

Tugas Sesi - 13
Melakukan Analisa Sistem IR

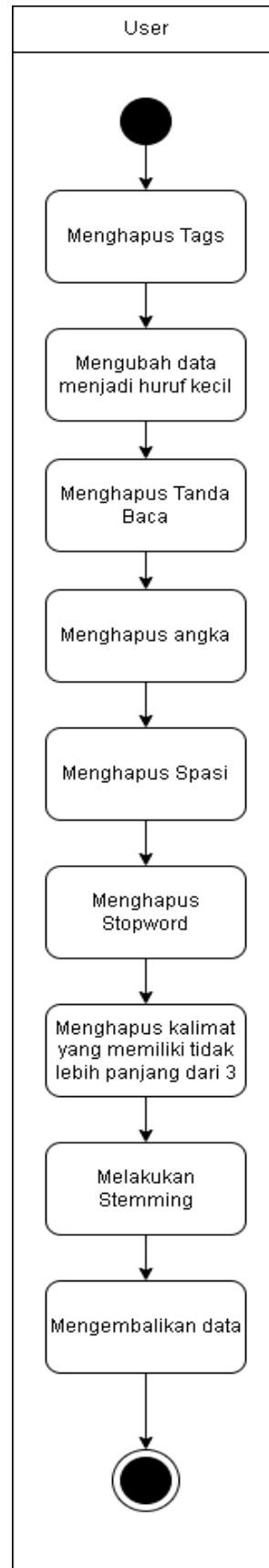
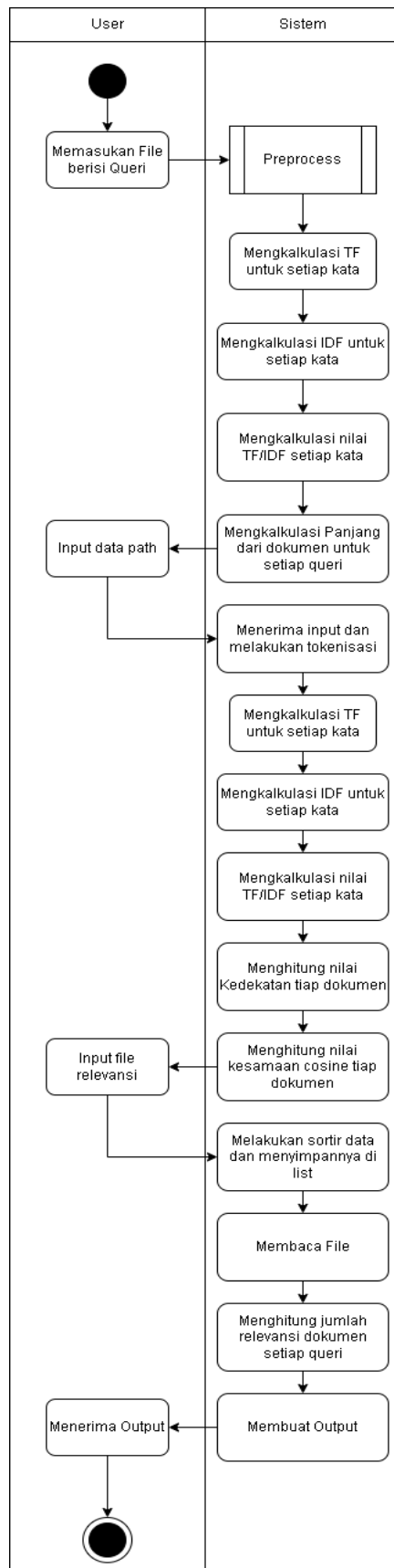
Jonathan Leonardi - 1872018

Ronaldo Tasman - 1872030

Billy Arya - 1872034

Steven Peter - 1872047

1. Berikut merupakan alur kerja sistem yang digambarkan dengan Activity Diagram



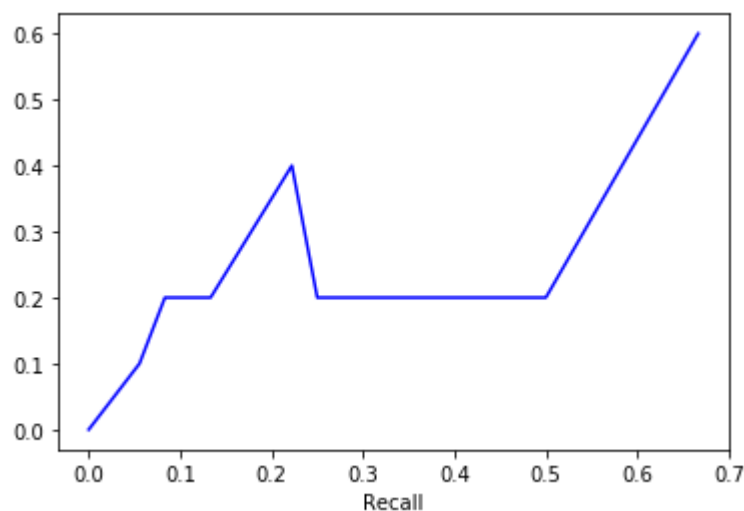
2. Jelaskan struktur data yang digunakan untuk representasi dokumen dan kueri, misalnya: matrix, dictionary atau lainnya:

Struktur data yang digunakan pada program IR ini menggunakan dictionary python yang mempunyai pasangan antara sebuah key dan value, dimana key pada kasus ini adalah sebuah kata yang sudah dipisahkan sesuai dengan frekuensi kata yang muncul pada setiap dokumen dan value nya adalah nilai dari TF/IDF dari kata tersebut.

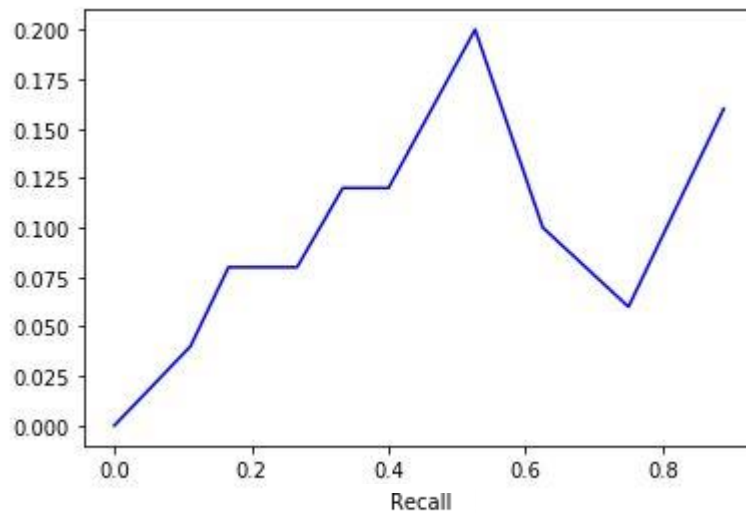
3. Bagaimanakah evaluasi performa temu balik dilakukan? Jelaskan perhitungan precision dan recall yang terjadi

Sebelumnya, Recall sendiri adalah perhitungan dokumen yang relevan yang terdapat didalam sistem dan precision adalah jumlah kelompok dokumen yang relevan dari jumlah keseluruhan dokumen. Kedua nilai ini yang dipakai sebagai patokan efektifitas suatu sistem temu balik.

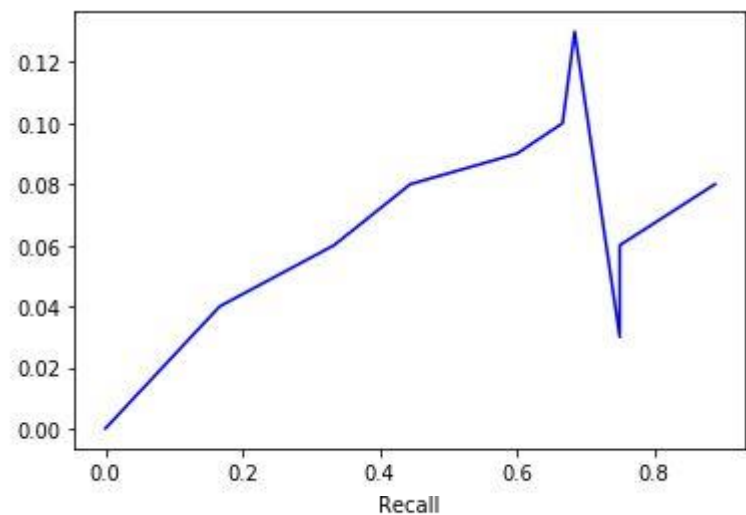
Adapun proses kalkulasi tersebut terjadi pada out.ipnyb dimana pada fungsi output, sistem akan menerima data yang telah tersortir. Kemudian akan dihitung untuk presisi dan recall nya. Dan dapat dilihat pada gambar gambar dibawah ini.



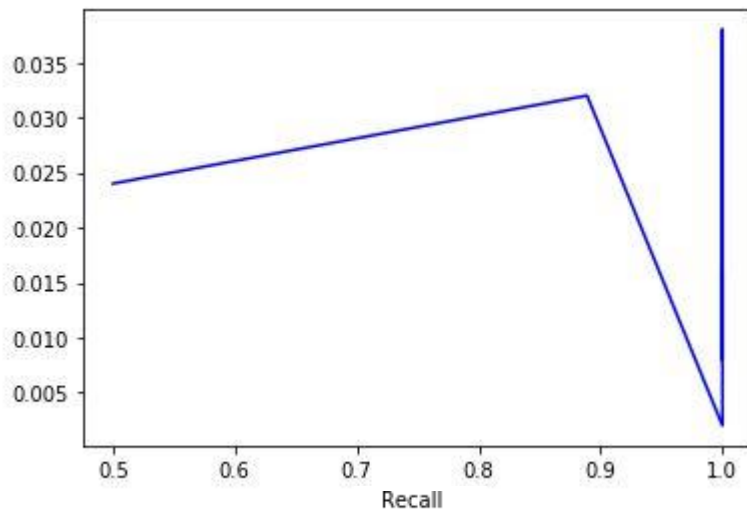
Gambar diatas merupakan top-10 untuk hasil temu balik, Adapun lonjakan yang signifikan membuat performa dari top 10 kurang baik.



Gambar selanjutnya adalah top 50 dimana performa menunjukan grafik yang stabil, oleh karena itu performa dari top 50 dapat dikatakan baik.

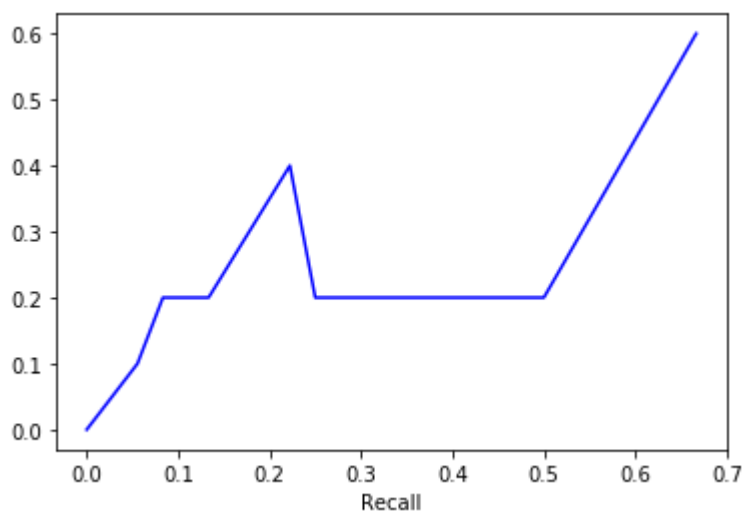


Gambar diatas merupakan hasil performa untuk top 100 dimana terjadi lonjakan yang tinggi sehingga membuat performa untuk top 100 kurang baik.



Gambar terakhir menunjukkan hasil performa untuk Top 500, dimana hasil menunjukkan grafik yang cukup stabil yang dapat dikatakan bahwa top 500 memiliki performa yang baik.

4. Gambarkan grafik P/R hasil interpolasi untuk top-10 hasil temu balik untuk kueri yang dicontohkan.



5. Adapun hasil yang kami dapatkan setelah berhasil mengeksekusi kode program antara lain:

FOR 10 MOST RELEVANT DOCUMENTS

Query 1			
Precision	0.0	Recall	0.0
Query 2			
Precision	0.2	Recall	0.13333333333333333
Query 3			
Precision	0.2	Recall	0.13333333333333333
Query 4			
Precision	0.1	Recall	0.05555555555555555
Query 5			
Precision	0.2	Recall	0.10526315789473684
Query 6			
Precision	0.4	Recall	0.22222222222222222
Query 7			
Precision	0.6	Recall	0.66666666666666666
Query 8			
Precision	0.2	Recall	0.5
Query 9			
Precision	0.2	Recall	0.25
Query 10			
Precision	0.2	Recall	0.08333333333333333
Average precision	0.23000000000000004		
Average recall	0.21497076023391815		

10 Most Relevant Document

FOR 50 MOST RELEVANT DOCUMENTS

Query 1			
Precision	0.0	Recall	0.0
Query 2			
Precision	0.08	Recall	0.26666666666666666
Query 3			
Precision	0.12	Recall	0.4
Query 4			
Precision	0.04	Recall	0.11111111111111111
Query 5			
Precision	0.2	Recall	0.5263157894736842
Query 6			
Precision	0.12	Recall	0.33333333333333333
Query 7			
Precision	0.16	Recall	0.88888888888888888
Query 8			
Precision	0.06	Recall	0.75
Query 9			
Precision	0.1	Recall	0.625
Query 10			
Precision	0.08	Recall	0.16666666666666666
Average precision	0.096		
Average recall	0.4067982456140351		

50 Most Relevant Document

FOR 100 MOST RELEVANT DOCUMENTS

Query 1			
Precision	0.0	Recall	0.0
Query 2			
Precision	0.1	Recall	0.6666666666666666
Query 3			
Precision	0.09	Recall	0.6
Query 4			
Precision	0.06	Recall	0.3333333333333333
Query 5			
Precision	0.13	Recall	0.6842105263157895
Query 6			
Precision	0.08	Recall	0.4444444444444444
Query 7			
Precision	0.08	Recall	0.8888888888888888
Query 8			
Precision	0.03	Recall	0.75
Query 9			
Precision	0.06	Recall	0.75
Query 10			
Precision	0.04	Recall	0.1666666666666666
Average precision	0.06700000000000002		
Average recall	0.528421052631579		

100 Most Relevant Document

FOR 500 MOST RELEVANT DOCUMENTS

Query 1			
Precision	0.002	Recall	1.0
Query 2			
Precision	0.03	Recall	1.0
Query 3			
Precision	0.03	Recall	1.0
Query 4			
Precision	0.032	Recall	0.8888888888888888
Query 5			
Precision	0.038	Recall	1.0
Query 6			
Precision	0.032	Recall	0.8888888888888888
Query 7			
Precision	0.018	Recall	1.0
Query 8			
Precision	0.008	Recall	1.0
Query 9			
Precision	0.016	Recall	1.0
Query 10			
Precision	0.024	Recall	0.5
Average precision	0.023		
Average recall	0.9277777777777778		

500 Most Relevant Document