

Kuis Besar PTBI
Genap 14/15
Rabu, 13 Januari 2016

Jawablah pertanyaan-pertanyaan berikut ini dengan tepat.

Gunakan kertas A4 dan jawaban ditulis tangan.

Jawablah secara berurutan.

Kumpulkan di hari UTS: Sabtu, 16 Januari 2016 (10.00 - 12.00)

1. Hal apakah menjadi perbedaan utama antara dokumen dalam konteks temu balik informasi dan *record* dalam konteks basis data?
2. Hal apakah yang menjadi permasalahan utama (*core issue*) dalam bidang ilmu temu balik informasi?
3. Berikan 1 contoh penerapan temu balik informasi dalam tugas-tugas (*tasks*) berikut ini:
 - a. *Ad-hoc search*
 - b. *Filtering*
 - c. *Classification*
 - d. *Clustering*
 - e. *Question answering*
4. Apakah yang dimaksud dengan 'relevansi' dalam temu balik informasi?
5. Apakah kaitan antara temu balik informasi dan *search engine*?
6. Jelaskan dan berikan contoh penerapan konsep dalam temu balik informasi berikut ini:
 - a. *Crawler*
 - b. *Indexer*
 - c. *Retriever*
7. Diketahui dua buah kueri dengan hasil temu balik sebagai berikut (x = hasil relevan):

Kueri 1

Ranking	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Relevansi	x		x			x				x										x

Kueri 2

Ranking	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Relevansi	x		x												x					

- a. Hitunglah nilai *precision* dan *recall* untuk setiap kueri
- b. Hitunglah nilai *average precision* untuk setiap kueri
- c. Hitunglah nilai *mean average precision* (MAP) untuk gabungan kedua kueri
- d. Gambarkan grafik *precision-recall* untuk kedua kueri tsb (dalam satu gambar)
8. Mengapa penerapan hukum Zipf penting dalam bidang temu balik informasi?
9. Jelaskan secara singkat tahapan pra-pemrosesan dokumen sebagai berikut: *tokenizing*, *stemming*, *remove stopwords*.
10. Apakah yang dimaksud dengan *inverted index*?

Jika diketahui tiga dokumen dengan isi sebagai berikut (untuk nomor 11 dan 16):

Dokumen 1:

Pada suatu hari yang ramai saya mengendarai sepeda motor beroda dua dan bertemu dengan dua polisi berbintang dua.

Dokumen 2:

Pengendara sepeda motor di jalan yang ramai wajib menaati peraturan lalu lintas dan menggunakan helm berkaca dua.

Dokumen 3:

Paman adalah seorang polisi yang hobinya mengendarai motor besar dengan dua roda yang membuat suara ramai.

11. Kembangkan sebuah *inverted index* (tanpa menggunakan *stemmer* dan tidak membuang *stopwords*), dengan menggunakan tiga dokumen di atas tersebut (*posting* berisi dokumen dan frekuensi kata).
12. Ubahlah nilai angka 20 dengan menggunakan Elias- γ dan Elias- δ .
13. Diketahui posisi sebuah *posting* kata dalam dokumen adalah: 1, 6, 13, 24, 68, 105. Apakah nilai *delta encoding* untuk posisi tersebut?
14. Jelaskan perbedaan implementasi temu balik dengan konsep: *document-at-a-time* dan *term-at-a-time*. Gambarkan perbedaan tersebut jika diperlukan.
15. Apakah yang dimaksud dengan *conjunctive search*?
16. Berikan indeks *boolean* untuk dokumen-dokumen yang disediakan. Jika seandainya ada sebuah kueri '**dua DAN polisi**', hasil apakah yang diberikan sistem temu balik dengan *boolean retrieval*?
17. Jika seandainya Anda diminta untuk memberikan preferensi dokumen yang ditemukan pada nomor 16, bagaimana Anda melakukannya dan apakah hasilnya?