

Nama : Raihan Rahmanda Junianto

NIM : 222112303

Kelas : 3SD2

Responsi Praktikum 2 Information Retrieval

Praktikum kali ini membahas mengenai term-vocabulary serta macam-macam teks processing, seperti tokenisasi, capitalization/case-folding, eliminasi stopword, normalisasi, dan stemming. Pada praktikum ini juga dibahas mengenai library-library yang digunakan untuk melakukan pemrosesan tersebut. Berikut disajikan jawaban dari masing-masing nomor penugasan yang terdapat pada modul 2 praktikum Information Retrieval.

- 1. Buat kode fungsi python untuk memisahkan dokumen sehingga menghasilkan variabel `list_paragraf` (nama fungsi: `paragraph_parsing`), dan masing-masing `paragraph` menjadi variabel `list_kalimat` (nama fungsi: `sentence_parsing`).**

Langkah pertama yang dilakukan untuk menjawab soal di atas adalah mendefinisikan dokumen/teks yang dimaksud. Pada kasus kali ini, setiap paragraf dalam dokumen tersebut dipisahkan oleh enter sebanyak dua kali. Hal tersebut bertujuan untuk memudahkan penulisan regular expressionnya karena paragraf yang didefinisikan tidak berada dalam satu baris lurus.

```
1  doctask = '''Mobilitas warga bakal diperketat melalui penerapan
2  PPKM level 3 se-Indonesia di masa libur Natal dan
3  tahun baru (Nataru). Rencana kebijakan itu dikritik
4  oleh Epidemiolog dari Griffith University Dicky
5  Budiman.
6
7  Dicky menyebut pembatasan mobilitas memang akan
8  memiliki dampak dalam mencegah penularan COVID-19.
9  Tapi, kata dia, dampaknya signifikan atau tidak akan
10 bergantung pada konsistensi yang mendasar yakni
11 testing, tracing, treatment (3T) hingga vaksinasi
12 COVID-19.'''
```

Proses selanjutnya adalah mendefinisikan fungsi yang akan digunakan pada program, yaitu fungsi `paragraph_parsing` dan `sentence_parsing`.

```

14 # fungsi untuk memisahkan paragraf dalam suatu dokumen
15 def paragraph_parsing(text):
16     paragraphs = text.split("\n\n")
17     return paragraphs
18
19 # fungsi untuk memisahkan kalimat dalam paragraf hasil parsing
20 def sentence_parsing(text):
21     sentences = text.split(". ")
22     return sentences

```

Pada fungsi `paragraph_parsing` di atas, setiap paragraf dalam dokumen akan dipisahkan (split) berdasarkan enter sebanyak dua kali yang dituliskan dengan regular expression “\n\n”. Sedangkan, pada fungsi `sentence_parsing`, setiap kalimat dalam paragraf akan dipisahkan berdasarkan tanda titik.

Selanjutnya, proses akan berlanjut ke dalam pemanggilan kedua fungsi tersebut dan mencetak output sesuai dengan yang tertera pada modul.

```

25 list_paragraph = paragraph_parsing(doctask)
26 print("List paragraf:")
27 for i in range(len(list_paragraph)):
28     print("P", i+1, ":", list_paragraph[i])
29     print("\n")
30
31 print("=====\n")
32
33 for i in range(len(list_paragraph)):
34     list_sentence = sentence_parsing(list_paragraph[i])
35     print("List kalimat pada paragraf ke-", i+1, ":\n")
36     for j in range(len(list_sentence)):
37         print("S", j+1, ":", list_sentence[j])
38         print("\n")

```

Sintaks di atas akan menampilkan output berupa masing-masing paragraf yang terkandung dalam dokumen serta kalimat-kalimat pada masing-masing paragraf tersebut. Berikut disajikan tampilan program secara keseluruhan beserta outputnya.

```

1 doctask = '''Mobilitas warga bakal diperketat melalui penerapan
2 PPKM level 3 se-Indonesia di masa libur Natal dan
3 tahun baru (Nataru). Rencana kebijakan itu dikritik
4 oleh Epidemiolog dari Griffith University Dicky
5 Budiman.
6
7 Dicky menyebut pembatasan mobilitas memang akan
8 memiliki dampak dalam mencegah penularan COVID-19.
9 Tapi, kata dia, dampaknya signifikan atau tidak akan
10 bergantung pada konsistensi yang mendasar yakni
11 testing, tracing, treatment (3T) hingga vaksinasi
12 COVID-19.'''
13
14 # fungsi untuk memisahkan paragraf dalam suatu dokumen
15 def paragraph_parsing(text):
16     paragraphs = text.split("\n\n")
17     return paragraphs
18
19 # fungsi untuk memisahkan kalimat dalam paragraf hasil parsing
20 def sentence_parsing(text):
21     sentences = text.split(". ")
22     return sentences
23
24
25 list_paragraph = paragraph_parsing(doctask)
26 print("List paragraf:")
27 for i in range(len(list_paragraph)):
28     print("P", i+1, ":", list_paragraph[i])
29     print("\n")
30
31 print("=====\n")
32
33 for i in range(len(list_paragraph)):
34     list_sentence = sentence_parsing(list_paragraph[i])
35     print("List kalimat pada paragraf ke-", i+1, ":\n")
36     for j in range(len(list_sentence)):
37         print("S", j+1, ":", list_sentence[j])
38     print("\n")

```

```

Anaconda Prompt
(base) D:\RAIHAN STIS\Perkuliah\SEMESTER 5\Praktikum INFORMATION RETRIEVAL\Pertemuan (2)>python penugasan2_1.py
List paragraf:
P 1 : Mobilitas warga bakal diperketat melalui penerapan
PPKM level 3 se-Indonesia di masa libur Natal dan
tahun baru (Nataru). Rencana kebijakan itu dikritik
oleh Epidemiolog dari Griffith University Dicky
Budiman.

P 2 : Dicky menyebut pembatasan mobilitas memang akan
memiliki dampak dalam mencegah penularan COVID-19.
Tapi, kata dia, dampaknya signifikan atau tidak akan
bergantung pada konsistensi yang mendasar yakni
testing, tracing, treatment (3T) hingga vaksinasi
COVID-19.

=====

```

```

List kalimat pada paragraf ke- 1 :

S 1 : Mobilitas warga bakal diperketat melalui penerapan
PPKM level 3 se-Indonesia di masa libur Natal dan
tahun baru (Nataru)

S 2 : Rencana kebijakan itu dikritik
oleh Epidemiolog dari Griffith University Dicky
Budiman.

List kalimat pada paragraf ke- 2 :

S 1 : Dicky menyebut pembatasan mobilitas memang akan
memiliki dampak dalam mencegah penularan COVID-19

S 2 :
Tapi, kata dia, dampaknya signifikan atau tidak akan
bergantung pada konsistensi yang mendasar yakni
testing, tracing, treatment (3T) hingga vaksinasi
COVID-19.

```

2. Lakukan case-folding, tokenisasi, eliminasi stopword dan stemming pada dokumen di folder “berita” menggunakan library yang sudah tersedia (nltk, spacy, sastrawi, etc).

- Mendefinisikan library/modul yang digunakan

```

1  # Import Module
2  import os
3  from spacy.lang.id import Indonesian
4  import spacy
5  from spacy.lang.id.stop_words import STOP_WORDS
6  from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
7

```

Pada program kali ini, modul yang digunakan adalah modulo os, spacy, dan sastrawi. **Import os** berfungsi untuk mengimpor modul os yang digunakan untuk berinteraksi dengan sistem operasi, seperti mengakses file dan direktori. Sedangkan, **from spacy.lang.id import Indonesian** berfungsi untuk mengimpor modul Indonesian dari paket spacy.lang.id. Modul ini menyediakan dukungan untuk pemrosesan bahasa Indonesia menggunakan library spaCy. Selanjutnya, **import spacy** digunakan mengimpor modul spacy yang merupakan library utama untuk pemrosesan bahasa alami. Lalu, **from spacy.lang.id.stop_words import STOP_WORDS** digunakan untuk mengimpor variabel STOP_WORDS dari modul stop_words yang terdapat dalam paket spacy.lang.id. Variabel ini berisi kumpulan kata-kata yang umumnya dianggap sebagai kata penghubung atau kata yang tidak memiliki makna penting dalam bahasa Indonesia. Dan yang terakhir, **from Sastrawi.Stemmer.StemmerFactory import StemmerFactory** berfungsi untuk mengimpor kelas StemmerFactory dari modul StemmerFactory yang terdapat dalam paket Sastrawi. Kelas ini digunakan untuk membuat objek stemmer yang digunakan untuk melakukan stemming pada kata-kata dalam bahasa Indonesia.

- Mendefinisikan fungsi dan lokasi folder berita

```

8  nlp = Indonesian() # Load Indonesian language model
9
10 # Folder Path
11 path = "D:/RAIHAN STIS/Perkuliahan/SEMESTER 5/Praktikum INFORMATION RETRIEVAL/Pertemuan (2)/berita"
12
13 # Function to read text file
14 def read_text_file(file_path):
15     with open(file_path, 'r') as f:
16         content = f.read()
17     return content

```

- Menampilkan seluruh list file yang terdapat pada folder berita dan menampilkan isi dari masing-masing file tersebut.

```

20 print("\n=====\\n")
21 print("Isi dari folder berita")
22 print("\n=====\\n")
23 # List all files in a directory using os.listdir
24 for file in os.listdir(path):
25     if os.path.isfile(os.path.join(path, file)):
26         print(file)
27
28
29 print("\n=====\\n")
30 print("isi dari berita.txt")
31 print("\n=====\\n")
32 # iterate through all file
33 for file in os.listdir(path):
34     # Check whether file is in text format or not
35     if file.endswith(".txt"):
36         #print kalimat berita ke-i
37         print("isi dari", file, ":")
38         file_path = f"{path}\\{file}"
39         # call read text file function
40         result1 = read_text_file(file_path)
41         print(result1)
42         print("\\n")
43

```

- Case Folding masing-masing isi dari folder berita

```

44 print("\n===== \n")
45 print("Case Folding masing-masing isi dari folder berita")
46 print("\n===== \n")
47 # iterate through all file
48 for file in os.listdir(path):
49     # Check whether file is in text format or not
50     if file.endswith(".txt"):
51         #print kalimat berita ke-i
52         print("isi dari", file, ":")
53         file_path = f"{path}\\{file}"
54         # call read text file function
55         result2 = read_text_file(file_path)
56         print("teks original:")
57         print(result2)
58         print("\n")
59         print("teks case folding (lowercase):")
60         print(result2.lower())
61         print("\n")
62         print("teks case folding (uppercase):")
63         print(result2.upper())
64         print("\n")

```

- Tokenisasi masing-masing isi dari folder berita

```

66 print("\n===== \n")
67 print("Tokenisasi masing-masing isi dari folder berita")
68 print("\n===== \n")
69 def tokenize_file(file_path):
70     with open(file_path, 'r', encoding='utf-8') as file:
71         content = file.read()
72         doc = nlp(content)
73         tokens = [token.text for token in doc]
74         return tokens
75
76 for filename in os.listdir(path):
77     if filename.endswith(".txt"):
78         file_path = os.path.join(path, filename)
79         tokens = tokenize_file(file_path)
80         print(f"Tokens in {filename}: {len(tokens)}")
81         print(tokens)

```

- Stopword elimination masing-masing isi dari folder berita

```

83 print("\n===== \n")
84 print("Stopword elimination masing-masing isi dari folder berita")
85 print("\n===== \n")
86 def eliminate_stopwords(tokens):
87     cleaned_tokens = [token for token in tokens if token.lower() not in STOP_WORDS]
88     return cleaned_tokens
89
90 for filename in os.listdir(path):
91     if filename.endswith(".txt"):
92         file_path = os.path.join(path, filename)
93         tokens = tokenize_file(file_path)
94         tokens_without_stopwords = eliminate_stopwords(tokens)
95
96         print(f"Tokens in {filename} after stopwords elimination: {len(tokens_without_stopwords)}")
97         print(tokens_without_stopwords)
98         print("-" * 60) # Add a separator between outputs

```

- Stemming masing-masing isi dari folder berita

```
101 print("\n===== \n")
102 print("Stemming masing-masing isi dari folder berita")
103 print("\n===== \n")
104 stemmer = StemmerFactory().create_stemmer()
105
106 for file in os.listdir(path):
107     # Check whether file is in text format or not
108     if file.endswith(".txt"):
109         #print kalimat berita ke-i
110         print("isi dari", file, "setelah stemming:")
111         file_path = f"{path}\\{file}"
112         # call read text file function
113         result3 = read_text_file(file_path)
114         print(stemmer.stem(result3))
115         print("\n")
```

- Sintaks program secara keseluruhan serta outputnya

```
1 # Import Module
2 import os
3 from spacy.lang.id import Indonesian
4 import spacy
5 from spacy.lang.id.stop_words import STOP_WORDS
6 from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
7
8 nlp = Indonesian() # Load Indonesian language model
9
10 # Folder Path
11 path = "D:/RAINAN STIS/Perkuliahan/SEMESTER 5/Praktikum INFORMATION RETRIEVAL/Pertemuan (2)/berita"
12
13 # Function to read text file
14 def read_text_file(file_path):
15     with open(file_path, 'r') as f:
16         content = f.read()
17         # content = content.replace('\n', ' ') # Replace newline with space
18     return content
19
20 print("\n=====")
21 print("Isi dari folder berita")
22 print("\n=====")
23 # List all files in a directory using os.listdir
24 for file in os.listdir(path):
25     if os.path.isfile(os.path.join(path, file)):
26         print(file)
27
28
29 print("\n=====")
30 print("Isi dari berita.txt")
31 print("\n=====")
32 # Iterate through all file
33 for file in os.listdir(path):
34     # Check whether file is in text format or not
35     if file.endswith(".txt"):
36         # Print kalimat berita ke-1
37         print("Isi dari", file, ":")
38         file_path = f"{path}\\{file}"
39         # Call read text file function
40         result1 = read_text_file(file_path)
41         print(result1)
42         print("\n")
43
44 print("\n=====")
45 print("Case Folding masing-masing isi dari folder berita")
46 print("\n=====")
47 # Iterate through all file
48 for file in os.listdir(path):
49     # Check whether file is in text format or not
50     if file.endswith(".txt"):
51         # Print kalimat berita ke-1
52         print("Isi dari", file, ":")
53         file_path = f"{path}\\{file}"
54         # Call read text file function
55         result2 = read_text_file(file_path)
56         print("Teks Original:")
57         print(result2)
58         print("\n")
59         print("Teks case folding (lowercase):")
60         print(result2.lower())
61         print("\n")
62         print("Teks case folding (uppercase):")
63         print(result2.upper())
64         print("\n")
65
66 print("\n=====")
67 print("Tokenisasi masing-masing isi dari folder berita")
68 print("\n=====")
69 def tokenize_file(file_path):
70     with open(file_path, 'r', encoding="utf-8") as file:
71         content = file.read()
72         doc = nlp(content)
73         tokens = [token.text for token in doc]
74         return tokens
75
76 for filename in os.listdir(path):
77     if filename.endswith(".txt"):
78         file_path = os.path.join(path, filename)
79         tokens = tokenize_file(file_path)
80         print(f"Tokens in {filename}: {len(tokens)}")
81         print(tokens)
82
83 print("\n=====")
84 print("Stopword elimination masing-masing isi dari folder berita")
85 print("\n=====")
86 def eliminate_stopwords(tokens):
87     cleaned_tokens = [token for token in tokens if token.lower() not in STOP_WORDS]
88     return cleaned_tokens
89
90 for filename in os.listdir(path):
91     if filename.endswith(".txt"):
92         file_path = os.path.join(path, filename)
93         tokens = tokenize_file(file_path)
94         tokens_without_stopwords = eliminate_stopwords(tokens)
95
96         print(f"Tokens in {filename} after stopwords elimination: {len(tokens_without_stopwords)}")
97         print(tokens_without_stopwords)
98         print("-" * 60) # Add a separator between outputs
99
100 print("\n=====")
101 print("Stemming masing-masing isi dari folder berita")
102 print("\n=====")
103 stemmer = StemmerFactory().create_stemmer()
104
105 for file in os.listdir(path):
106     # Check whether file is in text format or not
107     if file.endswith(".txt"):
108         # Print kalimat berita ke-1
109         print("Isi dari", file, "setelah stemming:")
110         file_path = f"{path}\\{file}"
111         # Call read text file function
112         result3 = read_text_file(file_path)
113         print(stemmer.stem(result3))
114         print("\n")
```



```

Anaconda Prompt x + v

(base) D:\RAIHAN STIS\Perkuliahan\SEMESTER 5\Praktikum INFORMATION RETRIEVAL\Pertemuan (2)>python penugasan2_2.py

=====

Isi dari folder berita

=====

berital.txt
berita2.txt
berita3.txt
berita4.txt
berita5.txt

=====

isi dari berita.txt

=====

isi dari berital.txt :
Wilayah Kamu Sudah 'Bebas' COVID-19? Cek 34 Kab/Kota Zona Hijau Terbaru

Jakarta - Pemerintah rencananya bakal menerapkan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) level 3 terhitung 24 Desember 2021 hingga 2 Januari 2022. Namun, Kementerian Kesehatan RI memastikan kebijakan PPKM level 3 ini masih dalam tahap kajian. Menurut Direktur Pencegahan dan Pengendalian Penyakit Menular Langsung (P2PML) Kemenkes RI dr Siti Nadia Tarmizi, PPKM level 3 bakal diterapkan jika kasus COVID-19 naik signifikan. Hal ini umumnya dipicu dengan peningkatan mobilitas dan pelanggaran protokol kesehatan.

https://health.detik.com/berita-detikhealth/d-5816690/wilayah-kamu-sudah-bebas-covid-19-cek-34-kabkota-zona-hijau-terbaru

isi dari berita2.txt :
Vaksin COVID-19 Bakal Rutin Setiap Tahun? Tergantung, Ini Penjelasan

Jakarta - Pemberian booster atau dosis ketiga vaksin COVID-19 di Indonesia direncanakan bakal berlangsung Januari 2022. Lantas adakah kemungkinan vaksinasi COVID-19 bakal harus dilakukan setiap tahun seperti vaksinasi influenza? Ketua Satgas COVID-19 Ikatan Dokter Indonesia (IDI) Prof Zubairi Djoerban menjelaskan hingga kini belum ada kepastian terkait hal tersebut. Menurutnya masih ada kemungkinan vaksin COVID-19 harus diberikan setiap tahun, ada juga kemungkinan cukup booster diberikan sekali kemudian vaksinasi COVID-19 tidak diperlukan lagi.

https://health.detik.com/berita-detikhealth/d-5816582/vaksin-covid-19-bakal-rutin-setiap-tahun-tergantung-ini-penjelasan

isi dari berita3.txt :
RI Mulai Suntikkan Booster di 2022, Masihkah Ampuh Lawan Varian Delta Cs?

Jakarta - Pakar mengakui vaksin-vaksin yang sudah digunakan untuk dosis 1-2 memang mengalami penurunan efektivitas terhadap varian baru Corona seperti varian Delta. Mengingat booster atau dosis ketiga vaksin COVID-19 di Indonesia disebut bakal dimulai awal 2022, apakah jenis vaksin yang digunakan bakal mengikuti strain virus terbaru? Menjawab pertanyaan tersebut, Ketua Satgas COVID-19 Ikatan Dokter Indonesia (IDI) Prof Zubairi Djoerban kembali menyinggung riset yang sudah berlangsung terkait efektivitas vaksin COVID-19 dosis 1 dan 2. Ia menyebut berdasarkan riset sejauh ini, efektivitas vaksin COVID-19 Pfizer dan Moderna terbukti menurun dalam melawan varian Delta.

https://health.detik.com/berita-detikhealth/d-5816534/ri-mulai-suntikkan-booster-di-2022-masihkah-ampuh-lawan-varian-delta-cs

isi dari berita4.txt :
Alert! Kasus Varian Delta COVID-19 di DKI Meningkat

Jakarta - Data terbaru dari Balitbangkes Kemenkes RI per 13 November, menunjukkan adanya penambahan varian Delta. Penambahan tersebut terjadi di Jawa Barat ada 165 kasus, DKI Jakarta 90 kasus, dan Sulawesi Utara 86 kasus. Dalam satu bulan terakhir, Balitbangkes menyebutkan DKI Jakarta mengalami peningkatan kasus varian Delta yang signifikan. Sementara, pada varian baru seperti varian Alpha, varian Delta, hingga Beta terbanyak di Indonesia berasal dari DKI Jakarta, dengan total 1.327 kasus.

https://health.detik.com/berita-detikhealth/d-5812940/alert-kasus-varian-delta-covid-19-di-dki-meningkat

isi dari berita5.txt :
Corona di AS Mendadak Naik Lagi Usai Serangan Delta Sempat Mereda

Jakarta - Kasus COVID-19 kembali naik di sejumlah wilayah Amerika Serikat (AS). Padahal diketahui, COVID-19 sempat tercatat stabil pasca serangan varian Delta musim panas ini. Ada apa? Hal tersebut disampaikan oleh kepala penasihat medis Gedung Putih Dr. Anthony Fauci, Senin (15/11/2021). Diketahui, kasus nasional turun 57 persen minggu lalu dari puncak gelombang varian Delta pada musim panas. Namun jumlah pasien COVID-19 di area Barat Tengah dan Timur laut kini naik mendadak.

https://health.detik.com/berita-detikhealth/d-5813949/corona-di-as-mendadak-naik-lagi-usai-serangan-delta-sempat-mereda

=====

Case Folding masing-masing isi dari folder berita

=====

isi dari berital.txt :
teks original:
Wilayah Kamu Sudah 'Bebas' COVID-19? Cek 34 Kab/Kota Zona Hijau Terbaru

Jakarta - Pemerintah rencananya bakal menerapkan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) level 3 terhitung 24 Desember 2021 hingga 2 Januari 2022. Namun, Kementerian Kesehatan RI memastikan kebijakan PPKM level 3 ini masih dalam tahap kajian. Menurut Direktur Pencegahan dan Pengendalian Penyakit Menular Langsung (P2PML) Kemenkes RI dr Siti Nadia Tarmizi, PPKM level 3 bakal diterapkan jika kasus COVID-19 naik signifikan. Hal ini umumnya dipicu dengan peningkatan mobilitas dan pelanggaran protokol kesehatan.

```

n.

<https://health.detik.com/berita-detikhealth/d-5816690/wilayah-kamu-sudah-bebas-covid-19-cek-34-kabkota-zona-hijau-terbaru>

teks case folding (lowercase):
wilayah kamu sudah 'bebas' covid-19? cek 34 kab/kota zona hijau terbaru

Jakarta - pemerintah rencananya bakal menerapkan pemberlakuan pembatasan kegiatan masyarakat (ppkm) level 3 terhitung 24 desember 2021 hingga 2 januari 2022. namun, kementerian kesehatan RI memastikan kebijakan ppkm level 3 ini masih dalam tahap kajian. menurut direktur pencegahan dan pengendalian penyakit menular langsung (p2pml) kemenkes RI dr siti nadia tarmizi, ppkm level 3 bakal diterapkan jika kasus covid-19 naik signifikan. hal ini umumnya dipicu dengan peningkatan mobilitas dan pelanggaran protokol kesehatan.

<https://health.detik.com/berita-detikhealth/d-5816690/wilayah-kamu-sudah-bebas-covid-19-cek-34-kabkota-zona-hijau-terbaru>

teks case folding (uppercase):
WILAYAH KAMU SUDAH 'BEBAS' COVID-19? CEK 34 KAB/KOTA ZONA HIJAU TERBARU

JAKARTA - PEMERINTAH RENCANANYA BAKAL MENERAPKAN PEMBERLAKUAN PEMBATAAN KEGIATAN MASYARAKAT (PPKM) LEVEL 3 TERHITUNG 24 DESEMBER 2021 HINGGA 2 JANUARI 2022. NAMUN, KEMENTERIAN KESEHATAN RI MEMASTIKAN KEBIJAKAN PPKM LEVEL 3 INI MASIH DALAM TAHAP KAJIAN. MENURUT DIREKTUR PENCEGAHAN DAN PENGENDALIAN PENYAKIT MENULAR LANGSUNG (P2PML) KEMENKES RI DR SITI NADIA TARMIZI, PPKM LEVEL 3 BAKAL DITERAPKAN JIKA KASUS COVID-19 NAIK SIGNIFIKAN. HAL INI UMUMNYA DIPICU DENGAN PENINGKATAN MOBILITAS DAN PELANGGARAN PROTOKOL KESEHATAN.

[HTTPS://HEALTH.DETIK.COM/BERITA-DETIKHEALTH/D-5816690/WILAYAH-KAMU-SUDAH-BEBAS-COVID-19-CEK-34-KABKOTA-ZONA-HIJAU-TERBARU](https://health.detik.com/berita-detikhealth/d-5816690/wilayah-kamu-sudah-bebas-covid-19-cek-34-kabkota-zona-hijau-terbaru)

isi dari berita2.txt :
teks original:
Vaksin COVID-19 Bakal Rutin Setiap Tahun? Tergantung, Ini Penjelasannya

Jakarta - Pemberian booster atau dosis ketiga vaksin COVID-19 di Indonesia direncanakan bakal berlangsung Januari 2022. Lantas adakah kemungkinan vaksinasi COVID-19 bakal harus dilakukan setiap tahun seperti vaksinasi influenza?

Ketua Satgas COVID-19 Ikatan Dokter Indonesia (IDI) Prof Zubairi Djoerban menjelaskan hingga kini belum ada kepastian terkait hal tersebut. Menurutnya masih ada kemungkinan vaksin COVID-19 harus diberikan setiap tahun, ada juga kemungkinan cukup booster diberikan sekali kemudian vaksinasi COVID-19 tidak diperlukan lagi.

<https://health.detik.com/berita-detikhealth/d-5816582/vaksin-covid-19-bakal-rutin-setiap-tahun-tergantung-ini-penjelasannya>

teks case folding (lowercase):
vaksin covid-19 bakal rutin setiap tahun? tergantung, ini penjelasannya

Jakarta - pemberian booster atau dosis ketiga vaksin covid-19 di indonesia direncanakan bakal berlangsung januari 2022. lantas adakah kemungkinan vaksinasi covid-19 bakal harus dilakukan setiap tahun seperti vaksinasi influenza?

Ketua Satgas COVID-19 Ikatan Dokter Indonesia (IDI) Prof Zubairi Djoerban menjelaskan hingga kini belum ada kepastian terkait hal tersebut. Menurutnya masih ada kemungkinan vaksin covid-19 harus diberikan setiap tahun, ada juga kemungkinan cukup booster diberikan sekali kemudian vaksinasi covid-19 tidak diperlukan lagi.

<https://health.detik.com/berita-detikhealth/d-5816582/vaksin-covid-19-bakal-rutin-setiap-tahun-tergantung-ini-penjelasannya>

teks case folding (uppercase):
VAKSIN COVID-19 BAKAL RUTIN SETIAP TAHUN? TERGANTUNG, INI PENJELASANNYA

JAKARTA - PEMBERIAN BOOSTER ATAU DOSIS KETIGA VAKSIN COVID-19 DI INDONESIA DIRENCANAKAN BAKAL BERLANGSUNG JANUARI 2022. LANTAS ADAKAH KEMUNGKINAN VAKSINASI COVID-19 BAKAL HARUS DILAKUKAN SETIAP TAHUN SEPERTI VAKSINASI INFLUENZA?

KETUA SATGAS COVID-19 IKATAN DOKTER INDONESIA (IDI) PROF ZUBAIRI DJOERBAN MENJELASKAN HINGGA KINI BELUM ADA KEPASTIAN TERKAIT HAL TERSEBUT. MENURUTNYA MASIH ADA KEMUNGKINAN VAKSIN COVID-19 HARUS DIBERIKAN SETIAP TAHUN, ADA JUGA KEMUNGKINAN CUKUP BOOSTER DIBERIKAN SEKALI KEMUDIAN VAKSINASI COVID-19 TIDAK DIPERLUKAN LAGI.

[HTTPS://HEALTH.DETIK.COM/BERITA-DETIKHEALTH/D-5816582/VAKSIN-COVID-19-BAKAL-RUTIN-SETIAP-TAHUN-TERGANTUNG-INI-PENJELASANNYA](https://health.detik.com/berita-detikhealth/d-5816582/vaksin-covid-19-bakal-rutin-setiap-tahun-tergantung-ini-penjelasannya)

isi dari berita3.txt :
teks original:
RI Mulai Suntikkan Booster di 2022, Masihkah Ampuh Lawan Varian Delta Cs?

Jakarta - Pakar mengakui vaksin-vaksin yang sudah digunakan untuk dosis 1-2 memang mengalami penurunan efektivitas terhadap varian baru Corona seperti varian Delta. Mengingat booster atau dosis ketiga vaksin COVID-19 di Indonesia disebut bakal dimulai awal 2022, apakah jenis vaksin yang digunakan bakal mengikuti strain virus terbaru?

Menjawab pertanyaan tersebut, Ketua Satgas COVID-19 Ikatan Dokter Indonesia (IDI) Prof Zubairi Djoerban kembali menyinggung riset yang sudah berlangsung terkait efektivitas vaksin COVID-19 dosis 1 dan 2. Ia menyebut berdasarkan riset sejauh ini, efektivitas vaksin COVID-19 Pfizer dan Moderna terbukti menurun dalam melawan varian Delta.

<https://health.detik.com/berita-detikhealth/d-5816534/ri-mulai-suntikkan-booster-di-2022-masihkah-ampuh-lawan-varian-delta-cs>

teks case folding (lowercase):
ri mulai suntikkan booster di 2022, masihkah ampuh lawan varian delta cs?

Jakarta - pakar mengakui vaksin-vaksin yang sudah digunakan untuk dosis 1-2 memang mengalami penurunan efektivitas terhadap varian baru Corona seperti varian delta. mengingat booster atau dosis ketiga vaksin covid-19 di indonesia disebut bakal dimulai awal 2022, apakah jenis vaksin yang digunakan bakal mengikuti strain virus terbaru?

menjawab pertanyaan tersebut, ketua satgas covid-19 ikatan dokter indonesia (idi) prof zubairi djoerban kembali menyinggung riset yang sudah berlangsung terkait efektivitas vaksin covid-19 dosis 1 dan 2. ia menyebut berdasarkan riset sejauh ini, efektivitas vaksin covid-19 pfizer dan moderna terbukti menurun dalam melawan varian delta.

<https://health.detik.com/berita-detikhealth/d-5816534/ri-mulai-suntikkan-booster-di-2022-masihkah-ampuh-lawan-varian-delta-cs>

teks case folding (uppercase):
RI MULAI SUNTIKKAN BOOSTER DI 2022, MASIIKHAH AMPUH LAWAN VARIAN DELTA CS?

JAKARTA - PAKAR MENGAKUI VAKSIN-VAKSIN YANG SUDAH DIGUNAKAN UNTUK DOSIS 1-2 MEMANG MENGALAMI PENURUNAN EFEKTIVITAS TERHADAP VARIAN BARU

RU CORONA SEPERTI VARIAN DELTA. MENINGAT BOOSTER ATAU DOSIS KETIGA VAKSIN COVID-19 DI INDONESIA DISEBUT BAKAL DIMULAI AWAL 2022, APA KAH JENIS VAKSIN YANG DIGUNAKAN BAKAL MENGIKUTI STRAIN VIRUS TERBARU?
MENJAWAB PERTANYAAN TERSEBUT, KETUA SATGAS COVID-19 IKATAN DOKTER INDONESIA (IDI) PROF ZUBATRI DJOERBAN KEMBALI MENYINGGUNG RISET YANG SUDAH BERLANGSUNG TERKAIT EFEKTIVITAS VAKSIN COVID-19 DOSIS 1 DAN 2. IA MENYEBUT BERDASARKAN RISET SEJAUH INI, EFEKTIVITAS VAKSIN COVID-19 PFIZER DAN MODERNA TERBUKTI MENURUN DALAM MELAWAN VARIAN DELTA.

[HTTPS://HEALTH.DETIK.COM/BERITA-DETIKHEALTH/D-5816534/RI-MULAI-SUNTIKKAN-BOOSTER-DI-2022-MASIHKAH-AMPUH-LAWAN-VARIAN-DELTA-CS](https://health.detik.com/berita-detikhealth/d-5816534/ri-mulai-suntikkan-booster-di-2022-masihkah-ampuh-lawan-varian-delta-cs)

isi dari berita4.txt :
teks original:
Alert! Kasus Varian Delta COVID-19 di DKI Meningkat

Jakarta - Data terbaru dari Balitbangkes Kemenkes RI per 13 November, menunjukkan adanya penambahan varian Delta. Penambahan tersebut terjadi di Jawa Barat ada 165 kasus, DKI Jakarta 90 kasus, dan Sulawesi Utara 86 kasus. Dalam satu bulan terakhir, Balitbangkes menyebutkan DKI Jakarta mengalami peningkatan kasus varian Delta yang signifikan. Sementara, pada varian baru seperti varian Alpha, varian Delta, hingga Beta terbanyak di Indonesia berasal dari DKI Jakarta, dengan total 1.327 kasus.

<https://health.detik.com/berita-detikhealth/d-5812940/alert-kasus-varian-delta-covid-19-di-dki-meningkat>

teks case folding (lowercase):
alert! kasus varian delta covid-19 di dki meningkat

jakarta - data terbaru dari balitbangkes kemenkes ri per 13 november, menunjukkan adanya penambahan varian delta. penambahan tersebut terjadi di jawa barat ada 165 kasus, dki jakarta 90 kasus, dan sulawesi utara 86 kasus. dalam satu bulan terakhir, balitbangkes menyebutkan dki jakarta mengalami peningkatan kasus varian delta yang signifikan. sementara, pada varian baru seperti varian alpha, varian delta, hingga beta terbanyak di indonesia berasal dari dki jakarta, dengan total 1.327 kasus.

<https://health.detik.com/berita-detikhealth/d-5812940/alert-kasus-varian-delta-covid-19-di-dki-meningkat>

teks case folding (uppercase):
ALERT! KASUS VARIAN DELTA COVID-19 DI DKI MENINGKAT

JAKARTA - DATA TERBARU DARI BALITBANGKES KEMENKES RI PER 13 NOVEMBER, MENUNJUKKAN ADANYA PENAMBAHAN VARIAN DELTA. PENAMBAHAN TERSEBUT TERJADI DI JAWA BARAT ADA 165 KASUS, DKI JAKARTA 90 KASUS, DAN SULAWESI UTARA 86 KASUS. DALAM SATU BULAN TERAKHIR, BALITBANGKES MENYEBUTKAN DKI JAKARTA MENGALAMI PENINGKATAN KASUS VARIAN DELTA YANG SIGNIFIKAN. SEMENTARA, PADA VARIAN BARU SEPERTI VARIAN ALPHA, VARIAN DELTA, HINGGA BETA TERBANYAK DI INDONESIA BERASAL DARI DKI JAKARTA, DENGAN TOTAL 1.327 KASUS.

[HTTPS://HEALTH.DETIK.COM/BERITA-DETIKHEALTH/D-5812940/ALERT-KASUS-VARIAN-DELTA-COVID-19-DI-DKI-MENINGKAT](https://health.detik.com/berita-detikhealth/d-5812940/alert-kasus-varian-delta-covid-19-di-dki-meningkat)

isi dari berita5.txt :
teks original:
Corona di AS Mendadak Naik Lagi Usai Serangan Delta Sempat Mereda

Jakarta - Kasus COVID-19 kembali naik di sejumlah wilayah Amerika Serikat (AS). Padahal diketahui, COVID-19 sempat tercatat stabil pada serangan varian Delta musim panas ini. Ada apa? Hal tersebut disampaikan oleh kepala penasihat medis Gedung Putih Dr. Anthony Fauci, Senin (15/11/2021). Diketahui, kasus nasional turun 57 persen minggu lalu dari puncak gelombang varian Delta pada musim panas. Namun jumlah pasien COVID-19 di area Barat Tengah dan Timur laut kini naik mendadak.

<https://health.detik.com/berita-detikhealth/d-5813949/corona-di-as-mendadak-naik-lagi-usai-serangan-delta-sempt-mereda>

teks case folding (lowercase):
corona di as mendadak naik lagi usai serangan delta sempat mereda

jakarta - kasus covid-19 kembali naik di sejumlah wilayah amerika serikat (as). padahal diketahui, covid-19 sempat tercatat stabil pada serangan varian delta musim panas ini. ada apa? hal tersebut disampaikan oleh kepala penasihat medis gedung putih dr. anthony fauci, senin (15/11/2021). diketahui, kasus nasional turun 57 persen minggu lalu dari puncak gelombang varian delta pada musim panas. namun jumlah pasien covid-19 di area barat tengah dan timur laut kini naik mendadak.

<https://health.detik.com/berita-detikhealth/d-5813949/corona-di-as-mendadak-naik-lagi-usai-serangan-delta-sempt-mereda>

teks case folding (uppercase):
CORONA DI AS MENDADAK NAIK LAGI USAI SERANGAN DELTA SEMPAT MEREDA

JAKARTA - KASUS COVID-19 KEMBALI NAIK DI SEJUMLAH WILAYAH AMERIKA SERIKAT (AS). PADAHAL DIKETAHUI, COVID-19 SEMPAT TERCATAT STABIL PADA SERANGAN VARIAN DELTA MUSIM PANAS INI. ADA APA? HAL TERSEBUT DISAMPAIKAN OLEH KEPALA PENASIHAT MEDIS GEDUNG PUTIH DR. ANTHONY FAUCI, SENIN (15/11/2021). DIKETAHUI, KASUS NASIONAL TURUN 57 PERSEN MINGGU LALU DARI PUNCAK GELOMBANG VARIAN DELTA PADA MUSIM PANAS. NAMUN JUMLAH PASIEN COVID-19 DI AREA BARAT TENGAH DAN TIMUR LAUT KINI NAIK MENDADAK.

[HTTPS://HEALTH.DETIK.COM/BERITA-DETIKHEALTH/D-5813949/CORONA-DI-AS-MENDADAK-NAIK-LAGI-USAI-SERANGAN-DELTA-SEMPAT-MEREDA](https://health.detik.com/berita-detikhealth/d-5813949/corona-di-as-mendadak-naik-lagi-usai-serangan-delta-sempt-mereda)

=====
Tokenisasi masing-masing isi dari folder berita
=====

Tokens in berital.txt: 105
['Wilayah', 'Kamu', 'Sudah', '', 'Bebas', '', 'COVID', '-', '19', '?', 'Cek', '34', 'Kab', '/', 'Kota', 'Zona', 'Hijau', 'Terbaru', '\n\n', 'Jakarta', '-', 'Pemerintah', 'rencananya', 'bakal', 'menerapkan', 'Pemberlakuan', 'Pembatasan', 'Kegiatan', 'Masyarakat', '(', 'PPKM', ')', 'level', '3', 'terhitung', '24', 'Desember', '2021', 'hingga', '2', 'Januari', '2021', '.', 'Namun', ' ', 'Kementeri an', 'Kesehatan', 'RI', 'memastikan', 'kebijakan', 'PPKM', 'level', '3', 'ini', 'masih', 'dalam', 'tahap', 'kajian', ' ', 'Menu rut', 'Direktur', 'Pencegahan', 'dan', 'Pengendalian', 'Penyakit', 'Menular', 'Langsung', '(', 'P2PML', ')', 'Kemenkes', 'RI', 'dr', 'Siti', 'Nadia', 'Tarmizi', ' ', 'PPKM', 'level', '3', 'bakal', 'diterapkan', 'jika', 'kasus', 'COVID', '-', '19', 'naik', 'signifika']

```

[',', 'Hal', 'ini', 'umumnya', 'dipicu', 'dengan', 'peningkatan', 'mobilitas', 'dan', 'pelonggaran', 'protokol', 'kesehatan', '']
['\n\n', 'https://health.detik.com/berita-detikhealth/d-5816690/wilayah-kamu-sudah-bebas-covid-19-cek-34-kabkota-zona-hijau-terbaru']
Tokens in berita2.txt: 105
['Vaksin', 'COVID', '-', '19', 'Bakal', 'Rutin', 'Setiap', 'Tahun', '?', 'Tergantung', 'Ini', 'Penjelasannya', '\n\n', 'Jakarta', '-', 'Pemberian', 'booster', 'atau', 'dosis', 'ketiga', 'vaksin', 'COVID', '-', '19', 'di', 'Indonesia', 'direncanakan', 'bakal', 'berlangsung', 'Januari', '2022', 'Lantas', 'adakah', 'kemungkinan', 'vaksinasi', 'COVID', '-', '19', 'bakal', 'harus', 'dilakuka', 'n', 'setiap', 'tahun', 'seperti', 'vaksinasi', 'influenza', '?', '\n', 'Ketua', 'Satgas', 'COVID', '-', '19', 'Ikatan', 'Dokter', 'Indonesia', 'C', 'IDI', 'Prof', 'Zubairi', 'Djoerban', 'menjelaskan', 'hingga', 'kini', 'belum', 'ada', 'kepastian', 'terkait', 'hal', 'tersebut', 'Menurutnya', 'masih', 'ada', 'kemungkinan', 'vaksin', 'COVID', '-', '19', 'harus', 'diberikan', 'setiap', 'tahun', 'ada', 'juga', 'kemungkinan', 'cukup', 'booster', 'diberikan', 'sekali', 'kemudian', 'vaksinasi', 'COVID', '-', '19', 'tidak', 'diperlukan', 'lagi', '']
['\n\n', 'https://health.detik.com/berita-detikhealth/d-5816582/vaksin-covid-19-bakal-rutin-setiap-tahun-tergantung-ini-penjelasannya']
Tokens in berita3.txt: 130
['RI', 'Mulai', 'Suntikkan', 'Booster', 'di', '2022', 'Masihkah', 'Ampuh', 'Lawan', 'Varian', 'Delta', 'Cs', '?', '\n\n', 'Jakarta', '-', 'Pakar', 'mengakui', 'vaksin', '-', 'vaksin', 'yang', 'sudah', 'digunakan', 'untuk', 'dosis', '1', '-', '2', 'memang', 'mengalami', 'penurunan', 'efektivitas', 'terhadap', 'varian', 'baru', 'Corona', 'seperti', 'varian', 'Delta', 'Mengingat', 'booster', 'atau', 'dosis', 'ketiga', 'vaksin', 'COVID', '-', '19', 'di', 'Indonesia', 'disebut', 'bakal', 'dimulai', 'awal', '2022', 'pakah', 'jenis', 'vaksin', 'yang', 'digunakan', 'bakal', 'mengikuti', 'strain', 'virus', 'terbaru', '?', '\n', 'Menjawab', 'pertanyaan', 'tersebut', 'Ketua', 'Satgas', 'COVID', '-', '19', 'Ikatan', 'Dokter', 'Indonesia', 'C', 'IDI', 'Prof', 'Zubairi', 'Djoerban', 'kembali', 'menyinggung', 'riset', 'yang', 'sudah', 'berlangsung', 'terkait', 'efektivitas', 'vaksin', 'COVID', '-', '19', 'dosis', '1', 'dan', '2', 'Ia', 'menyebut', 'berdasarkan', 'riset', 'sejauh', 'ini', 'efektivitas', 'vaksin', 'COVID', '-', '19', 'Pfizer', 'dan', 'Moderna', 'terbukti', 'menurun', 'dalam', 'melawan', 'varian', 'Delta', '']
['\n\n', 'https://health.detik.com/berita-detikhealth/d-5816534/ri-mulai-suntikkan-booster-di-2022-masihkah-ampuh-lawan-varian-delta-cs']
Tokens in berita4.txt: 98
['Alert', 'Kasus', 'Varian', 'Delta', 'COVID', '-', '19', 'di', 'DKI', 'Meningkat', '\n\n', 'Jakarta', '-', 'Data', 'terbaru', 'dari', 'Balitbangkes', 'Kemenkes', 'RI', 'per', '13', 'November', 'menunjukkan', 'adanya', 'penambahan', 'varian', 'Delta', 'Penambahan', 'tersebut', 'terjadi', 'di', 'Jawa', 'Barat', 'ada', '165', 'kasus', 'DKI', 'Jakarta', '90', 'kasus', 'dan', 'Sulawesi', 'Utara', '86', 'kasus', '\n', 'Dalam', 'satu', 'bulan', 'terakhir', 'Balitbangkes', 'menyebutkan', 'DKI', 'Jakarta', 'mengalami', 'peningkatan', 'kasus', 'varian', 'Delta', 'yang', 'signifikan', 'Sementara', 'pada', 'varian', 'baru', 'seperti', 'varian', 'Alpha', 'varian', 'Delta', 'hingga', 'Beta', 'terbanyak', 'di', 'Indonesia', 'berasal', 'dari', 'DKI', 'Jakarta', 'dengan', 'total', '1.327', 'kasus', '\n\n', 'https://health.detik.com/berita-detikhealth/d-5812940/alert-kasus-varian-delta-covid-19-di-dki-meningkat']
Tokens in berita5.txt: 109
['Corona', 'di', 'AS', 'Mendadak', 'Naik', 'Lagi', 'Usai', 'Serangan', 'Delta', 'Sempat', 'Mereda', '\n\n', 'Jakarta', '-', 'Kasus', 'COVID', '-', '19', 'kembali', 'naik', 'di', 'sejumlah', 'wilayah', 'Amerika', 'Serikat', 'C', 'AS', 'Padahal', 'diketahui', 'COVID', '-', '19', 'sempat', 'tercatat', 'stabil', 'pasca', 'serangan', 'varian', 'Delta', 'musim', 'panas', 'ini', 'Ada', 'apa', '?', '\n', 'Hal', 'tersebut', 'disampaikan', 'oleh', 'kepala', 'penasihat', 'medis', 'Gedung', 'Putih', 'Dr', 'Anthony', 'Fauci', 'Senin', 'C', '15', '11', '2021', 'Diketahui', 'kasus', 'nasional', 'turun', '57', 'persen', 'minggu', 'lalu', 'dari', 'puncak', 'gelombang', 'varian', 'Delta', 'pada', 'musim', 'panas', 'Namun', 'jumlah', 'pasien', 'COVID', '-', '19', 'di', 'area', 'Barat', 'Tengah', 'dan', 'Timur', 'laut', 'kini', 'naik', 'mendadak', '']
['\n\n', 'https://health.detik.com/berita-detikhealth/d-5813949/corona-di-as-mendadak-naik-lagi-usai-serangan-delta-sempat-mereda']
=====
Stopword elimination masing-masing isi dari folder berita
=====
Tokens in berital.txt after stopwords elimination: 85
['Wilayah', 'Bebas', 'COVID', '-', '19', 'Cek', '34', 'Kab', '/', 'Kota', 'Zona', 'Hijau', 'Terbaru', '\n\n', 'Jakarta', '-', 'Pemerintah', 'rencananya', 'menerapkan', 'Pemberlakuan', 'Pembatasan', 'Kegiatan', 'Masyarakat', 'C', 'PPKM', 'level', '3', 'terhitung', '24', 'Desember', '2021', '2', 'Januari', '2021', 'Kementerian', 'RI', 'kebijakan', 'PPKM', 'level', '3', 'tahap', 'kajian', '\n', 'Direktur', 'Pencegahan', 'Pengendalian', 'Penyakit', 'Menular', 'Langsung', 'C', 'P2PM', 'Kemenkes', 'RI', 'dr', 'Siti', 'Nadia', 'Tarmizi', 'PPKM', 'level', '3', 'diterapkan', 'COVID', '-', '19', 'signifikan', 'dipicu', 'peningkatan', 'mobilitas', 'pelonggaran', 'protokol', 'kesehatan', '\n\n', 'https://health.detik.com/berita-detikhealth/d-5816690/wilayah-kamu-sudah-bebas-covid-19-cek-34-kabkota-zona-hijau-terbaru']
Tokens in berita2.txt after stopwords elimination: 66
['Vaksin', 'COVID', '-', '19', 'Rutin', '?', 'Tergantung', 'Penjelasannya', '\n\n', 'Jakarta', '-', 'Pemberian', 'booster', 'dosis', 'ketiga', 'vaksin', 'COVID', '-', '19', 'Indonesia', 'direncanakan', 'Januari', '2022', 'Lantas', 'adakah', 'vaksinasi', 'COVID', '-', '19', 'vaksinasi', 'influenza', '?', '\n', 'Ketua', 'Satgas', 'COVID', '-', '19', 'Ikatan', 'Dokter', 'Indonesia', 'C', 'IDI', 'Prof', 'Zubairi', 'Djoerban', 'kepastian', 'terkait', 'Menurutnya', 'vaksin', 'COVID', '-', '19', 'booster', 'vaksinasi', 'COVID', '-', '19', '']
['\n\n', 'https://health.detik.com/berita-detikhealth/d-5816582/vaksin-covid-19-bakal-rutin-setiap-tahun-tergantung-ini-penjelasannya']
Tokens in berita3.txt after stopwords elimination: 95
['RI', 'Suntikkan', 'Booster', '2022', 'Ampuh', 'Lawan', 'Varian', 'Delta', 'Cs', '?', '\n\n', 'Jakarta', '-', 'Pakar', 'mengakui', 'vaksin', '-', 'vaksin', 'dosis', '1', '-', '2', 'mengalami', 'penurunan', 'efektivitas', 'varian', 'Corona', 'varian', 'Delta', 'booster', 'dosis', 'ketiga', 'vaksin', 'COVID', '-', '19', 'Indonesia', 'jenis', 'vaksin', 'mengikuti', 'strain', 'virus', 'terbaru', '?', '\n', 'Ketua', 'Satgas', 'COVID', '-', '19', 'Ikatan', 'Dokter', 'Indonesia', 'C', 'IDI', 'Prof', 'Zubairi', 'Djoerban', 'menyinggung', 'riset', 'terkait', 'efektivitas', 'vaksin', 'COVID', '-', '19', 'dosis', '1', '2', 'menyebut', 'berdasarkan', 'riset', 'efektivitas', 'vaksin', 'COVID', '-', '19', 'Pfizer', 'Moderna', 'terbukti', 'menurun', 'melawan', 'varian', 'Delta', '\n\n', 'https://health.detik.com/berita-detikhealth/d-5816534/ri-mulai-suntikkan-booster-di-2022-masihkah-ampuh-lawan-varian-delta-cs']
Tokens in berita4.txt after stopwords elimination: 67
['Alert', 'Varian', 'Delta', 'COVID', '-', '19', 'DKI', 'Meningkat', '\n\n', 'Jakarta', '-', 'Data', 'terbaru', 'Balitbangkes', 'Kemenkes', 'RI', '13', 'November', 'penambahan', 'varian', 'Delta', 'Penambahan', 'Jawa', 'Barat', '165', 'kasus', 'DKI', 'Jakarta', '90', 'Sulawesi', 'Utara', '86', '\n', 'Balitbangkes', 'DKI', 'Jakarta', 'mengalami', 'peningkatan', 'varian', 'Delta', 'signifikan', 'varian', 'Alpha', 'varian', 'Delta', 'Beta', 'Indonesia', 'berasal', 'DKI', 'Jakarta', 'total', '1.327', '\n\n', 'https://health.detik.com/berita-detikhealth/d-5812940/alert-kasus-varian-delta-covid-19-di-dki-meningkat']
Tokens in berita5.txt after stopwords elimination: 77
['Corona', 'AS', 'Mendadak', 'Serangan', 'Delta', 'Mereda', '\n\n', 'Jakarta', '-', 'COVID', '-', '19', 'wilayah', 'Amerika', 'Serikat', 'C', 'AS', 'COVID', '-', '19', 'tercatat', 'stabil', 'pasca', 'serangan', 'varian', 'Delta', 'musim', 'panas', '?', '\n', 'kepala', 'penasihat', 'medis', 'Gedung', 'Putih', 'Dr', 'Anthony', 'Fauci', 'Senin', 'C', '15', '11', '2021', 'nasional', 'turun', '57', 'persen', 'minggu', 'puncak', 'gelombang', 'varian', 'Delta', 'musim', 'panas', 'pasien', 'COVID', '-', '19', 'area', 'Barat', 'Timur', 'laut', 'mendadak', '\n\n', 'https://health.detik.com/berita-detikhealth/d-5813949/corona-di-as-mendadak-naik-lagi-usai-serangan-delta-sempat-mereda']

```

=====

isi dari berital.txt setelah stemming:

wilayah kamu sudah bebas covid-19 cek 34 kab kota zona hijau baru jakarta - perintah rencana bakal terap laku batas giat masyarakat p
pkm level 3 hitung 24 desember 2021 hingga 2 januari 2021 namun menteri sehat ri pasti bijak ppkm level 3 ini masih dalam tahap kaji
turut direktur cegah dan kendali sakit tular langsung p2pml kemenkes ri dr siti nadia tarmizi ppkm level 3 bakal terap jika kasus cov
id-19 naik signifikan hal ini umum picu dengan tingkat mobilitas dan longgar protokol sehat https health detik com berita-detikhealth
d-5816690 wilayah-kamu-sudah-bebas-covid-19-cek-34-kabkota-zona-hijau-terbaru

isi dari berita2.txt setelah stemming:

vaksin covid-19 bakal rutin tiap tahun gantung ini jelas jakarta - beri booster atau dosis tiga vaksin covid-19 di indonesia rencana
bakal langsung januari 2022 lantas ada mungkin vaksinasi covid-19 bakal harus laku tiap tahun seperti vaksinasi influenza ketua satga
s covid-19 ikat dokter indonesia idi prof zubairi djoerban jelas hingga kini belum ada pasti kait hal sebut turut masih ada mungkin v
aksin covid-19 harus beri tiap tahun ada juga mungkin cukup booster beri sekali kemudian vaksinasi covid-19 tidak perlu lagi https he
alth detik com berita-detikhealth d-5816582 vaksin-covid-19-bakal-rutin-setiap-tahun-tergantung-ini-penjasannya

isi dari berita3.txt setelah stemming:

ri mulai suntik booster di 2022 masih ampuh lawan varian delta cs jakarta - pakar aku vaksin yang sudah guna untuk dosis 1-2 memang a
lami turun efektivitas hadap varian baru corona seperti varian delta ingat booster atau dosis tiga vaksin covid-19 di indonesia sebut
bakal mulai awal 2022 apakah jenis vaksin yang guna bakal ikut strain virus baru jawab tanya sebut ketua satgas covid-19 ikat dokter
indonesia idi prof zubairi djoerban kembali singgung riset yang sudah langsung kait efektivitas vaksin covid-19 dosis 1 dan 2 ia seb
ut dasar riset jauh ini efektivitas vaksin covid-19 pfizer dan moderna bukti turun dalam lawan varian delta https health detik com be
rita-detikhealth d-5816534 ri-mulai-suntikkan-booster-di-2022-masihkah-ampuh-lawan-varian-delta-cs

isi dari berita4.txt setelah stemming:

alert kasus varian delta covid-19 di dki tingkat jakarta - data baru dari balitbangkes kemenkes ri per 13 november tunjuk ada tambah
varian delta tambah sebut jadi di jawa barat ada 165 kasus dki jakarta 90 kasus dan sulawesi utara 86 kasus dalam satu bulan akhir ba
litbangkes sebut dki jakarta alami tingkat kasus varian delta yang signifikan sementara pada varian baru seperti varian alpha varian
delta hingga beta banyak di indonesia asal dari dki jakarta dengan total 1 327 kasus https health detik com berita-detikhealth d-5812
940 alert-kasus-varian-delta-covid-19-di-dki-meningkat

isi dari berita5.txt setelah stemming:

corona di as dadak naik lagi usai serang delta sempat reda jakarta - kasus covid-19 kembali naik di jumlah wilayah amerika serikat as
padahal tahu covid-19 sempat catat stabil pasca serang varian delta musim panas ini ada apa hal sebut sampai oleh kepala nasihat med
is gedung putih dr anthony fauci senin 15 11 2021 tahu kasus nasional turun 57 persen minggu lalu dari puncak gelombang varian delta
pada musim panas namun jumlah pasien covid-19 di area barat tengah dan timur laut kini naik dadak https health detik com berita-detik
health d-5813949 corona-di-as-mendadak-naik-lagi-usai-serangan-delta-sempat-mereda