PRZETWARZANIE JĘZYKA NATURALNEGO

Temat projektu: ***Text summarizer w Pythonie***

Dominika Borowska, Aleksandra Brymerska

Kraków, 23.05.2022

Cel projektu

Celem projektu było stworzenie narzędzia do streszczania tekstu z artykułów dostępnych w Internecie. Narzędzie to powinno przyjmować tekst, który następnie powinien być streszczonych w określonej ilości zdań – w efekcie czego uzyskujemy podsumowanie całego artykułu.

Zakres pracy

W zakres pracy wchodziło określenie wymagań potrzebnych do zrealizowania projektu. Pierwszym krokiem było zebranie zbioru danych potrzebnego do wytrenowania modelu metodą web scrapping'u. Drugim krokiem było przygotowanie krótkiego programu, który podsumowuje tekst nie znając już wyjściowego wyniku, czyli wykorzystanie uczenia nienadzorowanego. Ostatnim i głównym punktem projektu było stworzenie modelu korzystającego z uczenia nadzorowanego oraz transformatorów, który na wejściu przyjmuje dwie kolumny – tekst artykułu oraz nagłówek i po wytrenowaniu potrafi już streścić jakikolwiek inny tekst podany na wejściu (nie znając odpowiedzi).

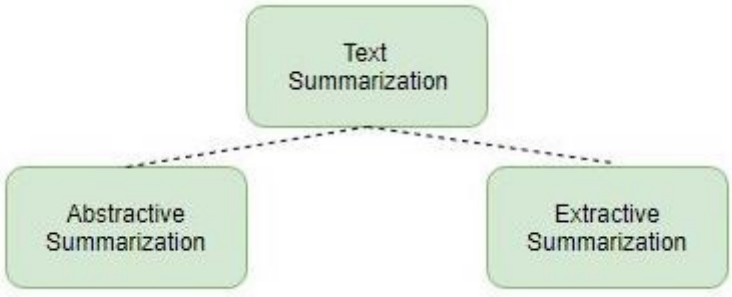
Wstęp teoretyczny

Text summarization (streszczenie tekstu) w NLP to proces tworzenia streszczeń z dużych ilości danych przy zachowaniu istotnych elementów informacyjnych i wartości treści. Język streszczenia powinien być zwięzły i prosty, tak aby przekazywał czytelnikowi znaczenie. Według Statista do 2025 r. łączna ilość danych tworzonych, rejestrowanych, kopiowanych i zużywanych na całym świecie ma przekroczyć 180 zettabajtów. Większość tych danych należy zminimalizować do prostszych, zwięzłych podsumowań zawierających istotne szczegóły, aby łatwiej je przeglądać i analizować. Istnieje duże zapotrzebowanie na algorytmy uczenia maszynowego, które mogą szybko podsumowywać długie artykuły i dostarczać dokładnych informacji. Właśnie tam pojawia się podsumowanie tekstu.

Podsumowanie tekstu jest korzystne w przypadku kilku zadań NLP, w tym klasyfikacji tekstu, streszczeń tekstów prawnych, streszczeń wiadomości, generowania nagłówków itp.

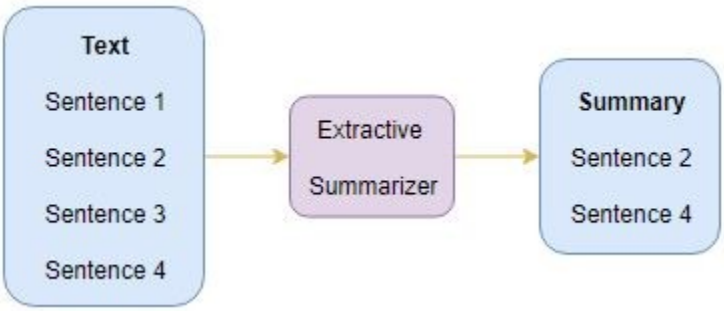
Podejścia do streszczania tekstu

Istnieją dwa główne podejścia do podsumowywania tekstu – ekstrakcyjne i abstrakcyjne.



Podsumowanie ekstrakcyjne

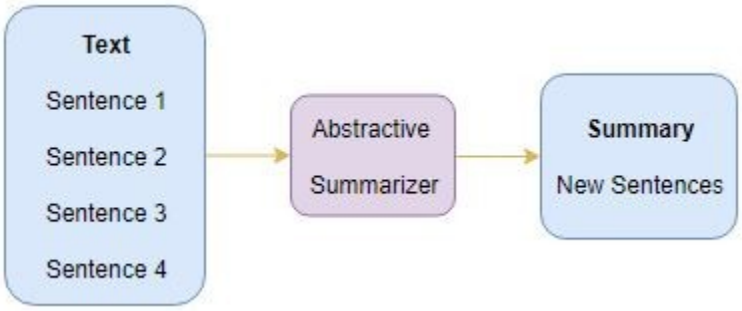
Podsumowywanie ekstrakcyjne polega na wyodrębnianiu podstawowych słów z oryginalnego dokumentu i łączeniu ich w celu stworzenia podsumowania. Wyodrębnianie podsumowań wykorzystuje mechanizm punktacji, aby uszeregować trafność fraz w celu wybrania tylko tych, które są najbardziej odpowiednie dla znaczenia dokumentu źródłowego. Ta technika wyodrębnia wymagany tekst zgodnie z określonymi kryteriami bez wprowadzania znaczących zmian w dokumentach. Ekstrakcyjna metoda podsumowania działa za pomocą algorytmów takich jak LexRank, Luhn, LSA itp., które są zaimplementowane przy użyciu bibliotek Pythona - Gensim lub Sumy.



Podsumowanie abstrakcyjne

Podsumowanie to koncentruje się na najważniejszych informacjach w tekście oryginalnym i tworzy nowy zestaw zdań do podsumowania. Nowe zdanie może nie być częścią tekstu źródłowego. Podejście to jest całkowitym przeciwieństwem podejścia ekstrakcyjnego, ponieważ to ostatnie generuje podsumowanie na podstawie dokładnego tekstu oryginalnego. Technika ta obejmuje identyfikację kluczowych elementów, interpretację kontekstu i odtworzenie ich w nowy sposób. Gwarantuje podanie najistotniejszych informacji w możliwie najkrótszych słowach. Abstrakcyjna metoda podsumowania działa dobrze z modelami

głębokiego uczenia, takimi jak model seq2seq, LSTM itp., a także z popularnymi pakietami Pythona (Spacy, NLTK itp.) i frameworkami (Tensorflow, Keras).



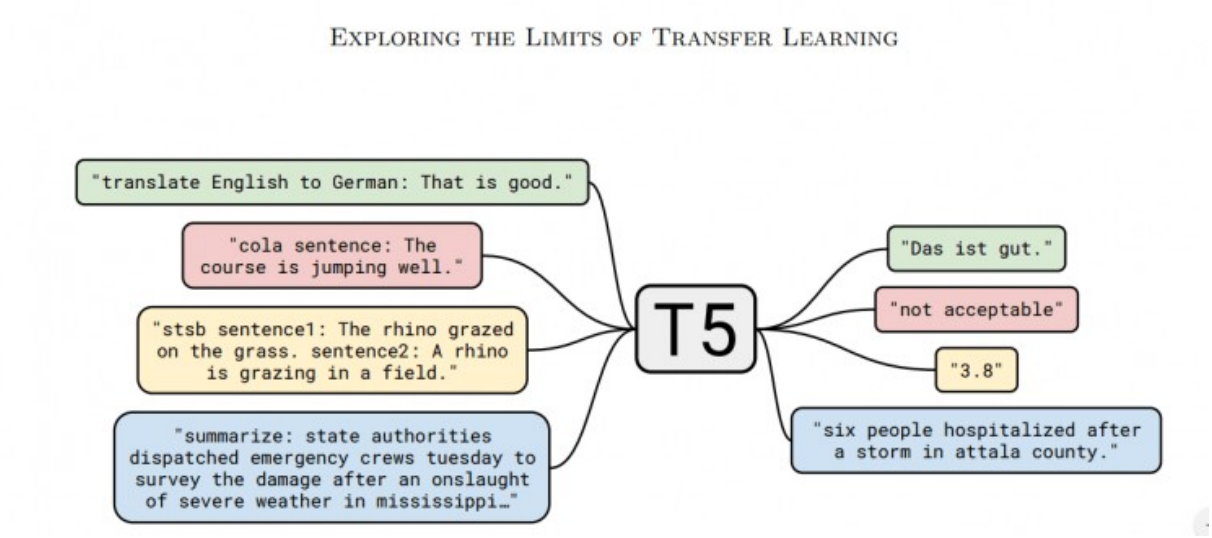
Transformer Simple T5

T5 (Text-To-Text Transfer Transformer) to model transformatora (transformera), który jest szkolony w sposób kompleksowy z tekstem jako wejściem i zmodyfikowanym tekstem jako wyjściem, w przeciwieństwie do modeli w stylu BERT, które mogą wyprowadzać tylko etykietę klasy lub rozpiętość wejścia. To formatowanie tekstu na tekst sprawia, że model T5 nadaje się do wielu zadań NLP, takich jak podsumowanie, odpowiadanie na pytania, tłumaczenie maszynowe i klasyfikowanie.

Model jest szkolony wykorzystując podejście MLM (masked language model). MLM to zadanie polegające na wypełnianiu pustych miejsc, w którym model maskuje część tekstu wejściowego i próbuje przewidzieć, jakie powinno być zamaskowane słowo. T5 zastępuje wiele kolejnych tokenów jednym słowem kluczowym maski.

Firma Google wydała wstępnie wytrenowane modele struktury text-to-text T5, które są szkolone na nieoznakowanym dużym korpusie tekstowym o nazwie C4 (Colossal Clean Crawled Corpus) przy użyciu głębokiego uczenia. C4 to oczyszczone dane tekstowe pobrane z sieci o pojemności 800 GB. Proces czyszczenia obejmuje reduplikację, odrzucanie niepełnych zdań oraz usuwanie obraźliwych lub hałaśliwych treści. T5 oczekuje prefiksu

przed tekstem wejściowym, aby zrozumieć zadanie zadane przez użytkownika. Na przykład „summarize:” w przypadku podsumowania, „cola sentence:” w przypadku klasyfikacji, „translate English to Spanish:” w przypadku tłumaczenia maszynowego itp.



Źródło: https://turbolab.in/abstractive-summarization-using-googles-t5/

W naszym projekcie model T5 wykorzystany został do uczenia nadzorowanego (simpleT5 jest stworzony tylko do tego). Jednak możliwe jest również wykorzystanie transformatora T5 do uczenia nienadzorowanego.

Technologia

W projekcie wykorzystano język Python oraz Jupyter Notebook. Skorzystano przede wszystkim z następujących bibliotek:

- *BeautifulSoup* - biblioteka Pythona do wyciągania danych z plików HTML i XML. Wykorzystana została do scrapping’u danych ze strony internetowej z codziennymi wiadomościami.

- *requests* – biblioteka Pythona obsługująca żądania i zapytania HTTP.

- *pandas* i *numpy* – jedne z najpopularniejszych bibliotek Pythona – wykorzystane do ogólnej obsługi danych.

- *SpaCy* – biblioteka Pythona przeznaczona do zaawansowanego przetwarzania języka naturalnego.

- *simpleT5* – biblioteka umożliwiająca korzystanie z transformatora simpleT5,

oraz wiele innych dodatkowych bibliotek.

Web scraping

Zdecydowano się na pobranie artykułów ze strony https://www.dailymail.co.uk. Strona ta nie blokuje dostępu do pełnych artykułów i nie wymaga logowania. Przez 3 dni pobrano po 150 artykułów ze strony głównej, co dało łącznie około 500 artykułów. Do pobierania danych wykorzystano wcześniej już opisane biblioteki BeautifulSoup oraz requests.

```
#request
r1 = requests.get(url)
r1.status_code

#strona główna
coverpage = r1.content

#stworzenie "zupy"
soup1 = BeautifulSoup(coverpage, 'html5lib')

#wyszukanie wszystkich wiadomości na stronie głównej
coverpage_news = soup1.find_all('h2', class_='linkro-darkred')
len(coverpage_news)
```

Nagłówki newsów oznaczone są tagiem HTML „h2”, dlatego wyszukano wszystkie takie tagi na stronie głównej witryny.

Przykład nagłówka pierwszego z dnia 21.05.2022:

```
<h2 class="linkro-darkred">
      <a href="/news/article-10838305/Four-men-murdered-university-student-24-New-Years-Eve-party-jailed-life.html" itemprop="u
rl">
            Four men who savagely beat a university student, 24, in 'sadistic' attack at New Year's Eve party in Brighton and thr
ew him to his death from a balcony are jailed for life over his murder
      </a>
  </h2>
```

Większa część kodu scrapera prezentowana jest poniżej. Jego głównym zadaniem jest znalezienie linku do artukułu, tytułu newsa oraz całego tekstu. Wszystko to wykonywane jest dzięki użyciu tagów HTML oraz ich wyszukiwaniu. Ponieważ paragraf tekstu znajduje się w tagu „p”, tak też jest on wyszukiwany. Ostatecznie dane zostały zapisane w dwóch zbiorach: pierwszym, zawierającym wyłącznie sam tekst artykułu (wykorzystany zostanie tylko do

porównania dwóch metod streszczania tekstu) oraz drugim, zawierającym nagłówek oraz główny tekst, który zostanie wykorzystany do stworzenia głównego modelu.

	Article_Title	Article_Content
0	Wagatha Christie cost me my career, says Rebek...	Rebekah Vardy has claimed that the Wagatha Chr...
1	Four men who savagely beat a university studen...	Three men and a teenager convicted of murderin...
2	My neighbour's dog keeps jumping up our fence ...	In lockdown, our neighbours took ownership of ...
3	How to BEAT the squeeze! We're all feeling the...	Growing up I was never taught, either at home ...
4	Diners' outrage as cost-of-living crisis TREBL...	Diners have been left outraged after a restaur...
...	...	...
145	Netflix's woke purge: Troubled streaming giant...	Netflix's recent axing of 290 staff targeted m...
146	Could YOU design a Martian metaverse? NASA cha...	NASA is offering $70,000 (£56,000) in cash pri...
147	Seven nights (with flights) in Corfu for £183!...	Tour operators have reductions of around a thi...
148	Farewell to 'One Eyed Baz': Frank Bruno among ...	The funeral of one-eyed anti-gang mentor Barri...
149	Living it BARGE! Tyson Fury and wife Paris swa...	Tyson Fury and his family have rounded off the...

Przykład newsów z dnia 21.05.2022.

Ostatecznie, dane zostały zapisane w odpowiednich plikach CSV.

```
news_contents = []
list_links = []
list_titles = []

for n in np.arange(0, number_of_articles):

    #odnośnik do artykułu
    link = url + coverpage_news[n].find('a')['href']
    list_links.append(link)

    #tytuł
    title = coverpage_news[n].find('a').get_text()
    list_titles.append(title)

    #znalezienie paragrafów "p"
    article = requests.get(link)
    article_content = article.content
    soup_article = BeautifulSoup(article_content, 'html5lib')
    body = soup_article.find_all('p', class_='mol-para-with-font')

    list_paragraphs = []
    for p in np.arange(0, len(body)):
        paragraph = body[p].get_text()
        list_paragraphs.append(paragraph)
        final_article = " ".join(list_paragraphs)

    #usunięcie dodatkowych znaków
    final_article = re.sub("\\xa0", "", final_article)

    news_contents.append(final_article)
```

Streszczenie artykułów z wykorzystaniem SpaCy oraz liczności słów

Zamiast rozumieć tekst, podsumowanie ekstrakcyjne, które zostało tutaj wykorzystane, opiera się na metrykach ilościowych skonstruowanych na podstawie samego tekstu. Podejście to składa się z poniższych kroków:

- spojrzenie na częstotliwość używania określonych słów,

- zsumowanie częstotliwości w każdym zdaniu

- uporządkowanie zdania na podstawie tej sumy.

Oczywiście zakładamy, że użycie słowa o wyższej częstotliwości implikuje bardziej „istotne” znaczenie. Może się to wydawać zbyt uproszczone, ale takie podejście często daje zaskakująco dobre wyniki.

Do tego podejścia wykorzystano jeden z już gotowych kodów dostępnych w Internecie. Zmodyfikowano go o drobny preprocessing danych, czyli usunięcie odnośników.

Wykorzystano, już wcześniej opisaną, bibliotekę SpaCy do zaimportowania wstępnie wytrenowanego potoku NLP, aby pomóc w interpretacji struktury gramatycznej tekstu. Do pozbycia się słów stopu wykorzystano zbudowane już w SpaCy STOP_WORDS, a także pozbyto się interpunkcji (biblioteka punctuation). Użyto również funkcji nlargest, aby wyodrębnić procent najważniejszych zdań.

Zasada działania algorytmu jest dość prosta i została opisana poniżej:

- tokenizacja tekstu z wykorzystaniem SpaCy.

- wektoryzacja i normalizacja liczby wystąpień – policzenie ile razy słowo zostało użyte w zdaniu (w tym wypadku pojedynczym wierszu). Słowa używane częściej mają wyższą wartość.

- obliczenie sumy znormalizowanej liczby dla każdego zdania.

- wyodrębnienie procentu najwyżej ocenionych zdań, które stanowią podsumowanie.

Przykład tekstu przed podsumowaniem:

299	Tyson Fury and his family have rounded off their holiday on the French Rivera with a stay in a luxury hotel after leaving their £18,000-a-night superyacht. The reigning WBC heavyweight champion and his entourage of family and friends have been living it up at the five-star Hotel Boscolo in one of the most fashionable areas of Nice just a stone's throw from the sea. Fury, 33, and his father John were pictured today as they returned to the five star hotel after a 90-minute early morning jog around the city centre while the boxer's wife Paris tucked into breakfast with their six children. The 6ft 9ins retired fighter, who was wearing blue WBC shorts and a black vest emblazoned with his Gypsy King nickname, went largely unrecognised as he pounded around the nearby streets. But he later posted a social media update showing a selfie of him and his bare-chested father, saying they had done a 'little four miler' run followed by 350 steps up to the top of a local landmark and some shadow boxing to keep in shape. A caption to his Instagram story, added: 'Feeling bless to be able to train daily with my old man, special moments.' He also added a new Instagram post featuring a video of him performing sparring moves on an idyllic spot beside a balustrade overlooking the sea, looking a world away from the views around his new £1.7million family home near Morecambe, Lancashire. A caption to the video stated: 'The man in the shadow is always following me! SOUTH OF FRANCE 2022. 1 of the most beautiful places on earth BLESSED'. Tyson and Paris have been constantly posting details of their holiday and pictures of themselves enjoying their break on Instagram. The boxer revealed to his 5.6million followers last night that he was due to return home today, saying in a video: 'It's been a blast'. As his father John nodded in agreement, he added: 'We have had plenty to drink, plenty of food and plenty of good times.' Fury and his family started off their lavish sunshine getaway on the 137ft luxury yacht iRama which can sleep up to 12 guests in six cabins including a master suite and three state rooms. He and Paris were pictured earlier this week speeding around on one of the yacht's two jetskis after mooring up at a beauty spot near Cannes, and later enjoying drinks on board with their friends. But they later swapped life on the ocean wave with up to seven crew looking after them to stay at the acclaimed Hotel Boscolo which has 113 rooms over five floors. Rooms at the hotel cost around £400 a night and are said to appeal to well-heeled guests 'looking for a certain discretion and tranquility'. The hotel on the stylish Boulevard Victor Hugo was built in 1913 in a classic Bella Epoque style, but has a modern largely white interior including a marble-clade reception. Paris, 32, who is reported to be pregnant with her seventh child and her children Venezuela, 12, Prince John James nine, Prince II, four, Amber, three, Prince Adonis Amaziah, two, and eight month old Athena have been making full use of the rooftop swimming pool during their stay. The hotel's restaurant La Pescheria specialises in Italian cuisine and serves mainly fish, but Tyson and his friends and family have also been visiting bars and eateries in the surrounding area. He has been happily posing for selfies with boxing fans and restaurant staff. Friends who have been enjoying the seaside break with Tyson and Paris and their family include Gabrielle Briggs. She has been posting pictures of herself on the couple's yacht, beside the pool at the Boscalo Hotel and on a girls' night out with Paris. Tyson admitted yesterday that he he'd had 'too many to drink' after video emerged of him kicking out at a taxi and struggling to stand while in Cannes following a day of boozing on Wednesday. The video appeared to show him being held up upright by his father as he stumbled towards the white taxi during daylight while the driver made apparent hand gestures out of the window, suggesting he was too drunk to enter. Tyson's friends seemed to try and insist that he should still be taken, but the driver sped away and Tyson appeared to kick out at the bumper as his father tried to restrain him. In an Instagram story post on Thursday morning, the boxer said: 'Nice run dad this morning through the city.' John replied: 'Back at it. Too many we had last night.' Tyson then asked his father: 'Strong beer wasn't it dad?' John replied: 'Strongest I have had in a while.' Fury has maintained he has retired from boxing since his sixth-round knockout of Dillian Whyte at Wembley Stadium last month. His multiple holiday clips of him and his wife on his Instagram stories, included one where he is seen asking her to hand him a beer while he relaxed on a sun lounger. As his blonde wife obliged with a smile on her face, a pleased Fury captioned the video: 'If Carlsberg did wives.' Paris who has been soaking up the sun in a blue and white striped bikini showing her deep tan has been seen enjoying mocktails. She also had no hesitation in fulfilling another one of her husband's requests by passing him a carrot stick dipped in hummus as she sipped her wine glass.

Podsumowanie dla 10 procent:

```
summarizer(df.Article_Content[299], 0.1)
```

```
"He also added a new Instagram post featuring a video of him performing sparring moves on an idyllic spot beside a balustrade o
verlooking the sea, looking a world away from the views around his new £1.7million family home near Morecambe, Lancashire.Fury,
33, and his father John were pictured today as they returned to the five star hotel after a 90-minute early morning jog around
the city centre while the boxer's wife Paris tucked into breakfast with their six children.But he later posted a social media u
pdate showing a selfie of him and his bare-chested father, saying they had done a 'little four miler' run followed by 350 steps
up to the top of a local landmark and some shadow boxing to keep in shape."
```

Podsumowanie dla 40 procent:

```
summarizer(df.Article_Content[299], 0.4)
```

```
"He also added a new Instagram post featuring a video of him performing sparring moves on an idyllic spot beside a balustrade o
verlooking the sea, looking a world away from the views around his new £1.7million family home near Morecambe, Lancashire.Fury,
33, and his father John were pictured today as they returned to the five star hotel after a 90-minute early morning jog around
the city centre while the boxer's wife Paris tucked into breakfast with their six children.But he later posted a social media u
pdate showing a selfie of him and his bare-chested father, saying they had done a 'little four miler' run followed by 350 steps
up to the top of a local landmark and some shadow boxing to keep in shape.The video appeared to show him being held up upright
by his father as he stumbled towards the white taxi during daylight while the driver made apparent hand gestures out of the win
dow, suggesting he was too drunk to enter.Tyson Fury and his family have rounded off their holiday on the French Rivera with a
stay in a luxury hotel after leaving their £18,000-a-night superyacht.Fury and his family started off their lavish sunshine get
away on the 137ft luxury yacht iRama which can sleep up to 12 guests in six cabins including a master suite and three state roo
ms.Rooms at the hotel cost around £400 a night and are said to appeal to well-heeled guests 'looking for a certain discretion a
nd tranquility'.The reigning WBC heavyweight champion and his entourage of family and friends have been living it up at the fiv
e-star Hotel Boscolo in one of the most fashionable areas of Nice just a stone's throw from the sea.He and Paris were pictured
earlier this week speeding around on one of the yacht's two jetskis after mooring up at a beauty spot near Cannes, and later en
joying drinks on board with their friends.The hotel's restaurant La Pescheria specialises in Italian cuisine and serves mainly
fish, but Tyson and his friends and family have also been visiting bars and eateries in the surrounding area.The boxer revealed
to his 5.6million followers last night that he was due to return home today, saying in a video: 'It's been a blast'.As his fath
er John nodded in agreement, he added: 'We have had plenty to drink, plenty of food and plenty of good times.'But they later sw
apped life on the ocean wave with up to seven crew looking after them to stay at the acclaimed Hotel Boscolo which has 113 room
s over five floors.Tyson's friends seemed to try and insist that he should still be taken, but the driver sped away and Tyson a
ppeared to kick out at the bumper as his father tried to restrain him."
```

Podsumowanie dla 3 procent:

```
summarizer(df.Article_Content[299], 0.03)
```

```
'He also added a new Instagram post featuring a video of him performing sparring moves on an idyllic spot beside a balustrade o
verlooking the sea, looking a world away from the views around his new £1.7million family home near Morecambe, Lancashire.'
```

Podsumowanie dodatkowego tekst:

```
summarizer(new_text, 0.1)
```

```
"Ukraine will battle to a bloody victory but the Russian invasion will end with diplomacy, Ukrainian's Volodymyr Zelensky said
in an hour-long TV interview to mark his third anniversary as President."
```

Streszczenie tekstu z wykorzystaniem transformatora i uczenia nadzorowanego

Do tej części projektu wykorzystano simpleT5. Jest to model zbudowany na bazie technologii PyTorch-lightning i Transformers, co pozwala szybko trenować modele T5.

T5 to model z koderem i dekoderem. Konwertuje wszystkie problemy NLP, takie jak tłumaczenie języka, podsumowania, generowanie tekstu, odpowiadanie na pytania, na zadanie zamiany tekstu na tekst.

Przygotowano klasy pomocnicze takie jak:

- ustawienia – klasa wskazująca ogólne elementy, takie jak rodzaj modelu, kolumny, wykorzystanie technologii CUDA (na gogle collab), wskazanie ilości epok, wielkości zbioru testowego oraz do treningu.

```python
class Settings:

    MODEL_TYPE = "t5"
    MODEL_NAME = "t5-base"

    DEVICE = torch.device("cuda" if torch.cuda.is_available() else "cpu")


    TRAIN_DATA = "df_info.csv"
```

```
columns_dict = {"Article_Title": "target_text", "Article_Content": "source_text"}
df_column_list = ['source_text', 'target_text']
SUMMARIZE_KEY = "summarize: "
SOURCE_TEXT_KEY = 'source_text'
TEST_SIZE = 0.2
BATCH_SIZE = 8
source_max_token_len = 128
target_max_token_len = 50
train_df_len = 240
test_df_len = 60
```

Model T5 wymaga wskazania słowa, które wskazuje na tekst będący podsumowanie. W naszym przypadku będzie to po prostu słowo summarizer. W następnych krokach zostanie ono dodane przed nagłówkiem w każdym wierszu.

- preprocessing – podstawowe operacje na zbiorze danych będące preprocessingiem, czyli pozbycie się linków, specjalnych znaków, dodatkowych spacji oraz \n. Następnie zmiana nazw kolumn (simpleT5 oczekuje od zbioru aby posiadał dwie kolumny: "source_text" i "target_text") oraz dodanie wcześniej wymienione słowa summarizer na początku nagłówków.

```
def preprocess_data(self, data_path):
    df = pd.read_csv(data_path, encoding=self.settings.encoding, usecols=self.settings.Columns)
    df = df.rename(columns=self.settings.columns_dict)
    df = df[self.settings.df_column_list]
    df[self.settings.SOURCE_TEXT_KEY] = self.settings.SUMMARIZE_KEY + df[self.settings.SOURCE_TEXT_KEY]

    return df
```

- klasa model – wskazująca na model T5

- klasa Train – tutaj inicjowane są wszystkie wcześniejsze klasy oraz następuje proces trenowania danych. Dodatkowo tutaj dodane zostały wszelkie ostrzeżenia oraz informacje drukujące się podczas trenowania.

```
def train(self, df):
    try:
        train_df, test_df = train_test_split(df, test_size=self.settings.TEST_SIZE)

        self.t5_model.model.train(train_df=train_df[:self.settings.train_df_len],
                                  eval_df=test_df[:self.settings.test_df_len],
                                  source_max_token_len=self.settings.source_max_token_len,
                                  target_max_token_len=self.settings.target_max_token_len,
                                  batch_size=self.settings.BATCH_SIZE, max_epochs=self.settings.EPOCHS,
                                  use_gpu=self.settings.USE_GPU)
```

Po odpowiednim trenowaniu wyniki zapisują się w folderze outputs. Następnie najlepsze wagi wykorzystujemy do podsumowania nowego kawałka tekstu, tym razem wziętego ze strony CNN.

Ukraine will battle to a bloody victory but the Russian invasion will end with diplomacy, Ukrainian's Volodymyr Zelensky said in an hour-long TV interview to mark his third anniversary as President. "We did not start this war. But we have to finish it," Zelensky said in an interview that was recorded Friday and broadcast Saturday. "Victory will be bloody in battle. But the end will be in diplomacy. We want everything back. Russia does not want to give anything away," Zelensky said. The broadcast also included an interview with Zelensky's wife Olena Zelenska, who said the war had not changed her husband. "I can't say that he has changed. As he was a reliable husband and man, so he is," she said. But she lamented not being able to see him for months. "Our family, like all Ukrainian families, is now torn. We didn't see each other for two and a half months, we only talked on the phone. Thank you for this opportunity [this interview], because we are spending time together now -- dating on TV," she said.

Nagłówek ze strony CNN: After bloody battles, Ukraine war will end with diplomacy, Zelensky vows in anniversary interview

Podsumowanie wykonane przez model: Ukraine will fight to a bloody victory but the Russian invasion will end with diplomacy, Ukrainian President Volodymyr Zelensky says

Porównując to z podsumowaniem stworzonym przez poprzedni model: Ukraine will battle to

a bloody victory but the Russian invasion will end with diplomacy, Ukrainian's Volodymyr Zelensky said in an hour-long TV interview to mark his third anniversary as President.

Podsumowanie

Oba streszczenia są bardzo podobne do siebie i całkowicie oddają sens artykułu. Obie metody, zarówno ekstrakcyjna jak i abstrakcyjna bez problemu radzą sobie z podsumowywaniem artykułów. Metoda abstrakcyjna jest o wiele bardziej czasochłonna ponieważ w naszym przypadku długo zajmuje trenowanie modelu, jej zaletą jest jednak możliwość odpowiedniego dostosowania parametrów. Metoda ekstrakcyjna bazowała wyłącznie na zliczaniu częstości słów w zdaniu.