

# Detektor języków

Michał Kowalski



HR EXCELLENCE IN RESEARCH



Politechnika Wroclawska

# Opis problemu

## Zastosowania

- Rozpoznanie, w jakim języku jest napisany dany tekst,
- Porównanie stopnia podobieństwa języków,
- Wykorzystanie sieci neuronowej w celu klasyfikacji

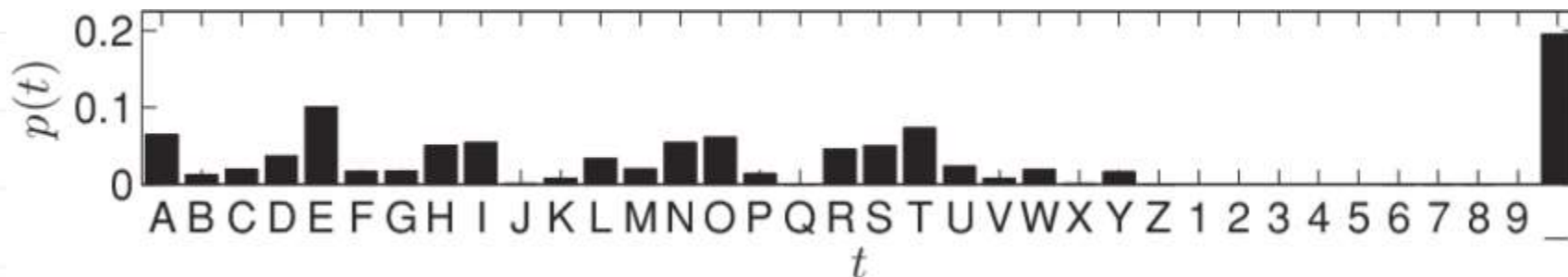
## Trudności

- Wiele języków ma część znaków spoza ASCII,
- Niektóre języki mają alfabetu spoza ASCII,
- Długi czas parsowania plików

# Rozwiązanie najprostsze

Badanie prawdopodobieństwa wystąpienia danej litery

- + Duża szybkość parsowania,
- Mała dokładność



# Rozwiązanie najdokładniejsze

Badanie prawdopodobieństwa wystąpienia ciągu 3-literowego

- + Duża dokładność detekcji,
- Bardzo powolne parsowanie  
(złożoność  $n^3$ )



# Rozwiązanie pośrednie

Badanie prawdopodobieństwa wystąpienia ciągu 2-literowego

- + Wystarczająca dokładność,
- + Złożoność  $n^2$



# Napisany program

## Funkcjonalności

Detektor języków

Uczenie sieci

Dzień dobry!

Detekcja języka wpisanego tekstu

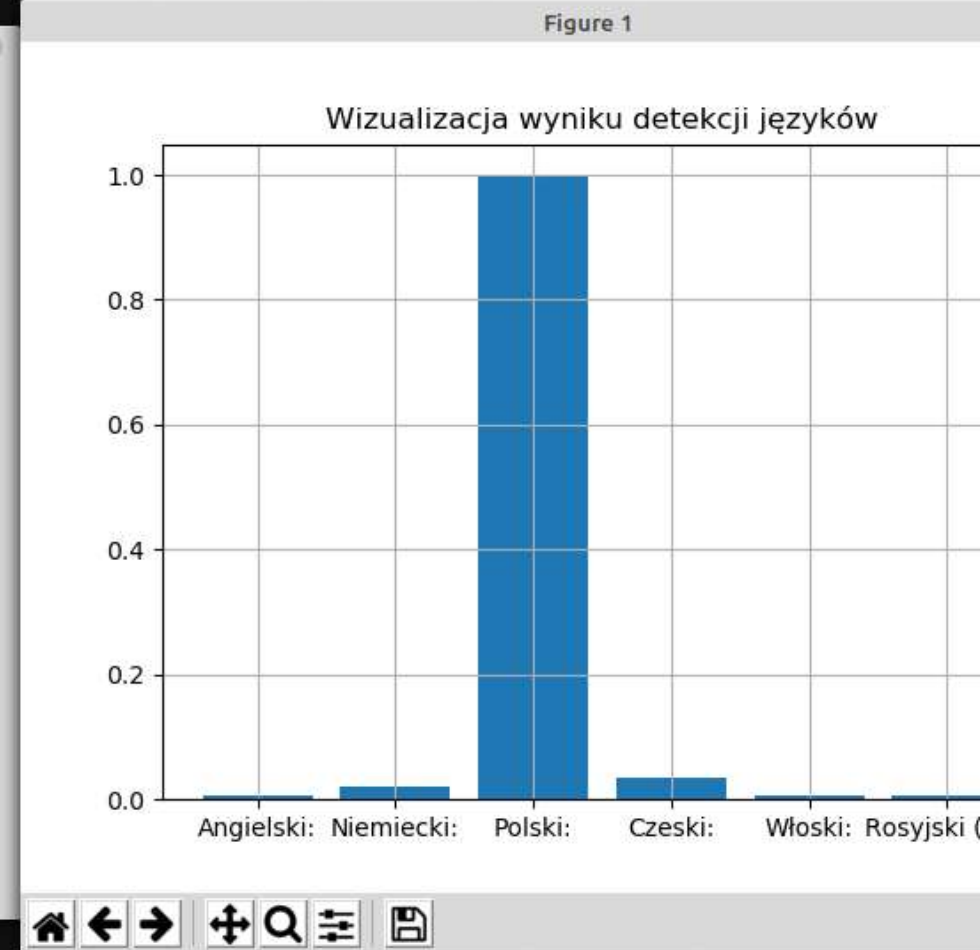
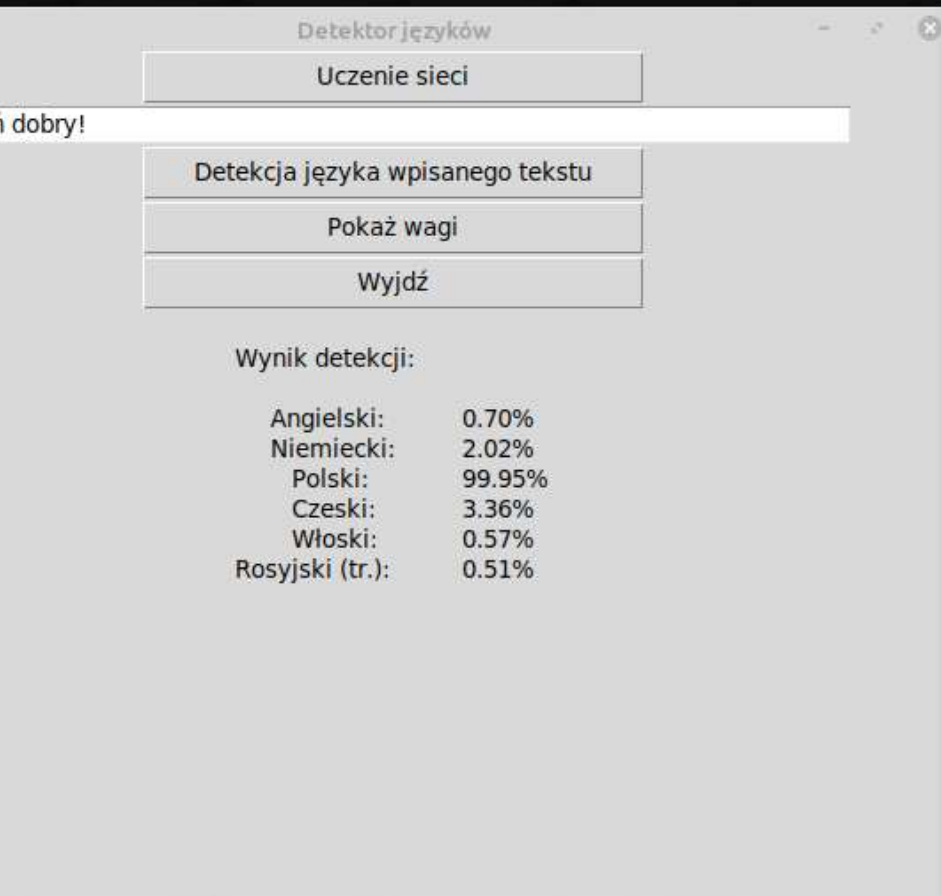
Pokaż wagi

Wyjdź

Wynik detekcji:

Angielski:	0.70%
Niemiecki:	2.02%
Polski:	99.95%
Czeski:	3.36%
Włoski:	0.57%
Rosyjski (tr.):	0.51%

- Parsowanie plików oraz uczenie sieci,
- Detekcja języka z wpisanego tekstu,
- Wizualizacja danych na wykresie



**Dziękuję za uwagę**  
Czy mają Państwo jakieś pytania?

# Źródła:

Sieci neuronowe:

<https://impicode.pl/blog/jak-stworzyc-siec-neuronowa/>

<https://www.blindtextgenerator.com/>

Python:

<https://docs.python.org/3/library/tkinter.html>

<https://matplotlib.org/stable/tutorials/introductory/pyplot.html>

Grafiki prezentacji:

[https://www.researchgate.net/profile/J-](https://www.researchgate.net/profile/J-Farquhar/publication/47518347/figure/fig2/AS:667872292503553@1536244571061/Probability-of-)

[Farquhar/publication/47518347/figure/fig2/AS:667872292503553@1536244571061/Probability-of-](https://www.researchgate.net/profile/J-Farquhar/publication/47518347/figure/fig2/AS:667872292503553@1536244571061/Probability-of-)  
[characters-in-English-The-SPACE-character-represented-by-has-the.png](https://www.researchgate.net/profile/J-Farquhar/publication/47518347/figure/fig2/AS:667872292503553@1536244571061/Probability-of-)

<https://wordart.com/>