KEMENTERIAN RISET TEKNOLOGI DAN PENDIDIKAN TINGGI PROGRAM STUDI ILMU KOMPUTER DEPARTEMEN ILMU KOMPUTER DAN ELEKTRONIKA FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS GADJAH MADA

TEMU KEMBALI INFORMASI

Tugas 6 Metode dan Hasil Kompresi Indeks

Sistem Rekomendasi Obat Berdasarkan Gejala



DISUSUN OLEH:

ADAM YOGISYAH PUTRA 20/455439/PA/19654

MUHAMMAD ARSYA PUTRA 20/462186/PA/20158

HIZKYA FIRSTADIPA HARTOKO 20/455447/PA/19662

DOSEN:

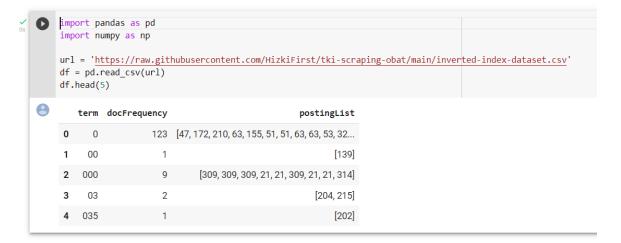
Dr. Lukman Heryawan, S.T., M.T.

I. Metode

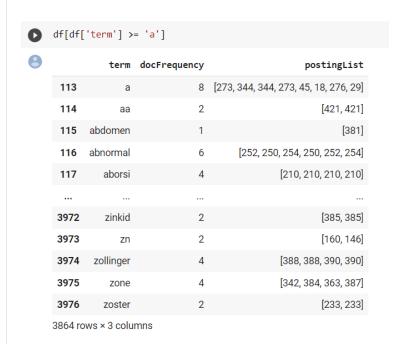
Metode yang digunakan untuk kompresi indeks sistem rekomendasi obat berdasarkan gejala adalah dengan metode Dictionary as a String, yaitu mengompres term list dengan menggabungkan semua term menjadi satu string. Dengan tidak menyimpan term pada setiap kolom di term, kita dapat menghemat file dari dictionary. Metode Dictionary as a String ini termasuk dalam metode Lossless compression karena tidak ada informasi yang terbuang. Selain itu, metode Lossy compression juga digunakan dengan cara membuang semua term angka karena term tersebut tidak digunakan pada sistem rekomendasi obat.

II. Proses Kompresi Indeks

- 1. Menghilangkan data yang tidak perlu
 - ▼ Import Library dan Dataset



- ▼ Penghapusan term berupa bilangan
- ▼ Mencari index term huruf pertama



▼ Mengambil data dengan term huruf saja

```
[ ] df_clean = df.iloc[113:]
[ ] df_clean = df_clean.reset_index(drop=True)
[ ] df_clean
```

	term	docFrequency	postingList
0	а	8	[273, 344, 344, 273, 45, 18, 276, 29]
1	aa	2	[421, 421]
2	abdomen	1	[381]
3	abnormal	6	[252, 250, 254, 250, 252, 254]
4	aborsi	4	[210, 210, 210, 210]
3859	zinkid	2	[385, 385]
3860	zn	2	[160, 146]
3861	zollinger	4	[388, 388, 390, 390]
3862	zone	4	[342, 384, 363, 387]
3863	zoster	2	[233, 233]

3864 rows × 3 columns

2. Pembuatan dictionary

▼ Pembuatan Dictionary

```
[ ] df_dict = pd.DataFrame(columns=['Freq','Posting_ptr','Term_ptr'])
[ ] df_dict
        Freq Posting_ptr Term_ptr
[ ] df_dict['Freq'] = df_clean['docFrequency']
[ ] df_dict['Posting_ptr'] = df_clean['postingList']
[ ] df_dict
            Freq
                                     Posting_ptr Term_ptr
               8 [273, 344, 344, 273, 45, 18, 276, 29]
                                                       NaN
                                        [421, 421]
                                                       NaN
       2
                                            [381]
                                                       NaN
       3
                       [252, 250, 254, 250, 252, 254]
               6
                                                       NaN
               4
                               [210, 210, 210, 210]
                                                       NaN
               2
                                        [385, 385]
      3859
                                                       NaN
      3860
               2
                                        [160, 146]
                                                       NaN
```

- 3. Penggabungan semua term menjadi 1 term string dan menentukan term pointer
 - ▼ Menentukan Term Pointer dan Membuat Term String

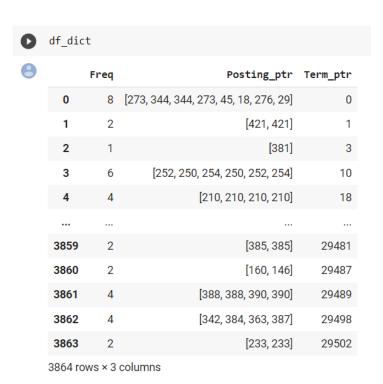
```
[ ] term_str = ""
  term_counter = 0
  for i in range(3864):
    term = df_clean['term'].iloc[i]
    term_str += term
    df_dict('Term_ptr'].iloc[i] = term_counter
    term_counter += len(term)
```

/usr/local/lib/python3.7/dist-packages/pandas/core/indexing.py:1732: SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy-self._setitem_single_block(indexer, value, name)

[] term_str

'aaaabdomenabnormalaborsiabsesabsolutabsopsiabsorpsinyaaceacetateacetonideacetylcysteineacetylsalicylicacidacnacareacneactacquiredacsbpactifedactingactionacycloviradadaadalahadany aadekuatadenosinadenosinaadenosineadhesiadhesiveadidasadipositadpadrenergicadrenergikadrenoceptoradrenokortikaladrenoreceptorsadrenoreseptoradukadvanceaereusaerobaerogenesaeruginosaaf ricanumaftosaagakagalactiaeagaragenagentagonisagonistagregasiahliainsairakanakarakhirakhirnyaakibatakilenakneaksiaktifaktifitasaktivatoraktivitaasaktivitasaktivitasaktaivitasaktuoitasaktuoitasaktivatoraktivitasaktiv



4. Export Dictionary ke CSV dan Term

▼ Export Hasil Dataset dan Term String

```
[ ] df_dict.to_csv('compressed-dictionary-dataset.csv',index=False)

text_file = open("term.txt", "w")
text_file.write(term_str)
text_file.close()
```

Link Dataset Hasil Kompresi Indeks:

 $\underline{https://github.com/HizkiFirst/tki-scraping-obat/blob/kumpul/tugas-6_7/compressed-dictionum-dataset.csv}$

Link Term String:

https://github.com/HizkiFirst/tki-scraping-obat/blob/kumpul/tugas-6 7/term.txt

IV. Lampiran

Link Repository: https://github.com/HizkiFirst/tki-scraping-obat

Link Google Collab Source Code:

https://colab.research.google.com/drive/1LLpyk1KC3retW3YzhQMabFoTOVaYRGPG?usp=sharing