



Konsep Dasar Sistem Temu Kembali Informasi

Oleh : Rahmat Robi Waliyansyah, M.Kom.

Database Retrieval vs IR

Database retrieval

- Dokumen mana yang mengandung himpunan keyword?
- Semantik didefinisikan dengan baik
- Error dari suatu obyek mengakibatkan kegagalan!

Information retrieval

- Informasi mengenai suatu subyek atau topik
- Semantik dapat bersifat lepas (longgar)
- Error kecil ditoleransi

Sistem temu kembali informasi

- Sistem temu kembali informasi merupakan sebuah sistem yang berguna dalam memanggil dan menempatkan dokumen dari/dalam basis data sesuai dengan permintaan pengguna.
- Sistem temu kembali informasi memiliki tujuan akhir, yaitu memberikan kepuasan informasi bagi pengguna sistem. Jadi, temu kembali informasi merujuk pada keseluruhan kegiatan yang meliputi pembuatan wakil informasi (representation), penyimpanan (storage), pengaturan (organization) sampai kepada pengambilan (access).



CARA SISTEM TEMU KEMBALI INFORMASI

- **Manual** : dengan menggunakan publikasi tercetak.
- **Offline** : dengan menggunakan pangkalan data terpasang yang dibangun sendiri atau CD ROM.
- **Online** : dengan akses langsung ke penyedia pangkalan data yang ada di dalam maupun luar negeri.



Information Retrieval

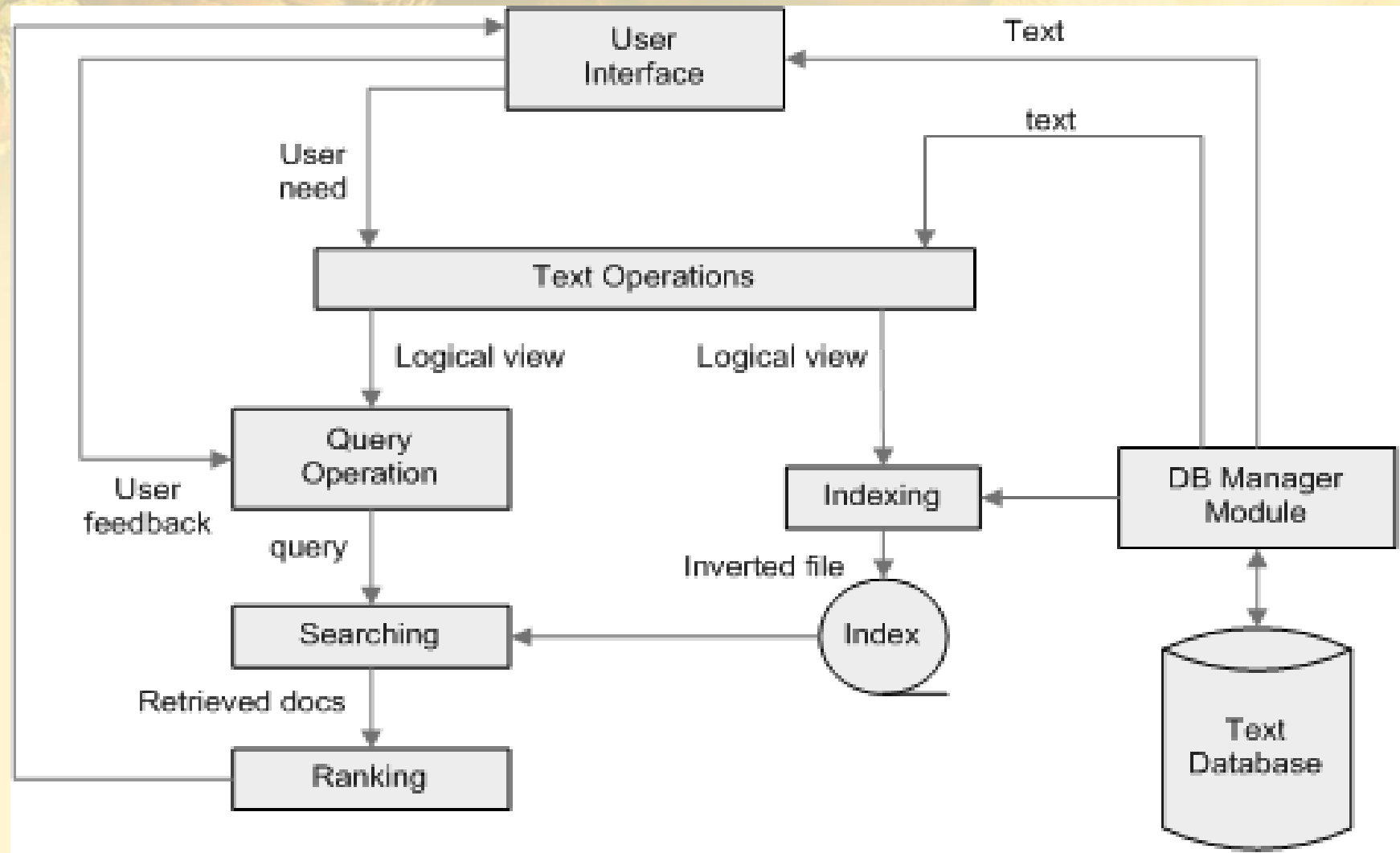
- Secara teknis: **indexing** (pembuatan *index*) dan **retrieval** (pencarian keterangan) dokumen *textual*.
- Pencarian halaman pada WWW adalah aplikasi paling “*ngetop*” saat ini.
- Fokus pertama: me-*retrieve* dokumen- dokumen yang **relevan** dengan query.
- Fokus kedua: me-*retrieve* himpunan **besar** dokumen secara **efisien**.
- Tujuan: **Me-*retrieve* semua dokumen yang **relevan** sekaligus me-*retrieve* sesedikit mungkin dokumen yang **tidak relevan****

Relevan

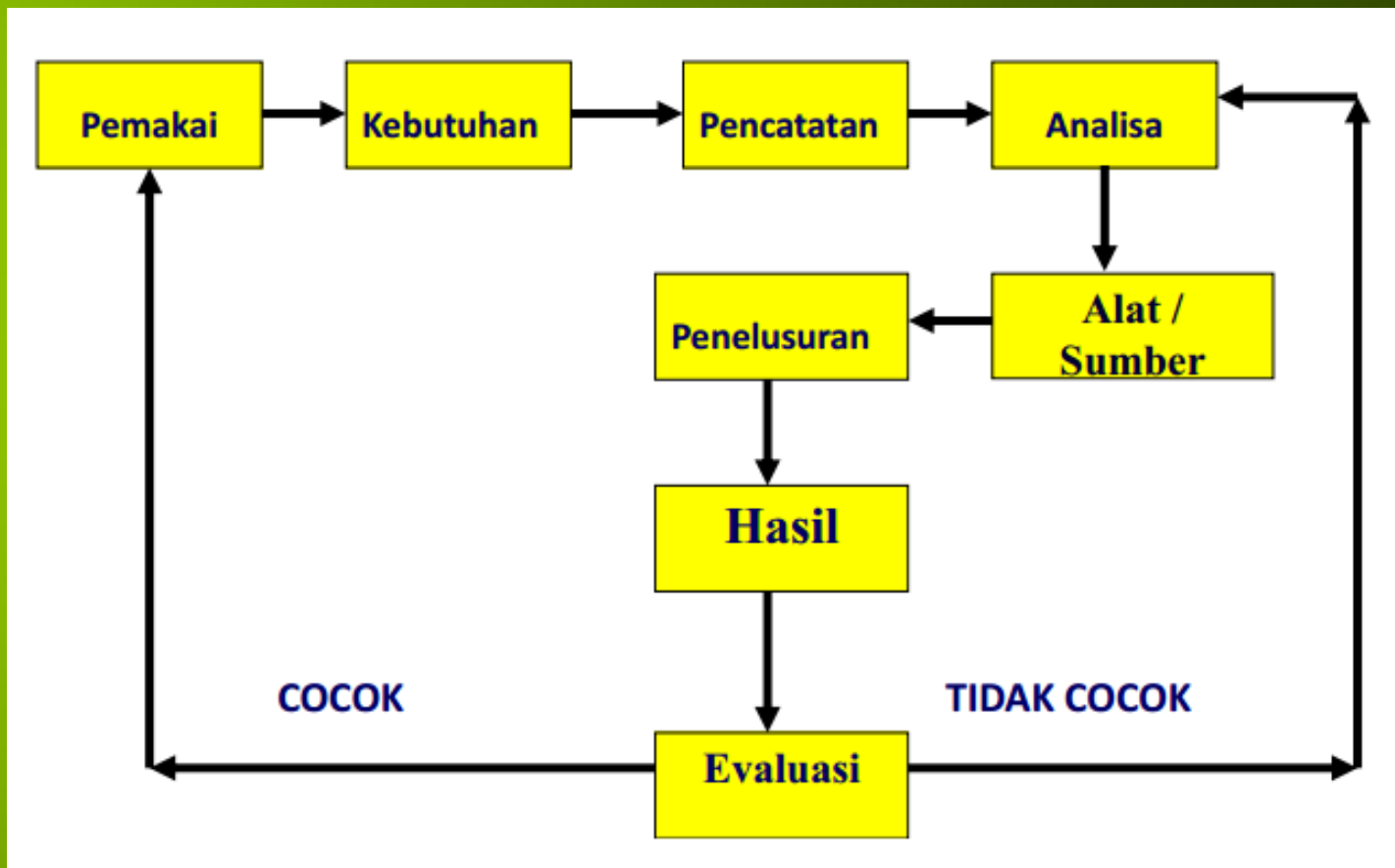


- Relevansi merupakan suatu *judgment* (keputusan) subyektif dan dapat didasarkan pada:
 - Topik yang tepat.
 - Waktu (informasi terbaru).
 - Otoritatif (dari suatu sumber terpercaya).
 - Kebutuhan informasi dari pengguna.
- **Kriteria relevansi** utama: suatu sistem IR sebaiknya (harus) memenuhi kebutuhan informasi pengguna.

Sistem IR



Tahapan



Konsep



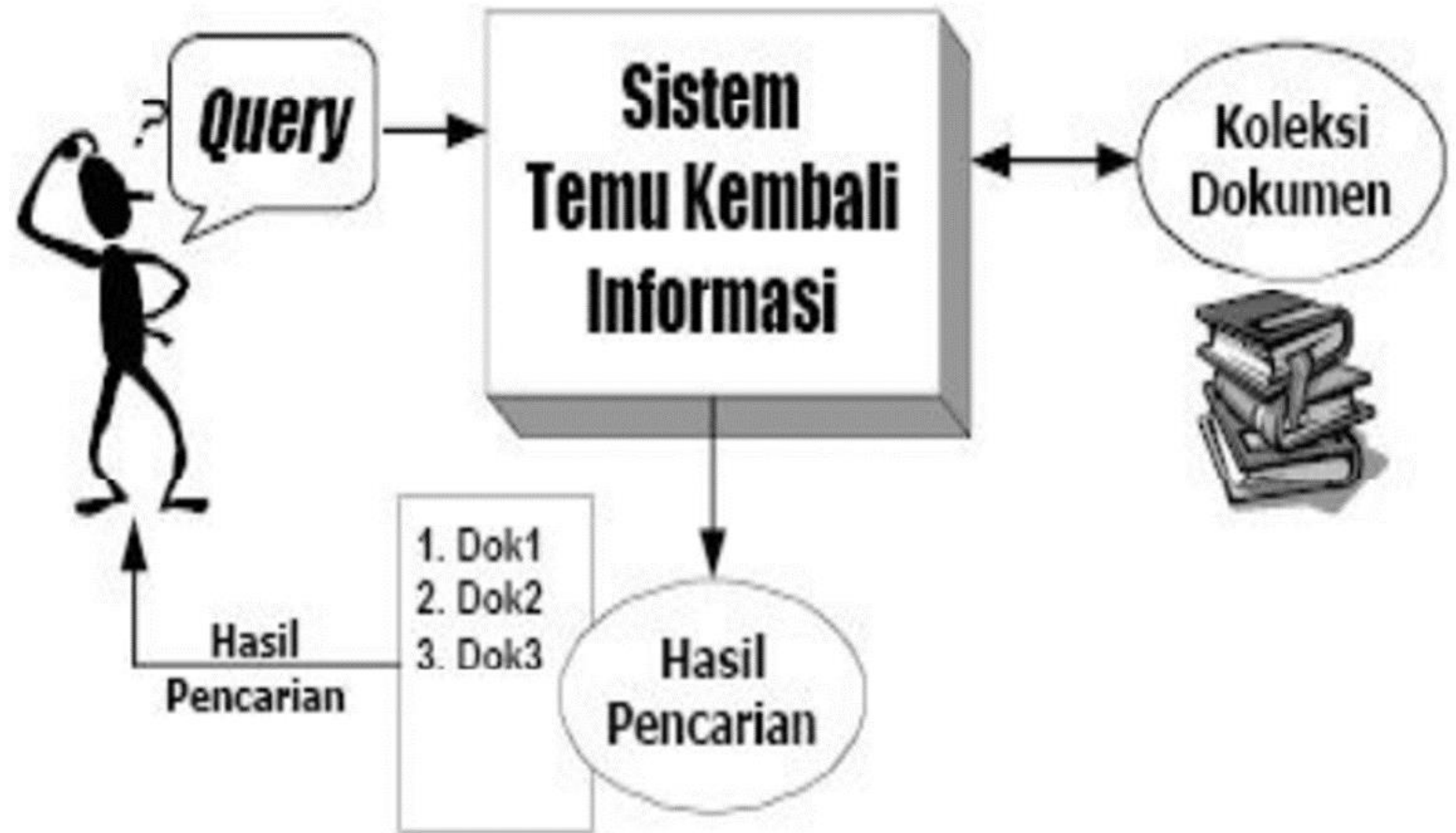
"Representation" dari gambar di sebelah kiri menunjukkan representasi dokumen data , data dan informasi. "Query" pada komponen sebelah 'kanan merupakan representasi dari pertanyaan pengguna, serta "matching function" komponen yang di tengah merupakan fungsi pencocokan antara representasi data/dokumen dengan pertanyaan.



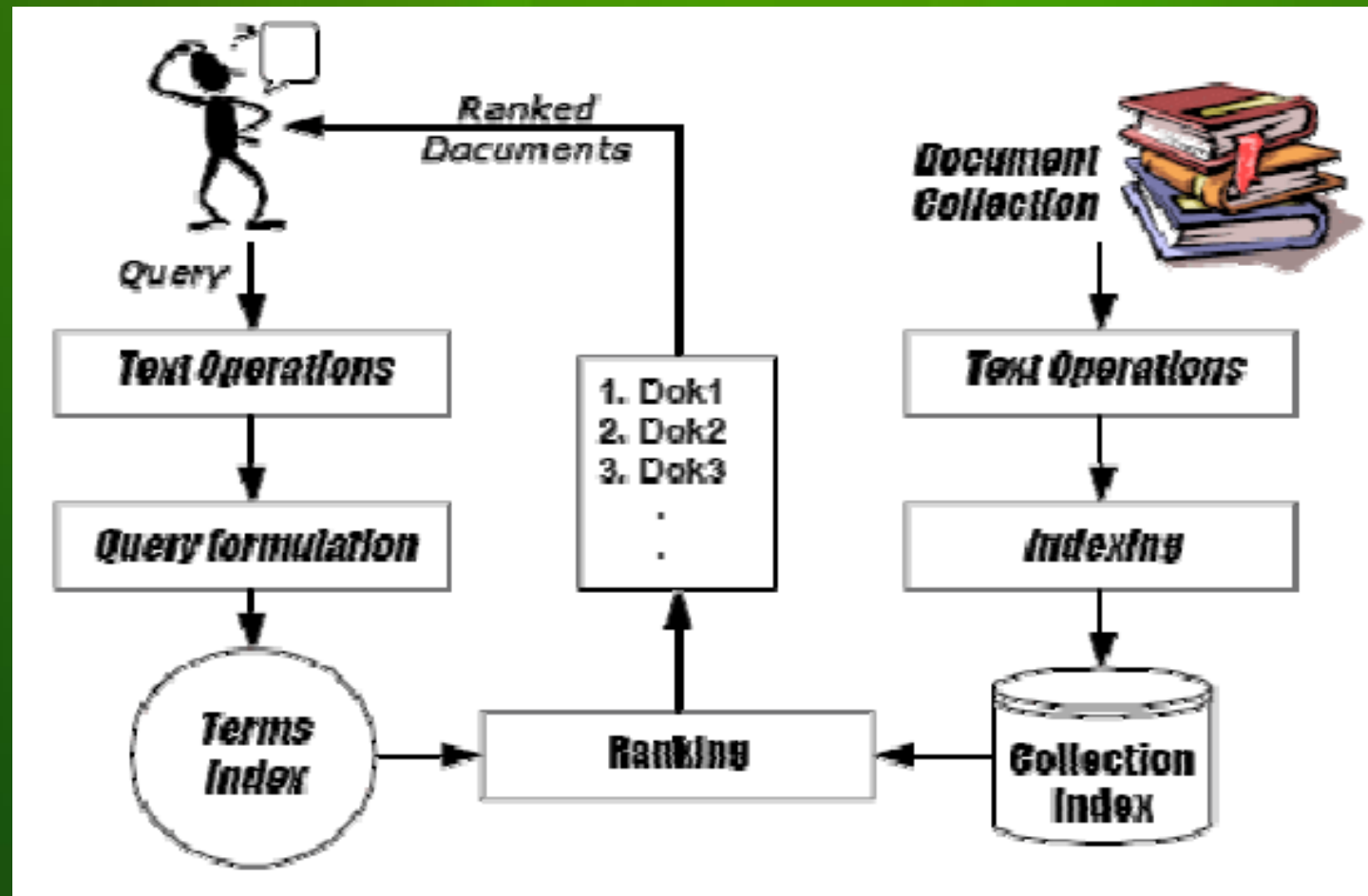
KONSEP SISTEM TEMU KEMBALI INFORMASI, BELKIN (1980) :

1. Penulis mempresentasikan sekumpulan ide dalam sebuah dokumen menggunakan sekumpulan konsep.
2. Terdapat beberapa pengguna yang memerlukan ide yang dikemukakan oleh penulis tersebut, tapi mereka tidak dapat mengidentifikasikan dan menemukannya dengan baik.
3. Sistem temu kembali kembali informasi bertujuan **untuk mempertemukan ide** yang dikemukakan oleh penulis dalam dokumen dengan kebutuhan informasi pengguna yang dinyatakan dalam bentuk pertanyaan (query)

Ilustrasi



BAGIAN-BAGIAN





SUBSISTEM SISTEM KEMBALI INFORMASI

1. Subsistem dokumen
2. Subsistem pengindeksan
3. Subsistem kosa kata
4. Subsistem pencarian
5. Subsistem antar muka pengguna sistem
6. Subsistem penyesuaian

KOMPONEN SISTEM KEMBALI INFORMASI

❖ Tague-Sutcliffe (1996) melihat Sistem Temu Kembali Informasi sebagai suatu proses yang terdiri dari 6 (enam) komponen utama yaitu:

1. Kumpulan dokumen
2. Pengindeksan
3. Kebutuhan informasi pemakai
4. Strategi pencarian
5. Kumpulan dokumen yang ditemukan
6. Penilaian relevansi



Pencarian Keyword

- Ide paling sederhana dari relevansi: **apakah string query ada di dalam dokumen** (kata demi kata, verbatim)?
- Ide yang lebih fleksibel: **Berapa sering kata-kata di dalam query muncul di dalam dokumen, tanpa melihat urutannya (*bag of words*)?**



Masalah dengan Keyword

- Mungkin tidak me-*retrieve* dokumen relevan yang menyertakan ***synonymous terms***.
 - “restaurant” vs. “café”
 - “NDHU” vs. “National Dong Hwa University”
- Mungkin me-*retrieve* dokumen tak-relevan yang menyertakan ***ambiguous terms***.
 - “bat” (baseball vs. mamalia)
 - “Apple” (perusahaan vs. buah-buahan)
 - “bit” (unit data vs. perilaku menggigit)



IR Cerdas

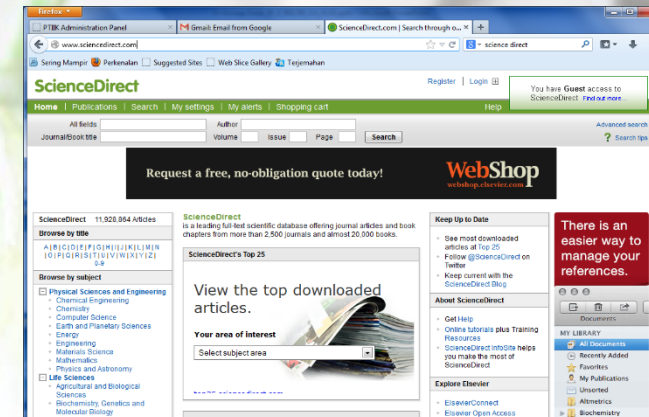
- Memanfaatkan **pengertian** atau makna dari kata yang digunakan.
- Melibatkan **urutan kata** di dalam query.
- Beradaptasi dengan pengguna berdasarkan pada ***feedback***, langsung atau tidak langsung.
- **Memperluas** pencarian dengan term terkait.
- Mengerjakan **pemeriksaan ejaan/perbaikan tanda pengenal** otomatis.
- Memanfaatkan **Otoritas** dari sumber.

Portal Jurnal

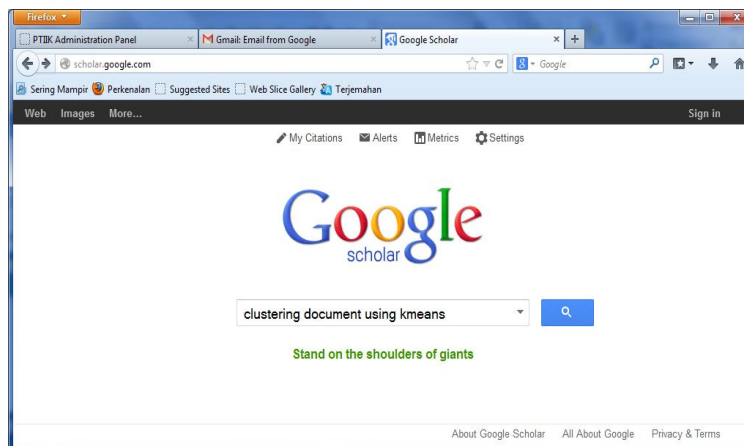
- Computer.org (gunakan proxy UB) :



- <http://www.sciencedirect.com/> :



- <http://scholar.google.com/>



- Other


Fungsi utama Sistem Temu Kembali Informasi menurut Lancaster (1979) dan Kent (1971) :

- Mengidentifikasi sumber informasi yang relevan dengan minat masyarakat pengguna yang ditargetkan.
- Menganalisis isi sumber informasi (dokumen).
- Merepresentasikan isi sumber informasi dengan cara tertentu yang memungkinkan untuk dipertemukan dengan pertanyaan (query) pengguna.
- Merepresentasikan pertanyaan (query) pengguna dengan cara tertentu yang memungkinkan untuk dipertemukan sumber informasi yang terdapat dalam basis data.
- Mempertemukan pernyataan pencarian dengan data yang tersimpan dalam basis data.
- Menemu-kembalikan informasi yang relevan.
- Menyempurnakan unjuk kerja sistem berdasarkan umpan balik yang diberikan oleh pengguna.



Perbandingan antara Sistem Temu Kembali Informasi (IRS), Sistem Manajemen Basis Data (DBMS) dan Sistem Kecerdasan Buatan (AI)

	OBJEK DATA	FUNGSI	UKURAN
IRS	Dokumen	Temu-kembali (Probabilistik)	Kecil-besar
DBMS	Tabel	Temu-kembali (Deterministik)	Kecil-besar
AI	Pernyataan Logika	Inferensia	Kecil



Sebuah IRS yang efektif dan terpercaya harus (Onwuchekwa, edeama O. & Olumakinde Richard Jegede) :

- Menyebarkan informasi dengan cepat
- Memfilter informasi
- Jumlah informasi yang tepat pada saat yang tepat
- Mendapatkan informasi secara ekonomis
- Literatur saat ini
- Membantu secara personal

Regular Expression (RegEx)

- Regular Expression atau yang lebih sering disebut regex merupakan sebuah teknik yang digunakan untuk mencocokkan string teks, seperti karakter tertentu, kata-kata, atau pola karakter.
- RegEx memiliki 2 fungsi utama yakni mencari dan mengganti, mencari suatu pola tertentu dalam text lalu menggantinya menjadi pola yang lain.



Prinsip Kerja RegEx

- Engine RegEx terdiri dari 2 jenis Text-directed engine dan regex-directed engine atau ada juga yang mengatakan DFA (Deterministic Finite Automaton) dan NFA (Nondeterministic Finite Automaton) engines.
- Jenis mesin engine yang lebih banyak diminati adalah regex-directed engine, disamping itu featurenya lebih hebat dari text-directed engine.



Prinsip Kerja RegEx

- Misalnya saja kita mencocokkan kata “**regex**” dengan kata “**belajar regex sekarang**”, pertama-tama regex akan memulai mencocokkan karakter pertama yaitu huruf “R” dengan “B”, karena tidak cocok maka akan dilanjutkan ke tempat selanjutnya yaitu “R” dengan “E”, karena masih belum cocok juga proses ini akan terus berlanjut, hingga akhirnya menemukan huruf “R” yaitu di posisi ke-7, hasil ini akan disimpan di memori bahwa telah valid pada posisi ke-7.
- Proses ini akan terus diulang hingga mencapai posisi akhir suatu himpunan string. Jika kata “regex” ditemukan oleh kata “belajar regex sekarang” maka engine akan melaporkan bahwa Regular Expression telah valid.

Selesai

