

Nama : Usama
NPM : 1606862766

1. Mula-mula stem ***korpus-tugas2.txt*** menggunakan ***stemmer.p***, saya melakukan stem menggunakan library sastrawi kemudian menyimpan hasil stem di ***korpus-tugas2-stem.txt***. Kemudian indeks dokumen berdasar <https://metacpan.org/pod/Plucene>. Untuk melakukan langkah-langkah :
 - a. Membuat dokumen.
 - b. Memilih *Analyser* dan menambah dokumen ke *Index Writer*.

Asumsi pengindeksan yang saya lakukan adalah menyamakan tag **<TITLE>** dan tag **<TEXT>** karena hasil yang didapatkan lebih baik daripada membedakan tag **<TITLE>** dengan tag **<TEXT>**.

Kemudian melakukan query

- a. Membuat parser dan query yang akan digunakan.
- b. Gunakan query di *IndexSearcher* dan *collect hits*.

Dokumen yang diterima tidak terurut *score*-nya, oleh karena itu sebelum memberikan dokumen yang diinginkan harus diurutkan berdasarkan dokumen yang memiliki *score* terbesar.

Berikut ini merupakan dokumen yang dianggap relevan untuk setiap query yang diberikan.

Query : pemilihan umum

SCORE	DOCUMENT ID
0.461140432608570	KPS-30062019-046
0.439018728652758	KPS-30062019-029
0.435672402470124	KPS-30062019-015
0.432341952488017	KPS-30082019-733
0.419179034027944	KPS-30062019-049
0.416719764645259	KPS-30062019-027
0.411881558478908	KPS-30062019-047
0.391379286587378	KPS-30062019-035
...	

Query : pemilihan umum presiden

SCORE	DOCUMENT ID
0.696140373309915	KPS-30062019-015
0.693890922386709	KPS-30062019-046
0.679237457880773	KPS-30062019-027
0.662597424947924	KPS-30062019-047
0.646294229556225	KPS-30062019-029
0.641407754379416	KPS-30062019-049
0.586611844202620	KPS-30062019-035
0.586223643793674	KPS-30062019-017
0.586223643793674	KPS-30062019-034
...	

Query : penetapan pemilihan umum presiden

SCORE	DOCUMENT ID
0.823249523626577	KPS-30062019-015
0.811618742851642	KPS-30062019-046
0.790884871074414	KPS-30062019-047
0.772269903743001	KPS-30062019-027
0.762512548646126	KPS-30062019-049
0.742298326305320	KPS-30062019-029
0.693787043655143	KPS-30062019-017
0.678449073488103	KPS-30062019-034
0.670458247704378	KPS-30062019-038
...	

Untuk hasil lebih lengkapnya dapat dilihat di “***pemilihan umum_documents.txt***”, “***pemilihan umum presiden_documents.txt***”, dan “***penetapan pemilihan umum presiden_documents.txt***” setelah menjalankan program **T3_1606862766_Usama.p**

2. Untuk mengerjakan soal nomor 2 ini, saya membuat program python **nomor2.py**, penjelasan di bawah mengasumsikan telah menjalankan program **nomor2.py**.

a. Hitunglah nilai *Mean Average Precision*

AP("pemilihan umum") = 0.187

AP("pemilihan umum presiden") = 0.268

AP("penetapan pemilihan umum presiden") = 0.281

MAP = $(0.187 + 0.268 + 0.281) / 3 = 0.245$

lebih jelasnya dapat dilihat di **MAP.txt**

b. Hitung nilai *Average Precision* pada *top-n* dokumen @10, @20, dan @100

top_n_average_precision.txt
<pre>pemilihan umum_documents.txt top 10 AP = 0.125 pemilihan umum_documents.txt top 20 AP = 0.201 pemilihan umum_documents.txt top 100 AP = 0.187 pemilihan umum presiden_documents.txt top 10 AP = 0.183 pemilihan umum presiden_documents.txt top 20 AP = 0.26 pemilihan umum presiden_documents.txt top 100 AP = 0.268 penetapan pemilihan umum presiden_documents.txt top 10 AP = 0.163 penetapan pemilihan umum presiden_documents.txt top 20 AP = 0.269 penetapan pemilihan umum presiden_documents.txt top 100 AP = 0.281</pre>

C.

