

**WYDZIAŁ**

**ELEKTROTECHNIKI, AUTOMATYKI,**

**INFORMATYKI I INŻYNIERII BIOMEDYCZNEJ**

Projekt dyplomowy

*Przewidywanie metryk psychologicznych tekstu za pomocą przetwarzania języka naturalnego*

*Predicting of psychological metrics of a text using natural language processing*

Autor: *Hubert Paweł Stępniewski*

Kierunek studiów: Automatyka i Robotyka

Opiekun pracy: dr hab. inż. Jerzy Baranowski

Kraków, 2022

[1. Wstęp 2](#_Toc89259430)

[2. Przetwarzanie języka naturalnego 2](#_Toc89259431)

[2.1. Trudności w przetwarzaniu języka naturalnego 3](#_Toc89259432)

[2.2. Podejście symboliczne, statystyczne i neuronowe 3](#_Toc89259433)

[2.3. Główne zastosowania NLP 4](#_Toc89259434)

[3. Problematyka rozpoznawania emocji w tekście 4](#_Toc89259435)

[3.1. Teorie Emocji 5](#_Toc89259436)

[3.2. Stosowane podejścia 5](#_Toc89259437)

[4. Projekt 5](#_Toc89259438)

[4.1. Koncepcja rozwiązania 5](#_Toc89259439)

[4.2. Zebranie danych 5](#_Toc89259440)

[4.3. Algorytm NLP 5](#_Toc89259441)

[4.4. Aplikacja webowa 5](#_Toc89259442)

[5. Implementacja 5](#_Toc89259443)

[5.1. Aplikacja do zbierania danych 5](#_Toc89259444)

[5.2. Algorytm NLP 5](#_Toc89259445)

[5.3. Aplikacja webowa 5](#_Toc89259446)

[5.4. Testy 5](#_Toc89259447)

[6. Analiza działania algorytmu 5](#_Toc89259448)

[6.1. Opis badań 5](#_Toc89259449)

[6.2. Stosowane miary 5](#_Toc89259450)

[6.3. Wyniki badań 5](#_Toc89259451)

[6.4. Analiza wyników 5](#_Toc89259452)

[7. Podsumowanie 5](#_Toc89259453)

[8. Literatura 5](#_Toc89259454)

# Wstęp

# Przetwarzanie języka naturalnego

Przetwarzanie języka naturalnego (ang. n*atural language processing, NLP)* to dziedzina zajmująca się automatycznym rozumieniem, analizowaniem i generowaniem tekstu przez komputer. Język naturalny jest rozumiany jako język stosowany przez ludzi do komunikacji, na tej podstawie odróżniamy go od języków formalnych np. języków programowania[1][[1]](#endnote-1). NLP stosowane jest w wielu zadaniach. Obejmuje zarówno proste zliczanie poszczególnych wyrazów jak i zaawansowane boty potrafiące wygenerować sensowną odpowiedź na pytanie użytkownika.

## Trudności w przetwarzaniu języka naturalnego

Ludzki układ nerwowy jest naturalnie przystosowany do przetwarzanie języka naturalnego. Jest to jedna z pierwszych umiejętności, którą nabywamy i błyskawicznie automatyzujemy[[2]](#endnote-2)[3]. Oznacza to, iż w większości przypadków analiza języka jest przeprowadzana w sposób nieuświadomiony. Nauczenie tego samego komputera okazuje się zadaniem niezwykle skomplikowanym. Wynika to z natury języka, który jest nieprecyzyjny i niejednoznaczny. Jest ona widoczna na wielu poziomach:

Składniowym - relacje słów w zdaniu można określić na wiele sposobów.

Semantycznym – jedno słowo może mieć wiele znaczeń, które zależą od kontekstu np. zamek, damy.

Pragmatycznym – istnienie ironii, metafor itd.

Te cechy języka powodują, że problemem jest opisanie go sztywnymi regułami. Ponadto do jego zrozumienia potrzebna jest znajomość kontekstu oraz ogólna wiedza o otaczającym świecie.

## Podejście symboliczne, statystyczne i neuronowe

Wyróżnia się trzy podejścia do przetwarzania języka naturalnego. Historycznie pierwsze jest podejście symboliczne. Polega ono na korzystaniu z bazy sztywnych reguł i faktów dotyczących języka. Wymaga ono eksperta, który taką bazę stworzy. Na podstawie tej bazy komputer analizuje język. Podejście to jest obecnie rzadko stosowane, gdyż zasady rządzące językami naturalnymi są zbyt skomplikowane by dało się je opisać jako zbiór logicznych reguł, często występują w nich też wyjątki. Ponadto samo tworzenie bazy reguł jest pracochłonne i bardzo podatne na błędy[2].[[3]](#endnote-3) Podejście statystyczne wykorzystuje uczenie maszynowe aby wytworzyć bazę wiedzy na temat języka na podstawie dużego korpusu tekstów, ocenionego wcześniej pod kątem potrzebnych metryk. Baza zasad jest w tym podejściu tworzona automatycznie. Korzystając z posiadanych danych konstruowane są modele statystyczne, odpowiadające zależnościom językowym. Człowiek odpowiada jedynie za dostarczenie odpowiednio ocenionego korpusu tekstów[2]. Rozwój nauki o przetwarzaniu danych (ang. *data science*) sprawił, iż podejście statystyczne jest obecnie często stosowane. Co więcej możliwości obliczeniowe współczesnych komputerów, pozwalają na analizowanie olbrzymich ilości danych. Jest to konieczne w omawianym zastosowaniu, a także w najmłodszym podejściu. Jest nim podejście neuronowe, polegające na uczeniu głębokich sieci neuronowych przetwarzania języka naturalnego. Obecnie odnosi ono duże sukcesy i jest bardzo intensywnie rozwijane[[4]](#endnote-4)[3].

## Główne zastosowania NLP

Przetwarzanie języka naturalnego jest obecnie szeroko stosowane. Praktycznie codziennie spotykamy się z jego zastosowaniami, kiedy korzystamy z Internetu. Do najczęściej spotykanych należą:

* Tłumaczenie maszynowe (machine translation) – automatyczne tłumaczenie tekstu z jednego języka na drugi.
* Rozpoznawanie tematów (topic modeling) – podzielenie tekstu na segmenty, które są poświęcone różnym tematom.
* Automatyczne podsumowanie (automatioc summarization) - wytwarzanie streszczenia podanego tekstu, zawierającego najistotniejsze informacje.
* Korekta tekstu - dotyczy szerokiego zakresu problemów. Od prostych błędów ortograficznych, przez gramatykę po pragmatykę.
* Generowanie języka naturalnego (Natural Language Generation) - przedstawienie informacji z komputerowych baz danych w czytelny dla ludzi sposób.
* Odpowiadanie na pytania (question answering) – generowanie odpowiedzi na zadane pytanie. Dotyczy zarówno pytań zamkniętych jak i otwartych.
* Rozumienie języka naturalnego (Natural Language Understanding) - przekonwertowanie języka naturalnego na formalne struktury logiczne.
* Analiza sentymentów (sentyment analysis) – określenie na podstawie tekstu nastawienia autora i wydźwięku wypowiedzi. Przede wszystkim określenie czy była ona pozytywna czy negatywna

# Problematyka rozpoznawania emocji w tekście

Naturalnym rozwinięciem analizy sentymentu, jest analiza emocji występujących w tekście. Jest to stosunkowo nowe zastosowanie NLP. Ze względu na przedmiot analizy jest to zadanie skomplikowane. Emocje są naturalną częścią funkcjonowania poznawczego ludzi. Wydaje się, iż wiemy o nich dużo. Jednak naukowe ich zdefiniowanie powoduje problemy wśród psychologów.

## Teorie Emocji

W kontekście rozpoznawania emocji w tekście kluczową sprawą wydaje się być określenie ile jest emocji oraz jakie to emocje. Okazuje się, że nawet w tej kwestii nie ma wśród psychologów zgody. Najczęściej pojawia się jednak odpowiedź, iż jest ich sześć. Wymieniane są: strach, radość, wstręt, złość, smutek i zdziwienie. Jest to pokłosie słynnych badań Ekmana, który poszukiwał emocji uniwersalnych – obecnych niezależnie od kontekstu społecznego. Przebadał on osoby z różnych kultur i na podstawie ekspresji mimicznych, wyróżnił powyższą listę. Zgodnie z jego badaniami są to emocje występujące u wszystkich ludzi. Choć jest to najpopularniejsze z koncepcji, badania Ekmana są krytykowane, a jego wnioski poddawane w wątpliwość przez część naukowców.

Ciekawą

## Stosowane podejścia

# Projekt

## Koncepcja rozwiązania

## Zebranie danych

## Algorytm NLP

## Aplikacja webowa

# Implementacja

## Aplikacja do zbierania danych

## Algorytm NLP

## Aplikacja webowa

## Testy

# Analiza działania algorytmu

## Opis badań

## Stosowane miary

## Wyniki badań

## Analiza wyników

# Podsumowanie

# Literatura

1. *John Lyons: Natural Language and Universal Grammar. Nowy Jork: Cambridge University Press, 1991, s. 68–70.* [*ISBN 978-0521246965*](https://pl.wikipedia.org/wiki/Specjalna:Ksi%C4%85%C5%BCki/9780521246965)*. (ang.)* [↑](#endnote-ref-1)
2. Zimbardo [↑](#endnote-ref-2)
3. Maciej Pienkosz praca magisterska (2)magisterska (1).pdf [↑](#endnote-ref-3)
4. How Transferable are Neural Networks in NLP Applications? Lili Mou,1 Zhao Meng,1 Rui Yan,2 Ge Li,1,† Yan Xu,1,∗ Lu Zhang,1 Zhi Jin1,† 1Key Laboratory of High Confidence Software Technologies (Peking University), MoE, China Institute of Software, Peking University, China †Corresponding authors 2 Insitute of Computer Science and Technology of Peking University, China [↑](#endnote-ref-4)