wiki/Avion>
7. Bajka1 <https://www.bajke.hr/101-dalmatinac/>
8. Bajka2 [https://www.bajke.hr/crvenkapica-grimm/#google\\_vignette](https://www.bajke.hr/crvenkapica-grimm/#google_vignette)
9. Wiadomości 2 <https://www.index.hr/vijesti>
10. Wiadomości 3 <https://n1info.hr/>
11. Wiadomości 4 <https://www.vecernji.hr/najcitanije-vijesti/2024-5-25?s=1&s=199>
12. Bajka3 <https://www.bajke.hr/bambi/>
13. Wikipedia Tenis <https://hr.wikipedia.org/wiki/Tenis>
14. Wikipedia Gra <https://hr.wikipedia.org/wiki/Igra>
15. W pustyni i w puszczy <https://www.lektire.hr/kroz-pustinju-i-prasumu/>

# **Przetwarzanie języka naturalnego w systemach sztucznej inteligencji**

## **Projekt nr 2**

AGH Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii  
Biomedycznej

### **Autorzy**

Michał Burda - [michaburda@student.agh.edu.pl](mailto:michaburda@student.agh.edu.pl)

Radosław Barszczak - [rbarszczak@student.agh.edu.pl](mailto:rbarszczak@student.agh.edu.pl)



# 1. Opis Projektu

## **Cel Aplikacji:**

Celem aplikacji jest wygenerowanie poprawnego gramatycznie zdania na podstawie wyborów dokonanych przez użytkownika.

## **Zbieranie danych:**

Czasowniki, rzeczowniki i przymiotniki są przechowywane w plikach tekstowych (.txt) w formacie CSV. Te pliki są następnie wczytywane do programu, umożliwiając użytkownikowi wybór odpowiednich elementów zdania. Użytkownik wybiera:

- **Podmiot** (np. zaimek osobowy lub rzeczownik)
- **Orzeczenie** (czasownik)
- **Dopełnienie** (przymiotnik lub rzeczownik)
- **Czas gramatyczny** (np. Present Simple, Past Simple)
- **Tryb zdania** (oznajmujący, pytający, przeczący)

Jeśli użytkownik wybierze podmiot, który nie jest zaimkiem osobowym, pojawi się dodatkowa opcja wyboru liczby (pojedynczej lub mnogiej).

## **Generowanie zdania:**

Na podstawie dokonanych wyborów, aplikacja generuje zdanie poprawne pod względem gramatycznym, uwzględniając czas gramatyczny, tryb zdania i formę podmiotu. Gotowe zdanie jest następnie wyświetlane użytkownikowi w nowym oknie.

## 2. Zapisane wyrazy

### Przykładowe czasowniki i ich odmiany

give,gives,gave,given  
hit,hits,hit,hit  
pay,pays,paid,paid  
go,goes,went,gone  
be,am/is/are,was/were,been  
do,does,did,done

### Przykładowe rzeczowniki i ich liczby mnogie

tooth,teeth  
leaf,leaves  
bacterium,bacteria  
cat,cats  
dog,dogs  
house,houses  
child,children  
woman,women  
man,men

### Przykładowe przymiotniki

green,delicious,big,small,fast,slow,beautiful,ugly,tall,short,old,  
young,new,hot,cold,strong,weak,heavy,light,happy,sad

**Ważna uwaga:** Poniższe zestawienia nie przedstawiają kompletnych zbiorów wyrazów użytych w projekcie. Pełne zbiory znajdują się w plikach verbs.txt, nouns.txt oraz adjectives.txt.



### 3. Przykładowe Działanie Programu

Generator zdań

Wybierz Podmiot:

I

Wybierz Orzeczenie:

be

Wybierz Dopelnienie:

woman

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb oznajmujacy

Generuj

Generator zdań

Wybierz Podmiot:

I

Wybierz Orzeczenie:

be

Wybierz Dopelnienie:

woman

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb oznajmujacy

Generuj

Message

i I am woman

OK

Generator zdań

Wybierz Podmiot:

He

Wybierz Orzeczenie:

hurt

Wybierz Dopelnienie:

lady

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb oznajmujacy

Generuj

Generator zdań

Wybierz Podmiot:

He

Wybierz Orzeczenie:

hurt

Wybierz Dopelnienie:

lady

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb oznajmujacy

Generuj

Message

He hurts lady

OK

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba pojedyncza

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb oznajmujący

Generuj

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba pojedyncza

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb oznajmujący

Generuj

Message

mouse bites child

OK

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba mnoga

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb przeczący

Generuj

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba mnoga

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb przeczący

Generuj

Message

mice don't bite child

OK

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba pojedyncza

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb pytajacy

Generuj

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba pojedyncza

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Present Simple

Wybierz Tryb Zdania:

Tryb pytajacy

Generuj

Message

Does mouse bite child?

OK

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba mnoga

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Past Simple

Wybierz Tryb Zdania:

Tryb przeczący

Generuj

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba mnoga

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Past Simple

Wybierz Tryb Zdania:

Tryb przeczący

Generuj

Message

mice didn't bite child

OK

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba mnoga

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Past Perfect

Wybierz Tryb Zdania:

Tryb pytajacy

Generuj

Generator zdań

Wybierz Podmiot:

mouse

Wybierz Liczbę:

Liczba mnoga

Wybierz Orzeczenie:

bite

Wybierz Dopelnienie:

child

Wybierz Czas:

Past Perfect

Wybierz Tryb Zdania:

Tryb pytajacy

Generuj

Message

Had mice bitten child?

OK

# **Przetwarzanie języka naturalnego w systemach sztucznej inteligencji**

## **Projekt nr 3**

AGH Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii  
Biomedycznej

### **Autorzy**

Michał Burda - [michaburda@student.agh.edu.pl](mailto:michaburda@student.agh.edu.pl)

Radosław Barszczak - [rbarszczak@student.agh.edu.pl](mailto:rbarszczak@student.agh.edu.pl)





# 1. Opis Projektu

## 1. Zbieranie danych wejściowych

- Użytkownik jest proszony o podanie dwóch czasowników. Program sprawdza, czy podane czasowniki znajdują się w słowniku **verbs**, który zawiera podstawowe formy czasowników oraz ich różne odmiany (np. przeszłe i imiesłowy).

## 2. Przygotowanie tekstu

- Program przeszukuje wszystkie pliki tekstowe (.txt) znajdujące się w określonym folderze. Tekst z tych plików jest odczytywany i przygotowywany do analizy. Zawiera to usunięcie znaków interpunkcyjnych, co pozwala na skupienie się wyłącznie na słowach.

## 3. Wyszukiwanie słów po czasownikach

- Dla każdego z dwóch zadanych czasowników program przeszukuje tekst, aby znaleźć wszystkie słowa, które występują bezpośrednio po tych czasownikach.

## 4. Zbieranie wyników

- Program zbiera wszystkie unikalne słowa, które pojawiają się po każdym z dwóch czasowników, a następnie tworzy zbiór wyników.

## 5. Prezentacja wyników

- Po przetworzeniu tekstów program może wyświetlić wyniki w czterech sekcjach:
  - Rzeczowniki występujące po pierwszym czasowniku.
  - Rzeczowniki występujące po drugim czasowniku.
  - Suma wszystkich znalezionych rzeczowników.
  - Część wspólna rzeczowników po obu czasownikach.

## 6. Interfejs użytkownika

- Program działa w trybie tekstowym, gdzie użytkownik wprowadza dane przez terminal.

## 2. Lista czasowników

Infinitive	Past Simple	Past Participle
be	was/were	been
have	had	had
do	did	done
say	said	said
go	went	gone
get	got	gotten
make	made	made
know	knew	known
think	thought	thought
take	took	taken
see	saw	seen
come	came	come
want	wanted	wanted
look	looked	looked
use	used	used
find	found	found
give	gave	given
tell	told	told
work	worked	worked
call	called	called
try	tried	tried
ask	asked	asked
need	needed	needed
feel	felt	felt
become	became	become
leave	left	left

put	put	put
mean	meant	meant
keep	kept	kept
let	let	let
begin	began	begun
seem	seemed	seemed
help	helped	helped
talk	talked	talked
turn	turned	turned
start	started	started
show	showed	shown
hear	heard	heard
play	played	played
run	ran	run
move	moved	moved
like	liked	liked
live	lived	lived
believe	believed	believed
hold	held	held
bring	brought	brought
happen	happened	happened
write	wrote	written
provide	provided	provided
sit	sat	sat
stand	stood	stood
lose	lost	lost
pay	paid	paid
meet	met	met
include	included	included

continue	continued	continued
set	set	set
lear	learned/learnt	learned/learnt
change	changed	changed
lead	led	led
understand	understood	understood
watch	watched	watched
follow	followed	followed
eat	ate	eaten
create	created	created
speak	spoke	spoken
read	read	read
allow	allowed	allowed
add	added	added
spend	spent	spent
grow	grew	grown
open	opened	opened
walk	walked	walked
win	won	won
offer	offered	offered
remember	remembered	remembered
love	loved	loved
consider	considered	considered
appear	appeared	appeared
buy	bought	bought
wait	waited	waited
serve	served	served
die	died	died
send	sent	sent

expect	expected	expected
built	built	built
stay	stayed	stayed
fall	fell	fallen
cut	cut	cut
reach	reached	reached
kill	killed	killed
remain	remained	remained
suggest	suggested	suggested
raise	raised	raised
pass	passed	passed
sell	sold	sold
required	required	required
report	reported	reported
decide	decided	decided
drink	drank	drunk

### 3. Przykładowe działanie programu

#### Wyszukiwarka Rzeczowników

##### Rzeczowniki po czasowniku 1:

tries, day, water, other, blood, chocolate, first, something, notice, devil

##### Rzeczowniki po czasowniku 2:

day, eggs, south, one, fish, daily, fear, chocolate, food, animal, apple, man, stale, everything, yer, followers, grass, myself, peace, meat, amounts, bread, reduce, japan, shelter, turkey, them, meals, breakfast, far, oats, ice, corn

##### Suma rzeczowników:

day, eggs, oats, south, one, fish, daily, fear, chocolate, food, apple, corn, notice, tries, water, man, stale, something, everything, yer, followers, grass, myself, peace, meat, amounts, bread, reduce, japan, first, shelter, turkey, other, blood, them, meals, breakfast, far, animal, ice, devil

##### Część wspólna:

day, chocolate

#### Wyszukiwarka Rzeczowników

##### Rzeczowniki po czasowniku 1:

ginger, tis, spell, building, millions, wornout, course, pony, five, two, mr, john, wand

##### Rzeczowniki po czasowniku 2:

city, improved, cars, own, building, themselves, apartment, skyscrapers, yet, bones, low, nicolasjoseph

##### Suma rzeczowników:

tis, improved, millions, themselves, course, pony, skyscrapers, yet, bones, mr, low, nicolasjoseph, city, ginger, spell, building, cars, own, wornout, apartment, five, two, john, wand

##### Część wspólna:

building

## Wyszukiwarka Rzeczowników

### Rzeczowniki po czasowniku 1:

flute, knees, hair, yourself, trace, end, own

### Rzeczowniki po czasowniku 2:

john, wand, tis, two, wornout, mr, ginger, spell, five, millions, course, pony

### Suma rzeczowników:

john, flute, tis, two, wornout, ginger, yourself, end, wand, pony, mr, hair, spell, five, millions, knees, trace, course, own

### Część wspólna:

Brak wyników

## Wyszukiwarka Rzeczowników

### Rzeczowniki po czasowniku 1:

villagers, mind, hell, desk, puzzlement, decoration, one, dear, rules, food, subject, plan, maybe, hands, color, sleeping

### Rzeczowniki po czasowniku 2:

horse, time, unclean, one, sir, himself

### Suma rzeczowników:

villagers, desk, one, subject, plan, maybe, hands, himself, horse, mind, puzzlement, time, decoration, unclean, dear, rules, food, sir, hell, color, sleeping

### Część wspólna:

one

## Wyszukiwarka Rzeczowników

### Rzeczowniki po czasowniku 1:

snape, harrys, percy, birds, gorilla, norbert, owl, horror, cat, goblin, birthday, hagrid, malfoy, television, quirrell, girl, harry, bated, recognized, professor, unwrap, redhaired

### Rzeczowniki po czasowniku 2:

losing, little, brass, forth, hed, white, gryffindors, factories, scales, ourselves, last, cabs, upon, boa, myself, tis, drawing, time, traffic, stairs, against, home, child, affairs, glass, drivers, carriage

### Suma rzeczowników:

snape, harrys, losing, little, brass, percy, birds, forth, gorilla, hed, norbert, owl, white, gryffindors, factories, horror, cat, goblin, scales, birthday, hagrid, ourselves, last, malfoy, cabs, upon, glass, television, quirrell, boa, myself, tis, drawing, girl, harry, bated, time, traffic, recognized, stairs, against, home, child, professor, unwrap, affairs, drivers, carriage, redhaired

### Część wspólna:

Brak wyników

## Wyszukiwarka Rzeczowników

### Rzeczowniki po czasowniku 1:

class, cloak, yourself, two, gryffindor, foot, stone, admirable, receiver, nose, meal, wand, harry, inside, cat, heads, anything, outside, bed, saddle, hopefully, hand, another, checkrein, jerry, money, something, end, lord, captain, harrys, pointed, straw, one, plates, charms, arm, horses, bottle, photos, wife, rein, face, stop, hagrids, merrylegs, hedwig, straps, robes, head, slytherin, hermione

### Rzeczowniki po czasowniku 2:

dick, somethin, cant, miss, someone, snape, hagrid, pointing, other, harry, seamus, woods, snake, price, questions, dear, oh, urgently, aunt, ing, looking, spells, hed, ron, jerry, madam, sticking, something, leading, shrilly, once, one, saw, wood, percy, question, fred, anymore, dumbledore, servant, conductor, hermione, man, trying, mr, hoping, hundreds, parents

### Suma rzeczowników:

dick, cloak, yourself, foot, stone, receiver, nose, wand, other, harry, seamus, heads, snake, price, questions, urgently, bed, aunt, hopefully, looking, another, spells, hed, something, jerry, ron, sticking, shrilly, pointed, one, plates, charms, arm, horses, bottle, photos, wife, rein, face, question, anymore, hagrids, servant, hedwig, man, head, trying, hoping, hundreds, hermione, class, somethin, cant, miss, someone, snape, two, hagrid, gryffindor, pointing, admirable, meal, inside, woods, cat, anything, outside, dear, oh, saddle, ing, hand, checkrein, madam, money, leading, end, lord, captain, harrys, straw, once, saw, wood, percy, fred, dumbledore, stop, merrylegs, conductor, straps, robes, slytherin, mr, parents

### Część wspólna:

one, harry, jerry, something, hermione



# Wyszukiwarka Rzeczowników

understand

fall

Szukaj

## Rzeczowniki po czasowniku 1:

secrets, importance, hell, harry, muggle, horses, professor, snape, jerrys, beauty, business, course, other, later, agree

## Rzeczowniki po czasowniku 2:

flinging, marcus, harry, quirrells, asleep, halloo, leaves, violence, malfoy, reuben, whether, blantyre, flat, promised, owing, trading, loose, backward, onto, bed, gripped, books, seemed, blackness, ahern, leaned, thirtysix, has, quirrell, upon, although, mortally

## Suma rzeczowników:

secrets, flinging, marcus, harry, quirrells, asleep, course, halloo, leaves, violence, malfoy, reuben, importance, snape, whether, professor, blantyre, flat, promised, owing, trading, loose, later, backward, onto, agree, bed, gripped, books, seemed, blackness, horses, business, ahern, beauty, leaned, thirtysix, hell, has, muggle, quirrell, jerrys, other, although, upon, mortally

## Część wspólna:

harry

## Źródła

Harry Potter:

<https://docenti.unimc.it/antonella.pascali/teaching/2018/19055/files/ultima-lezione/harry-potter-and-the-philosophers-stone>

Black Beauty:

<https://www.defence.lk/upload/ebooks/Black%20Beauty.pdf>

Bajki dla dzieci:

<https://andonovicmilica.wordpress.com/wp-content/uploads/2018/10/english-fairy-tales.pdf>

Food Article:

<https://education.nationalgeographic.org/resource/food/>

Car Article:

<https://kids.britannica.com/kids/article/car/352807>

City Article:

<https://www.english-online.at/geography/cities/cities-and-metropolitan-areas.htm>