

Nama : Muhamad Achir Suci Ramadhan

NPM : 1706979354

Laporan Tugas 3 Perolehan Informasi

1. Program dapat dilihat pada file T3_1706979354_MuhamadAchirSuciRamadhan.pl. Stemming dibuat ke dalam bahasa Perl menggunakan referensi <https://github.com/apache/lucene-solr/blob/master/lucene/analysis/common/src/java/org/apache/lucene/analysis/id/IndonesianStemmer.java> . Terdapat berbagai keanehan pada laptop saya saat saya ingin menggunakan langsung *library* lucene-solr sehingga saya memutuskan untuk menulis ulang kodingan Java tersebut ke dalam bahasa Perl. *Query* diproses terhadap isi teks dan judul setiap dokumen. *Search engine* yang digunakan berasal dari Library lucene. Dokumen relevan hasil query menggunakan program yang dibuat, diurutkan dari yang paling relevan, sebagai berikut:

a. pemilihan umum

Query setelah tokenisasi dan pengolahan sebelum masuk Plucene:

(text:ilih AND text:umum) OR (title:ilih AND title:umum)

53 Dokumen

KPS-31072019-356 (urutan ke-1)

KPS-30062019-046

KPS-30062019-029

KPS-30062019-049

KPS-30062019-017

KPS-30062019-027 (6)

KPS-30062019-035

KPS-31072019-379

KPS-30062019-056

KPS-31072019-177

KPS-30062019-015 (11)

KPS-30062019-034

KPS-31072019-210

KPS-30062019-047

KPS-30062019-050

KPS-30082019-733 (16)

KPS-30062019-038

KPS-31072019-241

KPS-30082019-556

KPS-30062019-060

KPS-30062019-040 (21)

KPS-30062019-043

KPS-31072019-366

KPS-30082019-706

KPS-30062019-061

KPS-30082019-684 (26)

KPS-31072019-408

KPS-30062019-058

KPS-31072019-329

KPS-31072019-378

KPS-30062019-052 (31)

KPS-30062019-095

KPS-31072019-441

KPS-31072019-289

KPS-30082019-541

KPS-30082019-577 (36)

KPS-30062019-030

KPS-30062019-054

KPS-30062019-092

KPS-30062019-114

KPS-30062019-116 (41)

KPS-30062019-126

KPS-31072019-257

KPS-31072019-369

KPS-30082019-557

KPS-30082019-681 (46)

KPS-01092019-780

KPS-30062019-096

KPS-31072019-283

KPS-31072019-327

KPS-31072019-336 (51)

KPS-30082019-478

KPS-30062019-128

b. pemilihan umum presiden

Query setelah tokenisasi dan pengolahan sebelum masuk Plucene:

(text:ilih AND text:umum AND text:presiden) OR (title:ilih AND title:umum AND title:presiden)

30 Dokumen

KPS-30062019-046 (urutan ke-1)

KPS-30062019-027

KPS-30062019-029

KPS-30062019-049

KPS-30062019-017

KPS-30062019-015 (6)

KPS-30062019-035

KPS-30062019-047

KPS-30062019-034

KPS-30062019-038

KPS-30082019-556 (11)

KPS-30062019-043

KPS-30062019-050

KPS-30062019-056

KPS-30062019-061

KPS-30062019-058 (16)

KPS-30062019-040

KPS-30082019-706

KPS-30062019-054

KPS-30062019-095

KPS-30062019-096 (21)

KPS-30062019-030

KPS-30062019-092

KPS-30062019-114

KPS-30062019-116

KPS-30062019-060 (26)

KPS-30062019-052

KPS-31072019-366

KPS-30062019-126

KPS-30082019-681

c. penetapan pemilihan umum presiden

Query setelah tokenisasi dan pengolahan sebelum masuk Plucene:

(text:tetap AND text:ilih AND text:umum AND text:presiden) OR (title:tetap AND title:ilih AND title:umum AND title:presiden)

27 Dokumen

KPS-30062019-046 (urutan ke-1)

KPS-30062019-029

KPS-30062019-027

KPS-30062019-017

KPS-30062019-015

KPS-30062019-049 (6)

KPS-30062019-047

KPS-30062019-035

KPS-30062019-034

KPS-30062019-038

KPS-30062019-043 (11)

KPS-30062019-058

KPS-30062019-050

KPS-30062019-040

KPS-30062019-061

KPS-30062019-056 (16)

KPS-30062019-052

KPS-30062019-096

KPS-30062019-054

KPS-30062019-095

KPS-30062019-116 (21)

KPS-30082019-706

KPS-30062019-030

KPS-30062019-092

KPS-30062019-126

KPS-30062019-114 (26)

KPS-31072019-366

2. Analisis

a. Mean Average Precision

i. Average Precision Query 1

Dokumen relevan muncul pada urutan ke- 7, 12, 15, 22, 25, 37, 41, 48.
Berarti, recall naik hanya pada urutan-urutan tersebut. Precision pada

urutan-urutan tersebut adalah: 1/7, 2/12, 3/15, 4/22, 5/25, 6/37, 7/41, 8/48.

$$AP = (1/7 + 2/12 + 3/15 + 4/22 + 5/25 + 6/37 + 7/41 + 8/48) / 8 \\ = 0.1738628$$

ii. Average Precision Query 2

Dokumen relevan muncul pada urutan ke- 7, 9, 12, 13, 15, 21, 22, 25. Precision pada urutan-urutan tersebut adalah: 1/7, 2/9, 3/12, 4/13, 5/15, 6/21, 7/22, 8/25.

$$AP = (1/7 + 2/9 + 3/12 + 4/13 + 5/15 + 6/21 + 7/22 + 8/25) / 8 \\ = 0.2725001$$

iii. Average Precision Query 3

Dokumen relevan muncul pada urutan ke- 8, 9, 11, 13, 15, 18, 21, 23. Precision pada urutan-urutan tersebut adalah: 1/8, 2/9, 3/11, 4/13, 5/15, 6/18, 7/21, 8/23.

$$AP = (1/8 + 2/9 + 3/11 + 4/13 + 5/15 + 6/18 + 7/21 + 8/23) / 8 \\ = 0.2844335$$

$$MAP = (0.1738628 + 0.2725001 + 0.2844335) / 3 = 0.2435988$$

b. Average Precision pada

i. top 10

1. pemilihan umum, dokumen relevan muncul di urutan ke-7

$$AP = (1/7)/1 = 0.1428571$$

2. pemilihan umum presiden, dokumen relevan muncul di urutan ke- 7, 9

$$AP = (1/7 + 2/9)/2 = 0.1825397$$

3. penetapan pemilihan umum presiden, dokumen relevan muncul di urutan ke- 8, 9

$$AP = (1/8 + 2/9)/2 = 0.1736111$$

ii. top 20

1. pemilihan umum, dokumen relevan muncul di urutan ke- 7, 12, 15

$$AP = (1/7 + 2/12 + 3/15)/3 = 0.1698413$$

2. pemilihan umum presiden, dokumen relevan muncul di urutan ke- 7, 9, 12, 13, 15

$$AP = (1/7 + 2/9 + 3/12 + 4/13 + 5/15)/5 = 0.2512210$$

3. penetapan pemilihan umum presiden, dokumen relevan muncul di urutan ke- 8, 9, 11, 13, 15, 18

$$AP = (1/8 + 2/9 + 3/11 + 4/13 + 5/15 + 6/18)/6 = 0.2657181$$

iii. top 100

1. pemilihan umum, dokumen relevan muncul di urutan ke- 7, 12, 15, 22, 25, 37, 41, 48

$$AP = (1/7 + 2/12 + 3/15 + 4/22 + 5/25 + 6/37 + 7/41 + 8/48) / 8 \\ = 0.1738628$$

2. pemilihan umum presiden, dokumen relevan muncul di urutan ke- 7, 9, 12, 13, 15, 21, 22, 25

$$AP = (1/7 + 2/9 + 3/12 + 4/13 + 5/15 + 6/21 + 7/22 + 8/25) / 8 \\ = 0.2725001$$

3. penetapan pemilihan umum presiden, dokumen relevan muncul di urutan ke- 8, 9, 11, 13, 15, 18, 21, 23

$$AP = (1/8 + 2/9 + 3/11 + 4/13 + 5/15 + 6/18 + 7/21 + 8/23) / 8 \\ = 0.2844335$$

c. Grafik Precision dan Recall untuk mengevaluasi top 200 dokumen

Tabel agar angka yang ditunjuk oleh plot lebih jelas:

Recall	Query 1	Query 2	Query 3	
0	0.1428571429	0.1428571429	0.125	interpolasi
0.1	0.1428571429	0.1428571429	0.125	interpolasi
0.125	0.1428571429	0.1428571429	0.125	
0.2	0.1666666667	0.2222222222	0.2222222222	interpolasi
0.25	0.1666666667	0.2222222222	0.2222222222	
0.3	0.2	0.25	0.2727272727	interpolasi
0.375	0.2	0.25	0.2727272727	
0.4	0.1818181818	0.3076923077	0.3076923077	interpolasi
0.5	0.1818181818	0.3076923077	0.3076923077	
0.6	0.2	0.3333333333	0.3333333333	interpolasi
0.625	0.2	0.3333333333	0.3333333333	
0.7	0.1621621622	0.2857142857	0.3333333333	interpolasi
0.75	0.1621621622	0.2857142857	0.3333333333	
0.8	0.1707317073	0.3181818182	0.3333333333	interpolasi
0.875	0.1707317073	0.3181818182	0.3333333333	
0.9	0.1666666667	0.32	0.347826087	interpolasi
1	0.1666666667	0.32	0.347826087	

Grafik:

Grafik Recall - Precision

