

INFORMATIKA

DIKLAT
HMIF 2019

**TEMU BALIK
INFORMASI**

AIK21444

MATA KULIAH PILIHAN

DAFTAR ISI

DAFTAR ISI	2
SOAL UJIAN TENGAH SEMESTER 2013/2014	3
SOAL UJIAN AKHIR SEMESTER 2013/2014	4
SOAL UJIAN TENGAH SEMESTER 2014/2015	5
SOAL UJIAN TENGAH SEMESTER 2017/2018	6
SOAL UJIAN AKHIR SEMESTER 2017/2018	7

SOAL UJIAN TENGAH SEMESTER 2013/2014



Ujian Tengah Semester Genap 2013/2014 Jurusan Teknik Informatika FSM UNDIP Semarang

Mata Kuliah	: Sistem Temu Balik Informasi	Dosen	: - Sukmawati Nur Endah, S.Si, M.Kom - Sutikno, M.Cs
Beban	: 3 SKS	Hari/Tgl	: Kamis / 24 April 2014
Sifat	: Closed Book	Waktu	: 90 menit

Catatan :

- Nilai 0 akan diberikan bagi mahasiswa yang berbuat curang dalam bentuk apapun!
- Masing-masing soal mempunyai bobot nilai 20.

Soal:

1. Gambarkan dan jelaskan konsep sebuah IR System!
2. Jelaskan tahapan pengindeksan dalam IR !
3. Berikut adalah indeks posisi dengan format sebagai berikut :
Kata : dokumen1 : posisi1, posisi 2, dst; dokumen2 : posisi1, posisi 2, dst

banjir	: 1 : 12, 45, 51;	2 : 23, 35;	3 : 27, 52, 60;	4 : 4, 21;
hujan	: 1 : 2, 18;	2 : -;	3 : 1, 7;	4 : 16, 19;
Jakarta	: 1 : 9, 20;	2 : 13, 34;	3 : 24;	4 : 11, 23, 32
bencana	: 1 : 4, 7, 44;	2 : 12, 48, 57	3 : 26, 30;	4 : -;
ibukota	: 1 : 7;	2 : 33, 41;	3 : -;	4 : 22;

Dokumen manakah yang memuat frase berikut ini?

- a. Ibukota Jakarta banjir
- b. Bencana banjir

Manakah dokumen yang akan terretrieve dengan query sebagai berikut :

- c. Ibukota (near 4) banjir
- d. Hujan (near 10) Jakarta
- e. Jakarta (with) banjir

4. Diketahui data sebagai berikut :

Kata Dok	X1	X2	X3	X4	X5
D1	1	1	1	0	1
D2	1	0	0	1	1
D3	0	1	0	0	0
D4	1	0	1	1	0
D5	0	1	0	0	1
D6	1	0	1	1	0

- a. Jika diberi query sebagai berikut:
 $(X1 \text{ OR } X5) \text{ AND } (\text{NOT } X3) \text{ OR } ((\text{NOT } X2) \text{ AND } X4)$
dokumen apa yang akan terretrieve?
 - b. Tuliskan sebuah query yang mempunyai hasil dokumen D4 dan D6, buktikan kebenarannya!
5. Gunakan *vector space model* (menggunakan pembobotan tf.idf) untuk meranking dokumen yang terretrieve dengan data berikut ini.

Query : Matakuliah Information Retrieval

Dokumen 1 : Matakuliah di Informatika sangat menarik
Dokumen 2 : Information Retrieval sangat diminati
Dokumen 3 : Retrieval bagian dari matakuliah pilihan

Daftar Stopword :

di
sangat
dari

&%% Selamat Mengerjakan &%%

SOAL UJIAN AKHIR SEMESTER 2013/2014



Ujian Akhir Semester Genap TA 2013/2014
Jurusan Ilmu Komputer / Informatika
Fakultas Sains dan Matematika Universitas Diponegoro Semarang

Mata Kuliah	: Sistem Temu Balik Informasi
Dosen	: Sutikno, S.T., M.Cs./Sukmawati N.E., S.Si., M.Kom
Hari/ Tanggal	: Kamis, 10 Juli 2014
Jam	: 13.00 – 14.30 WIB (90 Menit)
Sifat	: Open Book

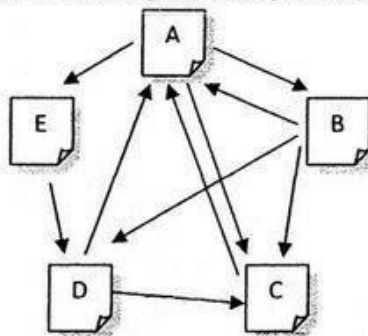
Perhatian: Segala bentuk **Kecurangan** (kerja sama, meminjam/memberi pinjaman catatan dan atau alat tulis dan lain-lain) akan diberikan nilai **0 (nol)**.

1. Jika suatu sistem temu balik informasi diketahui vektor dokumen relevance ($\bar{d}_n(R)$), vektor dokumen tidak relevance ($\bar{d}_n(TR)$), dan vektor query originalnya ($\bar{q}_0(R)$) dituliskan dalam bentuk matrik berikut:

	jalan	kota	indonesia	macet
$\bar{d}_1(R)$	9	5	4	2
$\bar{d}_2(R)$	2	1	4	0
$\bar{d}_3(R)$	3	4	1	0
$\bar{d}_1(TR)$	12	1	0	0
$\bar{d}_2(TR)$	0	1	0	3
\bar{q}_0	1	2	0	3

Berapa Query setelah dimodifikasi (setelah proses *relevance feedback*) dengan menggunakan *Rocchio Algorithm* ? (gunakan $\alpha=1$, $\beta=0,4$ dan $\gamma=0,1$)

2. Jika pada gambar berikut merupakan ilustrasi dari jaringan komputer, simpul mewakili halaman web dan busur mewakili link-link yang menghubungkan antara halaman web yang satu dengan halaman web yang lain. Hitung nilai PageRank setelah iterasi ke-3 dengan menggunakan algoritma PAGERANK? (gunakan nilai PageRank awal pada masing-masing halaman web 0,2)



3. Suatu sistem temu balik informasi dilakukan query sebanyak 6 kali dan setiap query didapatkan jumlah dokumen relevan terambil, jumlah dokumen tidak relevan terambil, dan jumlah dokumen relevan pada koleksi seperti pada tabel berikut:

Query Ke-	Jumlah Dokumen Relevan Terambil	Jumlah Dokumen Tidak Relevan Terambil	Jumlah Dokumen Relevan pada koleksi
1	30	10	33
2	12	2	20
3	43	3	50
4	4	5	5
5	10	1	12
6	17	0	20

Hitung *recall*, *precision* dan *F-Measure* menggunakan *microaveraging* ?

SOAL UJIAN TENGAH SEMESTER 2014/2015



Ujian Tengah Semester Genap 2014/2015
Jurusan Teknik Informatika
FSM UNDIP Semarang

9

Mata Kuliah	: Sistem Temu Balik Informasi	Dosen	: - Sukmawati Nur Endah, S.Si, M.Kom - Sutikno, M.Cs
Beban	: 3 SKS	Hari/Tgl	: Rabu / 22 April 2015
Sifat	: Closed Book	Waktu	: 90 menit

Catatan :

- Tidak boleh menggunakan HP, hanya boleh menggunakan kalkulator konvensional!
- Nilai 0 akan diberikan bagi mahasiswa yang berbuat curang dalam bentuk apapun!
- Masing-masing soal mempunyai bobot nilai 20

Soal:

- Jelaskan pengertian korpus dan jenis-jenis isi korpus!
 - Menurut distribusi Zipf, jika suatu kata pada urutan 15 mempunyai frekuensi 8000, berapa frekuensi kata yang berada pada urutan 300?
- Jelaskan dan gambarkan skema *logical view* sebuah dokumen dalam IR !
- Uraikan permasalahan yang dapat ditimbulkan akibat penggunaan stemmer dalam proses pengindeksan beserta contohnya!
- Diketahui data sebagai berikut :

Kata Dok	X1	X2	X3	X4	X5
D1	1	1	1	0	1
D2	1	0	1	1	1
D3	0	1	1	0	0
D4	0	0	0	1	0
D5	0	1	1	0	1
D6	1	0	1	1	0

Dengan menggunakan model Boolean, buatlah 2 (dua) query yang dapat menghasilkan dokumen D2 dan D6, buktikan kebenarannya!

- Gunakan *vector space model* (menggunakan pembobotan *tf.idf*) untuk merangking dokumen yang terretrieve dengan data berikut ini.

Query : Propinsi Jawa Tengah
Dokumen 1 : Propinsi di Indonesia ada puluhan
Dokumen 2 : Jawa Tengah ada di Pulau Jawa
Dokumen 3 : Peta propinsi Jakarta ada di tengah meja

&%& Selamat Mengerjakan &%&

SOAL UJIAN AKHIR SEMESTER 2014/2015

Mata Kuliah	: SISTEM TEMU BALIK INFORMASI
Kelas	: A dan B
Pengampu	: Sukmawati Nur Endah, S.Si., M.Kom. Dr. Retno Kusumaningrum, S.Si., M.Kom. ✓
Jurusan	: Informatika
Hari/Tanggal	: Rabu, 8 Juli 2015
Jam/Ruang	: 08.00 – 09.40 (100 menit) / E101 dan A101
Sifat Ujian	: TUTUP BUKU

Jawablah pertanyaan berikut pada lembar jawab yang telah disediakan!

1. (Nilai 30) Jika diketahui sebuah hasil perolehan informasi (R : Relevan dan TR : Tidak Relevan) adalah sebagai berikut :

Rank	1	2	3	4	5	6	7	8	9	10
Query1	R	R	R	TR	R	TR	R	TR	TR	TR
Query2	R	R	TR	R	R	R	TR	TR	TR	TR

Tentukan nilai 11-point Average Precision (P_{11-pt}) dari hasil perolehan di atas!

- (Nilai 5) Jelaskan apa fungsi/kegunaan penerapan **pengukuran Kappa**!
- (Nilai 20) Gambarkan perbedaan arsitektur *relevance feedback* dan *pseudo-relevance feedback*!
- (Nilai 10) Sebutkan 2 permasalahan “teknik berdasarkan *knowledge* untuk pelacakan *free text* pada CLIR” yang menggunakan **kamus dwibahasa** untuk menerjemahkan *query*! Berikan contohnya masing-masing!
- (Nilai 15) Jika diketahui sebuah *paragraph* yang berhasil diekstraksi oleh *Lasso QA System* (dimana k_1 , k_2 , dan k_3 adalah kata kunci yang ditemukan pada paragraph terkait) sebagai berikut :

Paragraph p1

-----2 words----- k_1 -----3 words-----
 -----2 word----- k_2 -----2 word-----1 word-----
 -----3 words----- k_3 -----

Tentukan nilai **max_distance_sequence** dari keseluruhan *paragraph window* yang mungkin terbentuk!

6. (Nilai 20) Isilah tabel berikut untuk menunjukkan perbedaan jenis ringkasan berbentuk **ekstrak** dan **abstrak**!

Perbedaan	Ekstrak	Abstrak
Bentuk luaran (<i>output</i>)		
Pendekatan (<i>approach</i>)		

~ SELAMAT MENGERJAKAN ~

SOAL UJIAN TENGAH SEMESTER 2017/2018



Ujian Tengah Semester Gasal 2017/2018 Departemen Ilmu Komputer/Informatika FSM UNDIP Semarang

Mata Kuliah	: Temu Balik Informasi	Dosen	: - Sukmawati Nur Endah, S.Si, M.Kom - Dr. Retno Kusumaningrum
Beban	: 3 SKS	Hari/Tgl	: Jumat / 13 Oktober 2017
Sifat	: Closed Book	Waktu	: 100 menit

Catatan :

- Nilai 0 akan diberikan bagi mahasiswa yang berbuat curang dalam bentuk apapun!

Soal:

1. Jelaskan dan gambarkan proses yang terjadi dalam menemukan kembali informasi (informaton retrieval) dalam domain teks !
Bobot nilai = 20
2. Jelaskan tahapan dalam pengindeksan?
Bobot nilai = 20
3. Jika A adalah matriks term-dokumen, jelaskan langkah-langkah pemodelan dengan menggunakan Latent Semantic Analysis!
Bobot nilai = 20

Untuk soal no 4 dan 5 gunakan koleksi dokumen berikut ini!

Dokumen	Isi Dokumen
D1	Publikasi artikel ilmiah wajib bagi mahasiswa dan dosen
D2	Dosen membimbing mahasiswa dalam penulisan artikel ilmiah
D3	Artikel ilmiah dapat dipublikasikan di jurnal nasional atau internasional
D4	Workshop penulisan artikel di jurnal nasional wajib bagi mahasiswa
D5	Dosen diwajibkan publikasi di jurnal internasional

4. Dengan menggunakan data di atas, dokumen apa saja yang akan terretrieve dengan query sebagai berikut ini:
 - a. Artikel AND jurnal AND (nasional OR internasional)
 - b. NOT publikasi AND (artikel OR ilmiah)
 - c. Dosen NEAR 6 artikel
 - d. Jurnal NEAR 2 internasional
 - e. Artikel WITH ilmiah
 Bobot nilai = 20
5. Gunakan *vector space model* (menggunakan pembobotan tf.idf) untuk merangking dokumen yang terretrieve dengan koleksi dokumen di atas untuk query dan daftar stopwords sebagai berikut.

Query : Publikasi jurnal nasional

Daftar Stopword : bagi, dan, dalam, dapat, di, atau

Catatan: Lakukan proses stemming dan penghilangan stopwords!

 Bobot nilai = 20

&%& **Selamat Menengerjakan** &%&

SOAL UJIAN AKHIR SEMESTER 2017/2018

Mata Kuliah	: SISTEM TEMU BALIK INFORMASI
Kelas	: A
Pengampu	: Sukmawati Nur Endah, S.Si., M.Kom. Dr. Retno Kusumaningrum, S.Si., M.Kom.
Departemen	: Informatika
Hari/Tanggal	: Jumat, 15 Desember 2017
Jam/Ruang	: 08.00 – 09.40 (100 menit) / B103
Sifat Ujian	: TUTUP BUKU

Jawablah pertanyaan berikut pada lembar jawab yang telah disediakan!

- (Nilai 30) Seorang pengguna diminta untuk menjalankan sebuah query ke dalam dua buah sistem dan menghasilkan dua hasil perolehan sebagai berikut:
 - Sistem A menghasilkan 5 buah dokumen yang relevan pada urutan ke-3, ke-6, ke-10, ke-15, dan ke-20.
 - Sistem B menghasilkan 8 buah dokumen yang relevan pada urutan ke-1, ke-2, ke-8, ke-16, ke-17, ke-18, ke-19, dan ke-20.

Diasumsikan tidak ada lagi dokumen yang relevan untuk masing-masing hasil perolehan tersebut pada korpus yang dimiliki.

 - Tentukan nilai *Precision* dan *Recall* untuk masing-masing kemunculan perolehan dokumen dari kedua sistem tersebut!
 - Tentukan mana sistem yang memiliki kinerja terbaik? Jelaskan!
- (Nilai 20) Jelaskan perbedaan *relevance feedback* dan *pseudo-relevance feedback* serta tunjukkan perbedaan gambar arsitekturnya!
- (Nilai 20) Salah satu cara penerjemahan pada *Cross-Language Information Retrieval* berbasis *Query Translation* adalah menggunakan mesin penerjemah. Sebutkan 2 (dua) kekurangan dari cara penerjemahan tersebut!
- (Nilai 20) Jelaskan 4 (empat) kesulitan / permasalahan yang muncul dalam pengindeksan pada *Geographic Information Retrieval* serta berikan masing-masing contoh kasusnya!
- (Nilai 10) Jelaskan apa yang Anda ketahui dengan *Music Recommender System* dan sebutkan salah satu contoh aplikasinya!

~ SELAMAT MENGERJAKAN ~

SOAL UJIAN TENGAH SEMESTER 2018/2019



Ujian Tengah Semester Gasal 2018/2019 Departemen Ilmu Komputer/Informatika FSM UNDIP Semarang

Mata Kuliah	: Temu Balik Informasi	Dosen	: - Sukmawati Nur Endah, S.Si, M.Kom - Dr. Retno Kusumaningrum
Beban	: 3 SKS	Hari/Tgl	: Jumat / 12 Oktober 2018
Sifat	: Closed Book	Waktu	: 100 menit

Catatan :

- Tidak boleh menggunakan HP, hanya boleh menggunakan kalkulator konvensional!
- Nilai 0 akan diberikan bagi mahasiswa yang berbuat curang dalam bentuk apapun!

Soal:

- a. Bagaimana bunyi hukum Zipf?
 - b. Menurut distribusi Zipf, jika suatu kata pada urutan 35 mempunyai frekuensi 12000, berapa frekuensi kata yang berada pada urutan 105?

Bobot nilai = 10

- a. Pembobotan kata mempunyai beberapa metode atau teknik.
 - b. Jelaskan metode dalam pembobotan kata yang Anda gunakan dalam mengerjakan tugas besar STBI!
 - c. Dari hasil eksperimen yang Anda lakukan, metode mana yang terbaik menurut Anda? Jelaskan!

Bobot nilai = 20

- a. Jika A adalah matriks term-dokumen, jelaskan langkah-langkah pemodelan dengan menggunakan Latent Semantic Analysis!

Bobot nilai = 20

- Berikut adalah indeks posisi dengan format sebagai berikut :

Kata : dokumen1 : posisi1, posisi 2, dst; dokumen2 : posisi1, posisi 2, dst

komputer	: 1 : 12, 45, 51;	2 : 23, 35;	3 : 27, 52, 60;	4 : 4, 21;
disiplin	: 1 : 2, 18;	2 : -;	3 : 1, 7;	4 : 16, 19;
ilmu	: 1 : 9, 20;	2 : 13, 34;	3 : 24;	4 : 11, 23, 32
kejahatan	: 1 : 4, 7, 44;	2 : 12, 48, 57	3 : 26, 30;	4 : -;
jurusan	: 1 : 8;	2 : 33, 41;	3 : -;	4 : 22;

Dokumen manakah yang memuat frase berikut ini?

- a. Kejahatan komputer → dok 1 2 3
- b. Jurusan Ilmu Komputer → 1 2, 4

Manakah dokumen yang akan terretrieve dengan query sebagai berikut :

- c. Jurusan (near 4) komputer → 1 2 4
- d. Ilmu (with) komputer → 2
- e. Disiplin (near 10) ilmu → 1 4

Bobot nilai = 20

- Gunakan *vector space model* (menggunakan pembobotan tf.idf) untuk meranking dokumen yang terretrieve dengan koleksi dokumen di bawah ini untuk query dan daftar stopwords sebagai berikut.

- D1 : Publikasi artikel ilmiah wajib bagi mahasiswa dan dosen
- D2 : Dosen membimbing mahasiswa dalam penulisan artikel ilmiah
- D3 : Artikel ilmiah dapat dipublikasikan di jurnal nasional atau internasional
- D4 : Workshop penulisan artikel di jurnal nasional wajib bagi mahasiswa

Query : Publikasi jurnal nasional

Daftar Stopword : bagi, dan, dalam, dapat, di, atau

Catatan: Lakukan proses stemming dan penghilangan stopwords!

Bobot nilai = 30