

ROZWIĄZYWANIE PROBLEMU MEDYCZNEGO NA DANYCH TEKSTOWYCH

GRUPA 2: LILIANA GMEREK, DARIUSZ IWANOW,
DOMINIKA SZOTTA



DATA PREPARATION

- Redukcja stron analizowanych CHPL do 4, usunięcie części strony zarezerwowanej dla stopki
- Przekształcenie zredukowanych plików pdf do DataFrame:

In [12]:	<code>text_df.head()</code>	
Out [12]:	Type to search	
	filename	CHPCL ↑
0	Charakterystyka-10025-2021-12-22-8574_N-2022-01-05.pdf	<p>CHARAKTERYSTYKA PRODUKTU LECZNICZEGO 1. NAZWA PRODUKTU LECZNICZEGO Metazydyna, 20 mg, tabletki powlekane 2. SKŁAD JAKOŚCIOWY I ILOŚCIOWY 1 tabletka zawiera 20 mg trimetazydyny dichlorowodorku (Trimetazidini dihydrochloridum). Substancje pomocnicze o znanym działaniu: 0,05 mg czerwieni koszenilowej. Pełny wykaz substancji pomocniczych, patrz: punkt 6.1. 3. POSTAĆ FARMACEUTYCZNA Tabletki powlekane. Tabletki barwy czerwonej, okrągłe, obustronnie wypukłe, gładkie. 4. SZCZEGÓŁOWE DANE KLINICZNE 4.1 Wskazania do stosowania Trimetazydyna jest wskazana u dorosłych w terapii skojarzonej w objawowym leczeniu pacjentów ze stabilną dławicą piersiową, u których leczenie przeciwdławicowe pierwszego rzutu jest niewystarczające lub nie jest tolerowane. 4.2 Dawkowanie i sposób podawania Dawka wynosi jedną tabletkę 20 mg trimetazydyny (w postaci dichlorowodorku) trzy razy na dobę, podczas posiłków. Szczególne grupy pacjentów Pacjenci z zaburzeniami czynności nerek U pacjentów z umiarkowanym zaburzeniem czynności nerek (klirens kreatyniny 30-60 ml/min; patrz punkty 4.4 i 5.2), zalecana dawka wynosi 1 tabletkę 20 mg dwa razy na dobę, tzn. raz rano i raz wieczorem, podczas posiłków. Pacjenci w podeszłym wieku U pacjentów w podeszłym wieku ekspozycja na trimetazydynę może być zwiększona wskutek związanego z wiekiem pogorszenia czynności nerek (patrz punkt 5.2). U pacjentów z umiarkowanym zaburzeniem czynności nerek (klirens kreatyniny 30-60 ml/min) zalecana dawka wynosi 1 tabletkę 20 mg dwa razy na dobę, tzn. raz rano i raz wieczorem, podczas posiłków. Ustalanie dawki u pacjentów w podeszłym wieku należy przeprowadzać z zachowaniem ostrożności (patrz punkt 4.4). Dzieci i młodzież Nie ustalono bezpieczeństwa stosowania ani skuteczności trimetazydyny u dzieci w wieku poniżej 18 lat. Brak dostępnych danych. 4.3 Przeciwwskazania Nadwrażliwość na substancję czynną lub którąkolwiek substancję pomocniczą wymienioną - Choroba Parkinsona, objawy parkinsonizmu, drżenia, zespół niespokojnych nóg oraz inne powiązane zaburzenia ruchowe. - Ciężkie zaburzenia czynności nerek (klirens kreatyniny <30 ml/min). 4.4 Specjalne ostrzeżenia i środki ostrożności dotyczące stosowania Produkt leczniczy Metazydyna nie służy do leczenia nowotworów dławicy piersiowej, ani też leczenia</p>

DATA EXTRACTION

- Po analizie danych i usunięciu oboczności wyrażeń inną funkcją, wykorzystanie indeksacji

```
def extract_columns(df, column):
    df = apply_replacements(df, column)
    for index, row in df.iterrows():
        #nazwa
        start_index_nazwa = row[column].find("1")
        end_index_nazwa = row[column].find("2.") if row[column].find("2.") != -1 else row[column].find("2")
        nazwa = row[column][start_index_nazwa:end_index_nazwa].strip()
        df.at[index, 'nazwa'] = nazwa
        #skład
        start_index_sklad = row[column].find("2.") if row[column].find("2.") != -1 else row[column].find("2 skład ilościowy i jakościowy ")
        end_index_sklad = row[column].find("3.") if row[column].find("3.") != -1 else row[column].find("4.1")
        sklad = row[column][start_index_sklad:end_index_sklad].strip()
        df.at[index, 'skład'] = sklad
        #wskazania
        start_index_wskazania = row[column].find("4.1") if row[column].find("4.1") != -1 else row[column].find("wskazania")
        end_index_wskazania = row[column].find("4.2")
        wskazania = row[column][start_index_wskazania:end_index_wskazania].strip()
        df.at[index, 'wskazania'] = wskazania

    return df[['filename', 'nazwa', 'skład', 'wskazania']]
```

DATA EXTRACTION

In [15]:

```
text_extracted.iloc[865:866,:]
```

Out [15]:

Type to search

	filename	nazwa	skład	wskazania
865	Charakterystyka-47283-2023-10-31-15084_D-2023-11-10.pdf	1. nazwa produktu leczniczego ebozan, 2,5 mg, tabletki ebozan, 5 mg, tabletki ebozan, 10 mg, tabletki ebozan 20 mg, tabletki	2. skład jakościowy i ilościowy ebozan, 2,5 mg, tabletki każda tabletki zawiera 2,5 mg torasemidu (torasemidum). substancja pomocnicza o znanym działaniu: każda tabletki zawiera 48 mg laktozy (w postaci laktozy jednowodnej). ebozan, 5 mg, tabletki każda tabletki zawiera 5 mg torasemidu (torasemidum). substancja pomocnicza o znanym działaniu: każda tabletki zawiera 97 mg laktozy (w postaci laktozy jednowodnej). ebozan, 10 mg, tabletki każda tabletki zawiera 10 mg torasemidu (torasemidum). substancja pomocnicza o znanym działaniu: każda tabletki zawiera 193 mg laktozy (w postaci laktozy jednowodnej). ebozan, 20 mg, tabletki każda tabletki zawiera 20 mg torasemidu (torasemidum). substancja pomocnicza o znanym działaniu: każda tabletki zawiera 386 mg laktozy (w postaci laktozy jednowodnej). pełny wykaz substancji pomocniczych, patrz punkt 6.1.	4.1 wskazania do stosowania pochodzenia nerkowego. ponadto produkt leczniczy ebozan 2,5 mg i 5 mg, tabletki są wskazane w leczeniu nadciśnienia tętniczego pierwotnego. produkt leczniczy ebozan jest wskazany u dorosłych i młodzieży w wieku powyżej 12 lat.

DATA CLEANING - FUNKCJE



Clean_formatting - Usunięcie znaków interpunkcyjnych, zamiana na małe litery, zastąpienie wielu spacji jedną spacją i usunięcie cyfr



Remove_stop_words - Usunięcie wskazanych słów dla danej kolumny, usunięcie nazw substancji pomocniczych z kolumny skład, dla każdej kolumny osobno



Remove_duplicated_words - usunięcie powtarzających się słów-tokenów



Tokenize_text - tokenizacja tekstu przy użyciu spaCy, model large



Lemmatize_text - lematyzacja tekstu dla skład i wskazania, pozostawienie jedynie rzeczowników

PROCESS TEXT COLUMNS




WYSZUKIWARKA 1 - PO WSKAZANIU

Początkowo wykorzystano Elastic Search.



Ostatecznie wykorzystano macierz TFIDF.



Znaczny wpływ na poprawienie wyników miało zastosowanie lematyzacji dla wskazań oraz zostawienie samych rzeczowników

PRZYKŁADOWE WYNIKI WYSZUKIWANIA

- Ból głowy,
- Astma.

Charakterystyka-12872-20191230000000-7999_A-20200418000213.pdf,"Ibum grip tabletki powlekane",0.214
Charakterystyka-12427-2022-10-21-9356_N-2022-11-04.pdf,"Ibupar tabletki drażowane",0.207
Charakterystyka-12872-20191230000000-7999_A-20200418000213.pdf,"Ibum grip tabletki powlekane",0.206
Charakterystyka-19363-2022-11-17-9432_N-2022-12-01.pdf,"Ibumax tabletki powlekane",0.206
Charakterystyka-19366-2022-11-17-9432_N-2022-12-01.pdf,"Ibumax tabletki powlekane",0.206
Charakterystyka-30097-2023-04-15-14173_B-2023-06-29.pdf,"Nalgesin mini tabletki powlekane",0.206
Charakterystyka-36541-2021-03-15-9949_B-2021-07-29.pdf,"Aponapro fast kapsułki miękkie",0.205
Charakterystyka-38098-2023-10-16-13202_N-2023-11-10.pdf,"Ibum supermax kapsułki miękkie",0.188
Charakterystyka-28723-2022-12-08-9433_N-2022-12-22.pdf,"Ibum express forte kapsułki miękkie",0.184
Charakterystyka-12819-2022-12-08-9433_N-2022-12-22.pdf,"Ibum forte kapsułki miękkie",0.181
Charakterystyka-24254-2022-11-28-9423_N-2022-12-12.pdf,"Ibufen junior kapsułki miękkie",0.176
Charakterystyka-42458-2024-02-10-16822_B-2024-03-16.pdf,"Ibuprofen dr max kapsułki miękkie",0.174
Charakterystyka-40885-2022-12-08-9433_N-2022-12-22.pdf,"Ibum comfort minicaps kapsułki miękkie",0.173
Charakterystyka-41158-2022-12-08-9433_N-2022-12-22.pdf,"Ibum femina kapsułki miękkie",0.173
Charakterystyka-22792-20200930000000-9281_B-202102090000500.pdf,"Ibalgin fast tabletki powlekane",0.168
Charakterystyka-26350-2024-02-26-16481_N-2024-03-11.pdf,"Ibufen baby czopki",0.167
Charakterystyka-1561-2024-02-21-16482_N-2024-03-06.pdf,"Metafen ibuprofen tabletki powlekane",0.165
Charakterystyka-19752-2024-03-05-16133_M-2024-03-06.pdf,"Mig tabletki powlekane",0.161
Charakterystyka-42702-2023-07-18-14496_B-2023-07-26.pdf,"Apap intense tabletki powlekane",0.158
Charakterystyka-43507-2023-02-22-9689_N-2023-03-08.pdf,"Kidofen max ml zawiesina doustna",0.152
Charakterystyka-35456-2024-03-14-16838_N-2024-03-28.pdf,"Ibum forte ml zawiesina doustna",0.149
Charakterystyka-9411-2024-03-14-16838_N-2024-03-28.pdf,"Ibum forte ml zawiesina doustna",0.149
Charakterystyka-35455-2022-03-15-8726_N-2022-03-29.pdf,"Ibum ml zawiesina doustna",0.147
Charakterystyka-42073-2023-03-06-14036_B-2023-08-22.pdf,"Kapsułki do żucia elastyczne chpl ibuprofen banner",0.143
Charakterystyka-36810-2021-03-31-7800_N-2021-04-14.pdf,"Ibenal max tabletki powlekane",0.141
Charakterystyka-16473-20180910000000-5254_A-201810300000829.pdf,"Agapurin sr tabletki o przedłużonym uwalnianiu",0.128
Charakterystyka-324-20201217000000-7326_N-202012310000646.pdf,"Agapurin tabletki drażowane",0.127
Charakterystyka-2133-2021-06-24-8139_N-2021-07-08.pdf,"Effortil mgg krople doustne",0.114
Charakterystyka-25096-2023-02-05-12593_B-2023-02-22.pdf,"Biofenac tabletki powlekane",0.110
Charakterystyka-40775-2022-12-09-12299_B-2023-01-04.pdf,"Axoprofen tabletki powlekane",0.103
Charakterystyka-8157-20200817000000-6722_N-20200831000113.pdf,"Polfilin mgml ml roztwór do wstrzykiwań",0.103
Charakterystyka-45506-2024-01-19-17070_B-2024-03-15.pdf,"Straben pastylki twarde",0.097
Charakterystyka-25109-2023-02-05-12593_B-2023-02-22.pdf,"Biofenac proszek do sporządzania zawiesiny doustnej",0.092
Charakterystyka-41866-2024-02-15-16877_B-2024-02-22.pdf,"Ibuprofen alkaloidint mgml zawiesina doustna",0.090
Charakterystyka-37993-20200807000000-8593_B-202009080000347.pdf,"Flustad max tabletki powlekane",0.088
Charakterystyka-39499-2023-02-10-12779_B-2023-03-10.pdf,"Dektac tabletki powlekane",0.087
Charakterystyka-45875-2023-01-19-13288_B-2023-04-19.pdf,"Dekenor tabletki powlekane",0.087
Charakterystyka-42894-2021-07-23-1571_D-2022-04-13.pdf,"Dexak granulaty w saszetce",0.080
Charakterystyka-36318-2023-01-30-4033_M-2023-01-31.pdf,"Septofar mięta pastylki twarde",0.079
Charakterystyka-40774-2024-03-06-17240_B-2024-03-22.pdf,"Ibuprofen aristo tabletki powlekane",0.079
Charakterystyka-22846-20201023000000-7001_N-20201106001109.pdf,"Diosminum aflofarm tabletki",0.076
Charakterystyka-25037-2023-04-27-13564_B-2023-05-10.pdf,"Ketesse sl granulaty do sporządzania roztworu doustnego",0.074
Charakterystyka-38665-20210201000000-7306_N-20210215000358.pdf,"Aflavic max tabletki",0.074
Charakterystyka-43451-20201223000000-2845_M-20210121000914.pdf,"Septogard smak miodowopomarańczowy pastylki twarde",0.074
Charakterystyka-11391-2024-03-04-17379_A-2024-04-03.pdf,"Aspicam bio tabletki",0.073
Charakterystyka-31740-2022-11-28-9423_N-2022-12-12.pdf,"Ibufen dla dzieci forte o smaku malinowym ml zawiesina doustna",0.072
Charakterystyka-36965-2021-06-08-7816_N-2021-06-22.pdf,"Ibenal czopki",0.072
Charakterystyka-8613-2024-02-13-17135_A-2024-03-16.pdf,"Rantudil forte kapsułki twarde",0.072
Charakterystyka-34491-2023-10-30-15963_A-2023-11-25.pdf,"Mel tabletki ulegające rozpadowi w jamie ustnej",0.071
Charakterystyka-10870-20200914000000-6904_N-20200928000217.pdf,"Rostil tabletki",0.070


```
(myenv) D:\nlp-group2-main\src>python main.py -q "astma"
Charakterystyka-14200-2021-02-13-9574_B-2022-07-20.pdf,"Ithib ribuspir mikrogramówdawkę odmierzoną aerozol inhalacyjny roztwór",0.233
Charakterystyka-10695-20170712000000-2873_N-20170802000128.pdf,"Miflonide breezhaler mikrogramówdawkę inhalacyjną proszek do inhalacji w kapsułkach twardych",0.187
Charakterystyka-6640-2023-08-29-14891_A-2023-08-31.pdf,"Theophyllinum tramco mgml roztwór do infuzji",0.165
Charakterystyka-12054-20200728000000-6434_N-20200814000113.pdf,"Budelin novolizer µgdawkę inhalacyjną proszek do inhalacji",0.156
Charakterystyka-13089-2021-07-27-9613_A-2021-09-28.pdf,"Asmanex twisthaler µgdawkę proszek do inhalacji",0.154
Charakterystyka-19543-2021-07-27-9613_A-2021-09-28.pdf,"Asmanex twisthaler µgdawkę proszek do inhalacji",0.154
Charakterystyka-42461-2021-04-15-9746_B-2021-05-20.pdf,"Soprobeć mikrogramówdawkę odmierzoną aerozol inhalacyjny roztwór",0.143
Charakterystyka-3065-2023-08-04-14806_A-2023-08-26.pdf,"Adrenalina wzf mgml roztwór do wstrzykiwań",0.141
Charakterystyka-24994-2019-09-20-11404_B-2022-05-19.pdf,"Asmenol pph tabletki do rozgryzania i żucia",0.138
Charakterystyka-24997-2019-09-20-11404_B-2022-05-19.pdf,"Asmenol pph tabletki do rozgryzania i żucia",0.136
Charakterystyka-8747-2023-01-03-9474_N-2023-01-17.pdf,"Baladex ml syrop",0.136
Charakterystyka-24839-2021-07-29-10166_B-2022-05-25.pdf,"Pulveril mikrogramówdawkę aerozol inhalacyjny zawiesina",0.135
Charakterystyka-24993-20190920000000-7405_B-20191005000511.pdf,"Asmenol pph tabletki powlekane",0.133
Charakterystyka-20966-2023-07-14-13552_M-2023-08-25.pdf,"Singulair mini granulat",0.130
Charakterystyka-8262-2023-07-14-13552_M-2023-08-25.pdf,"Singulair tabletki powlekane",0.130
Charakterystyka-43513-2021-12-07-1751_D-2022-01-03.pdf,"Montelukast bluefish pharma tabletki powlekane",0.129
Charakterystyka-24908-20200814000000-8560_B-20200829000422.pdf,"Montelukast bluefish tabletki do rozgryzania i żucia",0.128
Charakterystyka-36174-20180801000000-4736_A-20180901001429.pdf,"Flutixon neb ml zawiesina do nebulizacji",0.127
Charakterystyka-20126-20160617000000-2125_B-20160624000333.pdf,"Asmenol tabletki do rozgryzania i żucia",0.126
Charakterystyka-46025-2024-02-13-15193_D-2024-03-10.pdf,"Cyxodil mikrogramówdawkę inhalacyjną aerozol inhalacyjny roztwór",0.126
Charakterystyka-46026-2024-02-13-15193_D-2024-03-10.pdf,"Cyxodil mikrogramówdawkę inhalacyjną aerozol inhalacyjny roztwór",0.126
Charakterystyka-19910-2021-12-15-10764_B-2021-12-22.pdf,"Montelukast teva tabletki powlekane",0.124
Charakterystyka-10975-20201113000000-6968_N-20201127001028.pdf,"Flixotide ml zawiesina do nebulizacji",0.123
Charakterystyka-19475-20201113000000-6968_N-20201127001026.pdf,"Flixotide ml zawiesina do nebulizacji",0.123
Charakterystyka-24387-20200416000000-8757_B-20201014001316.pdf,"Montelukast sandoz granulat",0.123
Charakterystyka-45357-2023-09-05-15042_A-2023-09-09.pdf,"Flixodil combo mikrogramów mikrogramówdawkę odmierzoną aerozol inhalacyjny zawiesina",0.123
Charakterystyka-33837-20190701000000-7045_A-20190824000500.pdf,"Montelukast lekam tabletki do rozgryzania i żucia",0.121
Charakterystyka-46003-2023-04-17-2101_D-2023-05-14.pdf,"Ondemet mgml zawiesina do nebulizacji",0.118
Charakterystyka-36522-20190306000000-6515_B-20190406000134.pdf,"Airflusal mikrogramów mikrogramówdawkę odmierzoną aerozol inhalacyjny zawiesina",0.117
Charakterystyka-36523-20190306000000-6515_B-20190406000135.pdf,"Airflusal mikrogramów mikrogramówdawkę odmierzoną aerozol inhalacyjny zawiesina",0.117
Charakterystyka-13141-20200427000000-8025_A-20200429000451.pdf,"Foramed µgdawkę inhalacyjną proszek do inhalacji w kapsułkach twardych",0.108
Charakterystyka-22269-2023-12-01-16598_B-2024-02-02.pdf,"Sabumalin mikrogramówdawkę aerozol inhalacyjny zawiesina",0.108
Charakterystyka-46690-2023-08-22-13674_D-2023-09-12.pdf,"Befoair mikrogramów mikrogramówdawkę odmierzoną aerozol inhalacyjny roztwór",0.104
Charakterystyka-41610-2021-03-05-1431_D-2021-04-27.pdf,"Fostex nexthaler mikrogramów mikrogramówdawkę inhalacyjną proszek do inhalacji",0.103
Charakterystyka-28966-20170515000000-3434_B-20170719000421.pdf,"Flutiform mikrogramów mikrogramówdawkę inhalacyjną aerozol inhalacyjny zawiesina",0.101
Charakterystyka-28967-20170515000000-3434_B-20170719000422.pdf,"Flutiform mikrogramów mikrogramówdawkę inhalacyjną aerozol inhalacyjny zawiesina",0.101
Charakterystyka-28968-20170515000000-3434_B-20170719000423.pdf,"Flutiform mikrogramów mikrogramówdawkę inhalacyjną aerozol inhalacyjny zawiesina",0.101
Charakterystyka-37194-20191211000000-7548_A-20191221001417.pdf,"Budixon neb mgml zawiesina do nebulizacji",0.099
Charakterystyka-41188-2021-11-30-10786_B-2022-01-10.pdf,"Newair airmaster mikrogramów mikrogramówdawkę proszek do inhalacji podzielony",0.098
```



```
(lp) (base) MacBook-Pro:src LILI$ python main2.py -f "/Users/LILI/Projects/studia_grupowy/3_nlp_2/test_file_search/Results for file: Charakterystyka-9820-20181024000000-3466_N-20181115000255.pdf
Charakterystyka-17492-20201217000000-7158_N-20201231000627.pdf,"Setal mr o zmodyfikowanym",1.000
Charakterystyka-20544-20210212000000-7663_N-20210302000639.pdf,"Preductal mr o zmodyfikowanym",1.000
Charakterystyka-9820-20181024000000-3466_N-20181115000255.pdf,"Triacyt mr o zmodyfikowanym",1.000
Charakterystyka-10025-2021-12-22-8574_N-2022-01-05.pdf,"Metazydyna",0.982
Charakterystyka-12788-2023-03-02-13309_A-2023-04-20.pdf,"Cytoprotectin mr o zmodyfikowanym",0.982
Charakterystyka-35856-20171228000000-3233_N-20180426000114.pdf,"Protrimil o zmodyfikowanym",0.971
Charakterystyka-38272-20200226000000-7904_A-20200320000950.pdf,"Sivamat o zmodyfikowanym",0.955
Charakterystyka-38273-20200226000000-7902_A-20200320000937.pdf,"Zulkeza o zmodyfikowanym",0.933
Charakterystyka-38274-2021-11-05-11041_B-2022-03-03.pdf,"Trimetazidine zentiva o zmodyfikowanym",0.933
Charakterystyka-26235-20181203000000-5832_B-20181214001009.pdf,"Protevasc sr o",0.892
Charakterystyka-23947-20201021000000-8927_B-20201119000756.pdf,"Vascotazin o zmodyfikowanym",0.787
Charakterystyka-37589-2023-05-08-12559_N-2023-05-22.pdf,"Cirrus o",0.729
Charakterystyka-36343-20170830000000-3286_A-20170921000413.pdf,"Doxazosin genoptim",0.715
Charakterystyka-20736-2023-09-11-15743_A-2023-11-10.pdf,"Sudafed xylospray do",0.732
Charakterystyka-10872-2021-09-02-8214_N-2021-09-16.pdf,"Otrivin do",0.713
Charakterystyka-10873-2023-08-18-15276_A-2023-10-07.pdf,"Xylometazolin vibrocil do",0.730
Charakterystyka-7223-2021-08-09-8132_N-2021-08-23.pdf,"Xylometazolin vp do",0.694
Charakterystyka-37849-2023-07-17-5847_N-2023-08-10.pdf,"Xylopronol do",0.690
Charakterystyka-7224-2023-08-04-14808_A-2023-08-26.pdf,"Xylometazolin wzf do",0.695
Charakterystyka-7222-2023-03-20-8553_N-2023-04-03.pdf,"Xylorin µgml do",0.672
Charakterystyka-43316-2023-03-30-9541_N-2023-04-15.pdf,"Xylometazolin fortis do",0.711
Charakterystyka-13206-20200908000000-6614_N-20200922000726.pdf,"Acatar actitabs",0.683
Charakterystyka-17436-20191022000000-2401_M-20191023000804.pdf,"Doxagen o",0.709
Charakterystyka-45147-2023-03-24-13547_B-2023-05-09.pdf,"Ramidilan hct",0.689
Charakterystyka-12126-2021-10-04-9860_A-2021-12-08.pdf,"Actitrin syrop",0.650
Charakterystyka-44217-2023-02-28-12744_A-2023-03-07.pdf,"Doppelsept gardło do ssania",0.700
Charakterystyka-41301-2022-07-11-10714_A-2022-07-20.pdf,"Faringan do ssania",0.652
Charakterystyka-7218-2023-09-21-15165_A-2023-09-26.pdf,"Xylogel do",0.630
Charakterystyka-2972-2022-01-25-8665_N-2022-02-08.pdf,"Hydrochlorothiazidum polpharma",0.670
Charakterystyka-31568-2022-07-04-11775_B-2022-08-06.pdf,"Perindopril amlodipine teva pharmaceuticals",0.649
Charakterystyka-19227-2022-01-13-11150_B-2022-03-25.pdf,"Coprenessa",0.656
Charakterystyka-30754-2022-12-28-3915_M-2022-12-29.pdf,"Coamlessa",0.703
Charakterystyka-37685-2023-02-20-9665_N-2023-03-06.pdf,"Diclovit",0.680
Charakterystyka-25401-2024-04-02-17477_B-2024-04-11.pdf,"Doxazosin aurobindo",0.653
Charakterystyka-40518-2024-01-30-16757_B-2024-02-10.pdf,"Sinuzolin do",0.630
Charakterystyka-40081-2023-08-17-15127_B-2023-09-22.pdf,"Valtricom",0.647
Charakterystyka-40083-2023-08-17-15127_B-2023-09-22.pdf,"Valtricom",0.639
Charakterystyka-7704-2021-05-24-8131_N-2021-06-19.pdf,"Tenaxum",0.656
Charakterystyka-9405-2023-09-11-15736_A-2023-11-10.pdf,"Actifed syrop",0.642
Charakterystyka-39985-2022-07-07-12093_B-2022-10-21.pdf,"Dipperam hct",0.638
Charakterystyka-38982-2023-05-31-15297_B-2023-10-07.pdf,"Xedine ha do",0.673
Charakterystyka-36323-20191220000000-7802_B-20200122000925.pdf,"Toralis",0.670
Charakterystyka-38349-2023-07-03-13644_N-2023-07-17.pdf,"Envil katar do",0.637
Charakterystyka-25989-2022-08-22-12143_B-2022-11-04.pdf,"Perindoprilamlopidine krka",0.613
Charakterystyka-39259-2022-02-06-11346_B-2022-05-12.pdf,"Amlessini",0.661
Charakterystyka-30378-2023-12-14-16304_B-2024-01-06.pdf,"Telmisartanhydrochlorothiazide krka",0.648
Charakterystyka-41097-2022-06-02-8521_N-2022-06-16.pdf,"Sebidin intensive do ssania",0.614
Charakterystyka-41941-2022-04-28-8573_N-2022-05-23.pdf,"Gardimax medica junior truskawkowy do ssania",0.622
Charakterystyka-40806-2023-05-25-14751_B-2023-08-23.pdf,"Rosamera",0.643
```

WYSZUKIWARKA 2

- WYKORZYSTANIE MODELU DISTILROBERTA ORAZ INDEXACJI FAISS DO OPTIMALIZACJI.
- PRZYKŁADOWE WYNIKI WYSZUKIWANIA DLA TRIACYDU

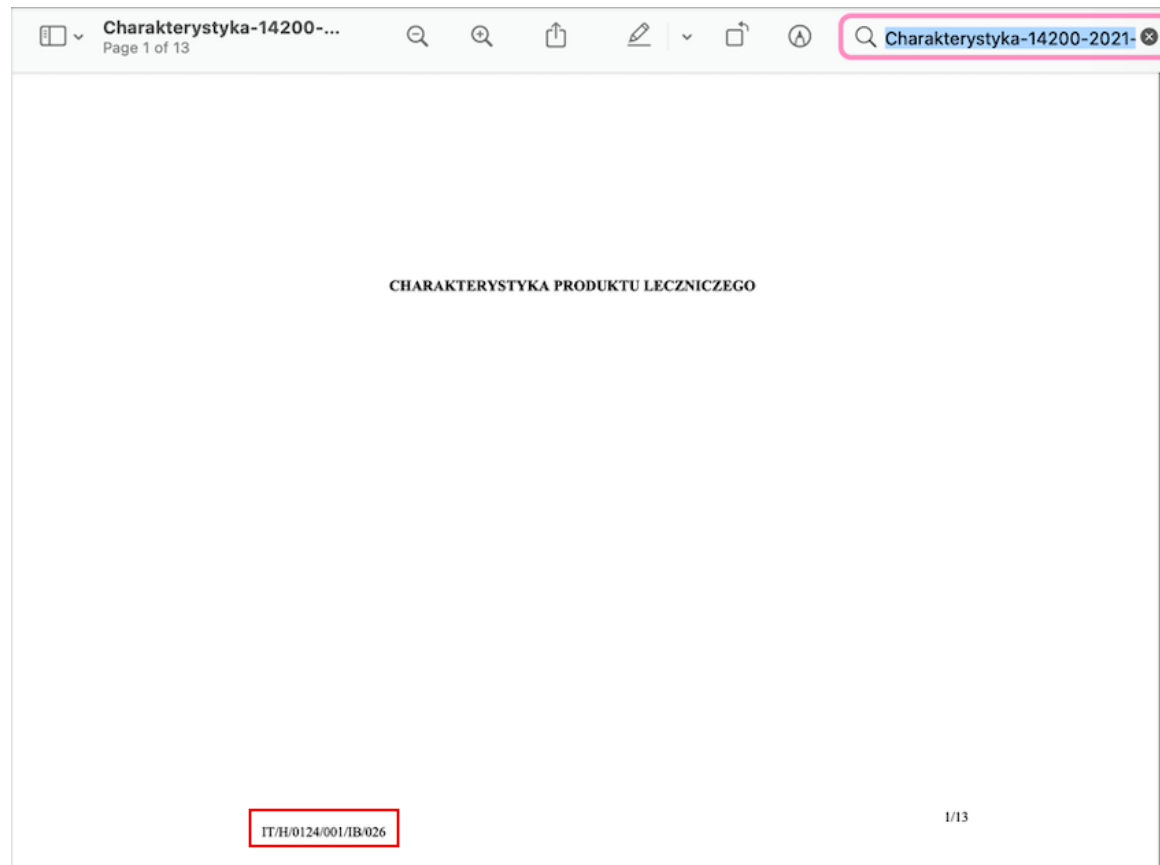
model =
SentenceTransformer('sdadas
/st-polish-paraphrase-from-
distilroberta')

Faiss (Facebook AI Similarity Search)

NAPOTKANE PROBLEMY

- Brak usunięcia części nagłówkowej, skutkowało że czasem otrzymywaliśmy takie wyniki:

```
(myenv) D:\nlp-group2-main\src>python main.py -q "astma"  
Charakterystyka-14200-2021-02-13-9574_B-2022-07-20.pdf, "Ithib ribuspir mikrogramówdawkę odmierzoną aerozol inhalacyjny roztwór",0.233
```



NAPOTKANE PROBLEMY

- Nie przemyślano czynników ryzyka, dlatego np. Leki na cukrzycę wyszukują się, ponieważ cukrzyca jest czynnikiem ryzyka choroby leczonej za pomocą danego leku a nie dlatego, że jest to lek na cukrzycę

Produkty pasujące do wskazania 'cukrzyca':

Nazwa produktu: rombidux | Plik: [Charakterystyka-42154-2020-12-14-1508_D-2023-05-02.pdf](#) | Wynik: 0.19660787643459798
Nazwa produktu: grindeks lisinopril | Plik: [Charakterystyka-46194-2023-06-26-12511_D-2023-07-14.pdf](#) | Wynik: 0.1847329401928125
Nazwa produktu: prinivil | Plik: [Charakterystyka-9408-2021-11-18-6230_N-2021-11-23.pdf](#) | Wynik: 0.15610605585983175
Nazwa produktu: synoptis rivaroxaban | Plik: [Charakterystyka-42935-2023-12-15-16277_A-2023-12-30.pdf](#) | Wynik: 0.154428893818765
Nazwa produktu: genoptim ramipril | Plik: [Charakterystyka-34173-20170224000000-2191_N-20170327000120.pdf](#) | Wynik: 0.15322812866267355
Nazwa produktu: rivaroxaban ranbaxy | Plik: [Charakterystyka-43349-2023-12-14-2109_D-2024-01-10.pdf](#) | Wynik: 0.15318809113237744
Nazwa produktu: aurobin ramipril | Plik: [Charakterystyka-22690-20190714000000-7053_B-20190806000734.pdf](#) | Wynik: 0.15190403611157893
Nazwa produktu: aporami | Plik: [Charakterystyka-34905-2022-04-01-11234_B-2022-04-16.pdf](#) | Wynik: 0.15190403611157893
Nazwa produktu: ampril | Plik: [Charakterystyka-15975-20190507000000-6766_B-20190528000822.pdf](#) | Wynik: 0.14612466129190052
Nazwa produktu: krka ramipril | Plik: [Charakterystyka-30041-20190913000000-7667_B-20191127001608.pdf](#) | Wynik: 0.1443060872839471
Nazwa produktu: oriram | Plik: [Charakterystyka-37560-20181221000000-7107_B-20190810001035.pdf](#) | Wynik: 0.1428617588231423
Nazwa produktu: captopril jelfa | Plik: [Charakterystyka-20171-20210204000000-7328_N-20210218001743.pdf](#) | Wynik: 0.13818326307461007
Nazwa produktu: dabigatran pharma etexilate kapsułki | Plik: [Charakterystyka-45664-2023-06-19-14384_B-2023-07-18.pdf](#) | Wynik: 0.1340043361011959

NAPOTKANE PROBLEMY

```
● (nlp) (base) MacBook-Pro-LILI:src LILI$ python main2.py -f "/Users/LILI/Projects_studia/projekt_grupowy  
Results for file: Charakterystyka-23642-20201203000000-9086_B-20210106000818.pdf  
Charakterystyka-23642-20201203000000-9086_B-20210106000818.pdf,"Telmisartanum teva bv",1.000  
Charakterystyka-30378-2023-12-14-16304_B-2024-01-06.pdf,"Telmisartanhydrochlorothiazide krka",0.731  
Charakterystyka-42458-2024-02-10-16822_B-2024-03-16.pdf,"Ibuprofen dr max miękkie",0.711  
Charakterystyka-38098-2023-10-16-13202_N-2023-11-10.pdf,"Ibum supermax miękkie",0.707  
Charakterystyka-24254-2022-11-28-9423_N-2022-12-12.pdf,"Ibufen junior miękkie",0.699
```

- Nieprecyzyjne wyniki wyszukiwania, które okazały się konsekwencją uwzględniania przez model w składzie również substancji pomocniczych, takich jak np. sorbitol. Problem udało się zlikwidować, po stworzeniu listy wykluczeń najczęściej stosowanych substancji pomocniczych

2. SKŁAD JAKOŚCIOWY I ILOŚCIOWY

Każda tabletka zawiera 40 mg telmisartanu.

Substancje pomocnicze o znanym działaniu:
Każda tabletka zawiera 19,2 mg sorbitolu.

2. SKŁAD JAKOŚCIOWY I ILOŚCIOWY

Każda tabletka zawiera 400 mg ibuprofenu.

Substancje pomocnicze o znanym działaniu
Każda kapsułka zawiera 50 mg sorbitolu (E 420).