

# Projekt PJN – część praktyczna

## Spis treści

<b>0. Tytuł, autorzy, abstrakt .....</b>	<b>2</b>
<b>1. Wstęp .....</b>	<b>2</b>
<b>2. Rozwinięcie.....</b>	<b>3</b>
<b>3. Podsumowanie .....</b>	<b>5</b>
<b>4. Bibliografia.....</b>	<b>6</b>

## 0. Tytuł, autorzy, abstrakt

**Tytuł:** Narzędzie do tworzenia podsumowań tekstu opierające się na text\_summarizer [1].

**Autorzy:**

- KK
- MG
- MK

**Abstrakt:**

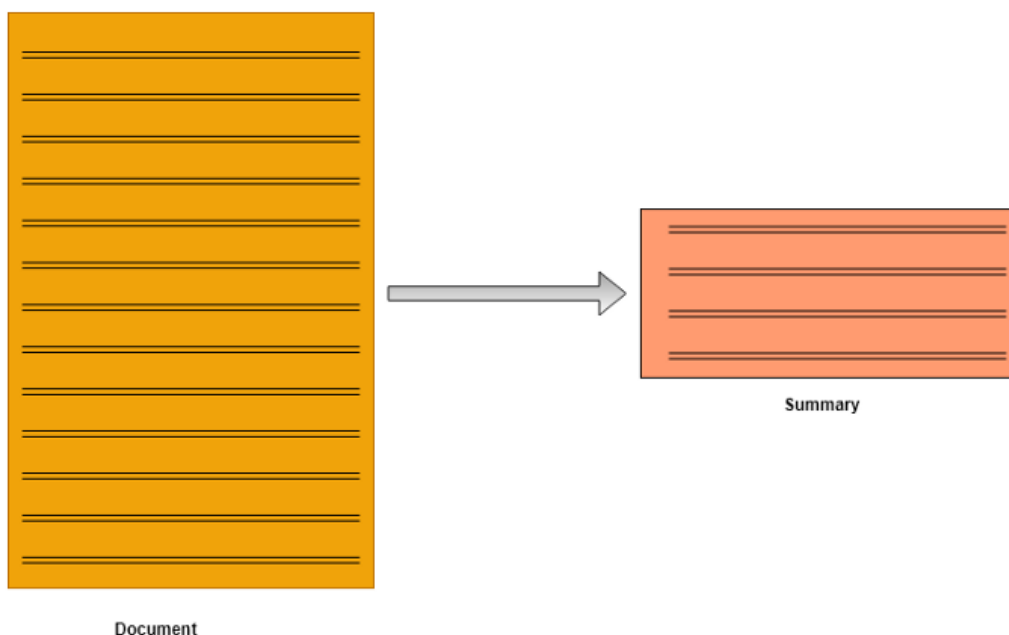
W tym raporcie przybliżyliśmy temat sumaryzacji, czyli podsumowań tekstu. Przedstawiliśmy schemat działania oraz zaimplementowaliśmy program tworzący sumaryzację tekstu opierający się na text\_summarizer [1]. Na końcu pokazaliśmy przykład działania i efekt końcowy.

## 1. Wstęp

Sumaryzacja może być zdefiniowana jako zadanie polegające na stworzeniu zwięzłego i płynnego streszczenia przy zachowaniu kluczowych informacji i ogólnego sensu. [2]

Często spotykamy się z sumaryzacją tekstu. Przydaje się ona, gdy chcemy pojąć jakąś informację nie musząc czytać całego wielostronnego artykułu. Na przykład w sporcie – możemy dowiedzieć się o głównych zdarzeniach meczu bez wchodzenia w szczegóły przebiegu spotkania.

Na rysunku poniżej przedstawiono rysunek schematyczny działania programu.



Z dużej ilości tekstu wyciągane są najistotniejsze informacje, dla których czytamy dany artykuł. Dostajemy dzięki temu krótszy tekst, który jesteśmy w stanie o wiele szybciej przeczytać i lepiej przyswoić jego treść.

Wyróżnić można dwa rodzaje sumaryzacji:

1. **Sumaryzacja abstrakcyjna** – wybiera słowa oparte na rozumieniu semantycznym, często wybierając takie, których nie ma w tekście. Jest to rodzaj sumaryzacji podobny do pisania streszczenia przez człowieka.
2. **Sumaryzacja ekstrakcyjna** – wybiera zbiór słów z tekstu, które zachowują najważniejszą treść.

W kolejnym rozdziale prezentujemy działanie programu wykorzystującego **sumaryzację ekstrakcyjną**.

## 2. Rozwinięcie

Implementacja i analiza kodu [2]:

1. Import bibliotek:

```
from nltk.corpus import stopwords
from nltk.cluster.util import cosine_distance
import numpy as np
import networkx as nx
```

- NLTK corpus – moduły w tym pakiecie dostarczają funkcje, które mogą być wykorzystane do odczytu plików korpusowych w różnych formatach.
- NLTK cluster – moduł ten zawiera szereg podstawowych algorytmów klastrowania. Klastrowanie opisuje zadanie odkrywania grup podobnych elementów z dużej kolekcji.
- „Numpy, czyli Numerical Python – podstawowy zestaw narzędzi, które umożliwiają zaawansowane obliczenia matematyczne na macierzach. Numpy to biblioteka open source, która zapewnia programistom wszechstronność w pracy z macierzami i wielowymiarowymi tablicami.” [3]
- NetworkX to biblioteka Pythona do tworzenia, manipulacji i badania struktury, dynamiki i funkcji złożonych sieci.

## 2. Funkcja do generowania czystych zdań:

```
def read_article(file_name):
    file = open(file_name, "r")
    filedata = file.readlines()
    article = filedata[0].split(". ")
    sentences = []

    for sentence in article:
        print(sentence)
        sentences.append(sentence.replace("[^a-zA-Z]", " ").split(" "))
        sentences.pop()

    return sentences
```

## 3. Macierz podobieństwa.

W tym miejscu będziemy używać podobieństwa cosinusowego, aby znaleźć podobieństwo między zdaniami.

```
def build_similarity_matrix(sentences, stop_words):
    # Stworzenie pustej macierzy podobieństwa
    similarity_matrix = np.zeros((len(sentences), len(sentences)))

    for idx1 in range(len(sentences)):
        for idx2 in range(len(sentences)):
            if idx1 == idx2: # jeśli oba zdania są takie same - są ignorowane
                continue
            similarity_matrix[idx1][idx2] = sentence_similarity(sentences[idx1], sentences[idx2], stop_words)

    return similarity_matrix
```

## 4. Generowanie sumaryzacji.

```
def generate_summary(file_name, top_n=5):
    nltk.download("stopwords")
    stop_words = stopwords.words('english')
    summarize_text = []

    # Krok 1 - Odczytanie tekstu i podzielenie
    sentences = read_article(file_name)

    # Krok 2 - Generowanie macierzy podobieństwa poprzez zdania
    sentence_similarity_matrix = build_similarity_matrix(sentences, stop_words)

    # Krok 3 - Uporządkowanie zdań w macierzy podobieństwa
    sentence_similarity_graph = nx.from_numpy_array(sentence_similarity_matrix)
    scores = nx.pagerank(sentence_similarity_graph)

    # Krok 4 - Sortowanie uporządkowania i wybranie najlepszych zdań
    ranked_sentence = sorted(((scores[i],s) for i,s in enumerate(sentences)), reverse=True)
    print("Indexes of top ranked_sentence order are ", ranked_sentence)

    for i in range(top_n):
        summarize_text.append(" ".join(ranked_sentence[i][1]))

    # Krok 5 - Wypisanie tekstu
    print("\nSummarize Text: \n", " ".join(summarize_text))
```

### 3. Podsumowanie

#### Przykładowy artykuł [4]:

Dec 19 (Reuters) - Former players came together in acclaiming the World Cup final in Qatar as one of the best ever after Argentina beat France on penalties following a pulsating 3-3 draw. Lionel Messi scored twice and Kylian Mbappe grabbed a hat-trick in 120 minutes of breathtaking action, before Gonzalo Montiel sent Hugo Lloris the wrong way to seal a shootout win and a third World Cup title for Argentina. Former England and Liverpool defender Jamie Carragher said it was one of the "best games of all time". "This game showed why football is the greatest game of all. This one is an all-time classic," he tweeted. Alan Shearer told the BBC "it was just an unbelievable final". "I've never seen anything like it and I don't think I'll ever see anything like it again. It was staggering," the former England striker said. Messi, in his 26th World Cup match at his fifth tournament, was finally able to get his hands on the elusive trophy and match his predecessor Diego Maradona. "For our country, we will see a picture of Maradona and Messi, two of the greatest players of all time with the World Cup trophy," former Argentina defender Pablo Zabaleta told the BBC. "That is something so incredible. We are so lucky." Gary Neville said Messi deserved to be the centre of attention. "During his whole career, he's lit up every game he's played in. If you've watched Lionel Messi play live you are truly blessed," the ex-England international Neville told ITV. Neville added that Argentina's win would inspire a new generation of players, pointing to his own experience of watching Maradona play at the 1986 World Cup. "We were watching the clips of Maradona back in '86 before the game and I was young watching it in my childhood," Neville added. "The kids who were my age back then will be watching this tonight thinking they want to become a footballer and become engrossed and embedded in the culture of football."

#### Wypisanie tekstu przez program:

Dec 19 (Reuters) - Former players came together in acclaiming the World Cup final in Qatar as one of the best ever after Argentina beat France on penalties following a pulsating 3-3 draw  
Lionel Messi scored twice and Kylian Mbappe grabbed a hat-trick in 120 minutes of breathtaking action, before Gonzalo Montiel sent Hugo Lloris the wrong way to seal a shootout win and a third World Cup title for Argentina  
Former England and Liverpool defender Jamie Carragher said it was one of the "best games of all time"  
"This game showed why football is the greatest game of all  
This one is an all-time classic," he tweeted  
Alan Shearer told the BBC "it was just an unbelievable final"  
"I've never seen anything like it and I don't think I'll ever see anything like it again  
It was staggering," the former England striker said  
Messi, in his 26th World Cup match at his fifth tournament, was finally able to get his hands on the elusive trophy and match his predecessor Diego Maradona  
"For our country, we will see a picture of Maradona and Messi, two of the greatest players of all time with the World Cup trophy," former Argentina defender Pablo Zabaleta told the BBC  
"That is something so incredible  
We are so lucky." Gary Neville said Messi deserved to be the centre of attention  
"During his whole career, he's lit up every game he's played in  
If you've watched Lionel Messi play live you are truly blessed," the ex-England international Neville told ITV  
Neville added that Argentina's win would inspire a new generation of players, pointing to his own experience of watching Maradona play at the 1986 World Cup  
"We were watching the clips of Maradona back in '86 before the game and I was young watching it in my childhood," Neville added  
"The kids who were my age back then will be watching this tonight thinking they want to become a footballer and become engrossed and embedded in the culture of football."

#### Uporządkowanie:

Indexes of top ranked\_sentence order are [(0.1297746657959731, [{"For", "our", "country", "we", "will", "see", "a", "picture", "of", "Maradona", "and", "Messi", "two", "of", "the", "greatest", "players", "of", "all", "time", "with", "the", "World", "Cup", "trophy", "former", "Argentina", "defender", "Pablo", "Zabaleta", "told", "the", "BBC"}]), (0.10062208972236078, [{"Neville", "added", "that", "Argentina's", "win", "would", "inspire", "a", "new", "generation", "of", "players", "pointing", "to", "his", "own", "experience", "of", "watching", "Maradona", "play", "at", "the", "1986", "World", "Cup"}]), (0.08790429785748931, [{"Dec", "19", "(Reuters)", "-", "Former", "players", "came", "together", "in", "acclaiming", "the", "World", "Cup", "final", "in", "Qatar", "as", "one", "of", "the", "best", "ever", "after", "Argentina", "beat", "France", "on", "penalties", "following", "a", "pulsating", "3-3", "draw"}]), (0.08679170353950438, [{"Former", "England", "and", "Liverpool", "defender", "Jamie", "Carragher", "said", "it", "was", "one", "of", "the", "best", "games", "of", "all", "time"}]), (0.08506296057702986, [{"We", "were", "watching", "the", "clips", "of", "Maradona", "back", "in", "'86", "before", "the", "game", "and", "I", "was", "young", "watching", "it", "in", "my", "childhood", "Neville", "added"}]), (0.07044362113962892, [{"It", "was", "staggering", "the", "former", "England", "striker", "said"}]), (0.06810107835018675, [{"If", "you've", "watched", "Lionel", "Messi", "play", "live", "you", "are", "truly", "blessed", "the", "ex-England", "international", "Neville", "told", "ITV"}]), (0.0675110439432148, [{"We", "are", "so", "lucky", "Gary", "Neville", "said", "Messi", "deserved", "to", "be", "the", "centre", "of", "attention"}]), (0.062276166758934284, [{"Lionel", "Messi", "scored", "twice", "and", "Kylian", "Mbappe", "grabbed", "a", "hat-trick", "in", "120", "minutes", "of", "breathtaking", "action", "before", "Gonzalo", "Montiel", "sent", "Hugo", "Lloris", "the", "wrong", "way", "to", "seal", "a", "shootout", "win", "and", "a", "third", "World", "Cup", "title", "for", "Argentina"}]), (0.05879862705651488, [{"Messi", "in", "his", "26th", "World", "Cup", "match", "at", "his", "fifth", "tournament", "was", "finally", "able", "to", "get", "his", "hands", "on", "the", "elusive", "trophy", "and", "match", "his", "predecessor", "Diego", "Maradona"}]), (0.05438753932396648, [{"This", "game", "showed", "why", "football", "is", "the", "greatest", "game", "of", "all"}]), (0.03695604697948592, [{"During", "his", "whole", "career", "he's", "lit", "up", "every", "game", "he's", "played", "in"}]), (0.03225832274698799, [{"Alan", "Shearer", "told", "the", "BBC", "it", "was", "just", "an", "unbelievable", "final"}]), (0.030391222765315758, [{"This", "one", "is", "an", "all-time", "classic", "he", "tweeted"}]), (0.018819623344397074, [{"I've", "never", "seen", "anything", "like", "it", "and", "I", "don't", "think", "I'll", "ever", "see", "anything", "like", "it", "again"}]), (0.00990099009009009, [{"That", "is", "something", "so", "incredible"}])]

#### Zsumaryzowany tekst:

Summarized text:  
"For our country, we will see a picture of Maradona and Messi, two of the greatest players of all time with the World Cup trophy," former Argentina defender Pablo Zabaleta told the BBC. Neville added that Argentina's win would inspire a new generation of players, pointing to his own experience of watching Maradona play at the 1986 World Cup

## Drugi przykładowy artykuł [1]:

For years, Facebook gave some of the world's largest technology companies more intrusive access to users' personal data than it has disclosed, effectively exempting those business partners from its usual privacy rules, according to internal records and interviews. The special arrangements are detailed in hundreds of pages of Facebook documents obtained by The New York Times. The records, generated in 2017 by the company's internal system for tracking partnerships, provide the most complete picture yet of the social network's data-sharing practices. They also underscore how personal data has become the most prized commodity of the digital age, traded on a vast scale by some of the most powerful companies in Silicon Valley and beyond. The exchange was intended to benefit everyone. Pushing for explosive growth, Facebook got more users, lifting its advertising revenue. Partner companies acquired features to make their products more attractive. Facebook users connected with friends across different devices and websites. But Facebook also assumed extraordinary power over the personal information of its 2 billion users - control it has wielded with little transparency or outside oversight. Facebook allowed Microsoft's Bing search engine to see the names of virtually all Facebook user's friends without consent, the records show, and gave Netflix and Spotify the ability to read Facebook users' private messages.

## Zsumaryzowany tekst:

Summarized text:

For years, Facebook gave some of the world's largest technology companies more intrusive access to users' personal data than it has disclosed, effectively exempting those business partners from its usual privacy rules, according to internal records and interviews. Facebook users connected with friends across different devices and websites

## 4. Bibliografia

[1] <https://github.com/edubey/text-summarizer>

[2] <https://towardsdatascience.com/understand-text-summarization-and-create-your-own-summarizer-in-python-b26a9f09fc70>

[3] <https://www.datasciencerobie.pl/frameworki-data-science-w-open-source/>

[4] <https://www.reuters.com/lifestyle/sports/all-time-classic-former-players-blown-away-by-thrilling-world-cup-final-2022-12-19/>