

Sistem Temu Kembali Informasi/ Information Retrieval

Konsep Dasar Sistem Temu Kembali Informasi





Dokumen





Penyimpanan yang Terorganisasi





Database

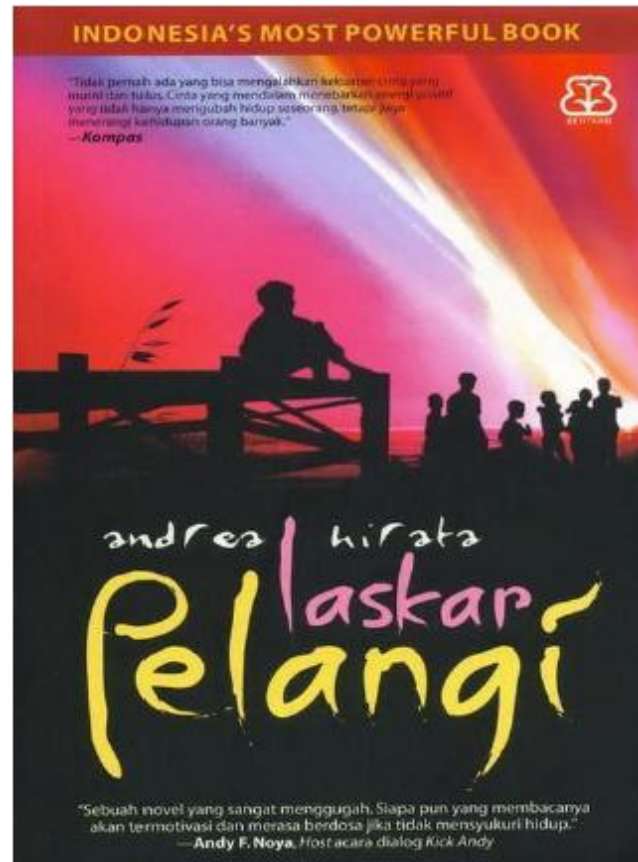
Database Mahasiswa

Buku ID	Nama Buku	Pengarang
001	Information Retrieval	Ricardo baeza
002	Matematika Diskrit	Rinaldi Munir
003	Pengenalan Java	Abdul Kadir
004	Pengenalan C++	Abdul Kadir
005	Design Pattern	Arnold



Unstructured

Siapa pemain dalam novel tersebut ?





Digital Library UB

www.digilib.ub.ac.id/opac/search.aspx

digilib ub

Central Library

[Home](#) [My library](#) [Administrative](#) [Services](#) [News](#) [Ask Librarian](#) [Sitemap](#)

OPAC (Online Public Access Catalog)

Search : by : Year : Exact word

Location : Criteria :

Search Result

Record found : 0 record(s)

[Page navigation](#)

No Record Found !

Student Administration

[Login](#)

[Register](#)

Other Services

[Knowledge Garden](#)

[InterLibrary Loan](#)

[Lentera](#)

[News](#)

[OPAC](#)



Google

Browser address bar: <https://www.google.com/#q=data+retrieval>

Navigation: +You Search Images Maps Play YouTube News Gmail Drive Calendar More -

Search bar: data retrieval [SEARCH] [SIGN IN]

Filters: Web Images Maps Shopping Patents More Search tools

About 14,400,000 results (0.21 seconds)

Ads related to **data retrieval**

- Database Recovery** 6319 0088
www.ontrackdatarecovery.com.sg/
Specialised in Recover All Complex Failed Media Types. No Data No Fee!
HDD Recovery Laptop Recovery
Apple Recovery Mobile Recovery
- Penyelamat Data Indonesia - penyelamatdata.com**
www.penyelamatdata.com/
Data Recovery Specialist Indonesia, Data Hilang/Rusak, Serahkan Ahlinya
- Recovery Data Hardisk Jkt** 0819-868-155
datarecoverynergicom.blogspot.com/
Harddisk Rusak. Perlu Data Penting? Menyelamatkan Data Harddisk 24hours

Data retrieval - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Data_retrieval
Data retrieval, in database management, involves extracting the wanted data from a database. The two primary forms of the retrieved data are reports and ...

2013-2014 IRS Data Retrieval Tutorial - Financial Aid Office
www.finaid.ucsb.edu/media/videos/1314/english/irsdrf2013.html



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



Pengertian IR

- Pencarian materi (biasanya dokumen) dari sesuatu yang sifatnya tak-terstruktur (*unstructured*, biasanya teks) untuk memenuhi kebutuhan informasi dari dalam koleksi besar (biasanya disimpan dalam komputer).
- Representasi, penyimpanan, organisasi, pencarian dan akses ke item informasi untuk memenuhi kebutuhan informasi pengguna.
- Penekanan pada proses *retrieval* informasi (bukan data).
- Karakterisasi kebutuhan informasi tidaklah mudah. Harus ditranslasi ke dalam suatu query terlebih dahulu.



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**

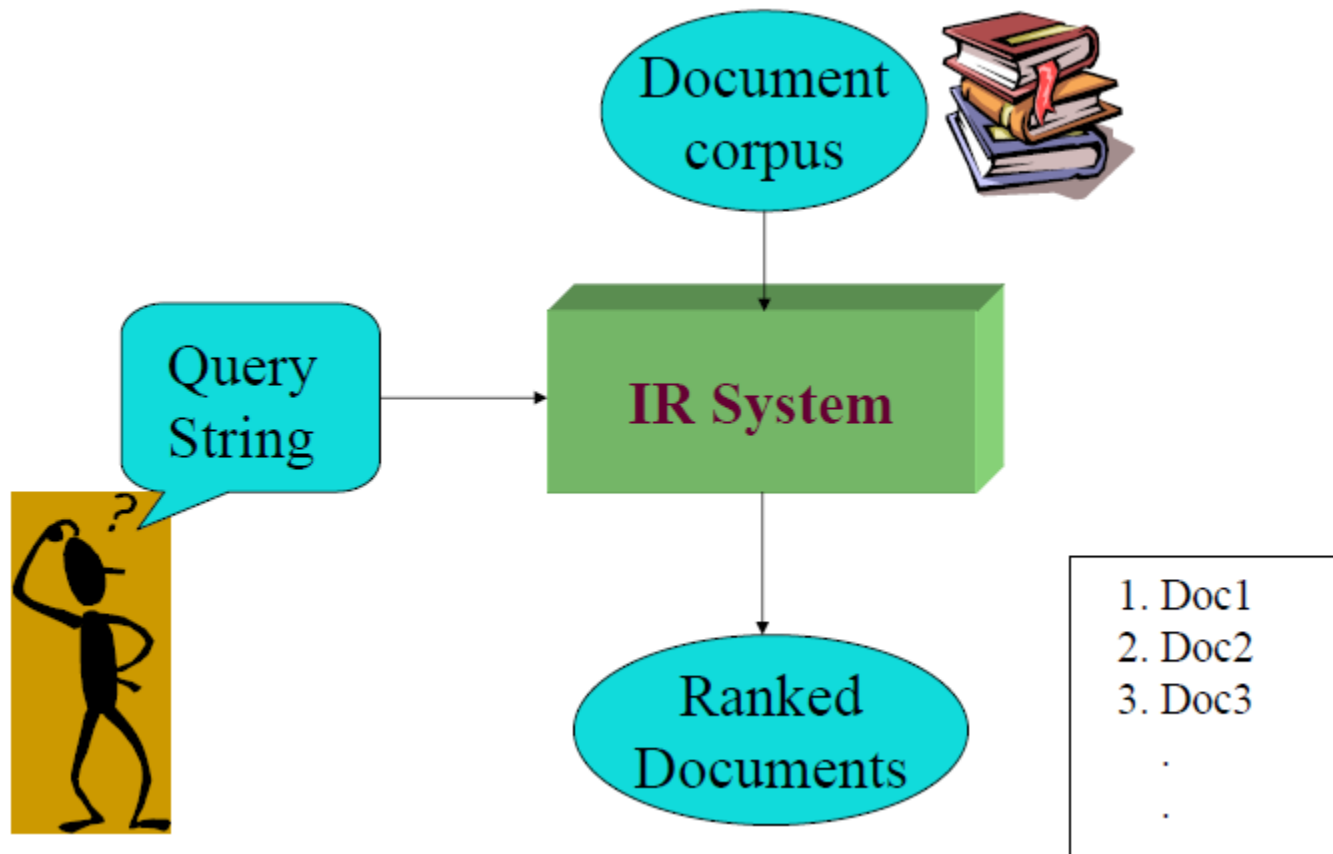


Relevan

- 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.

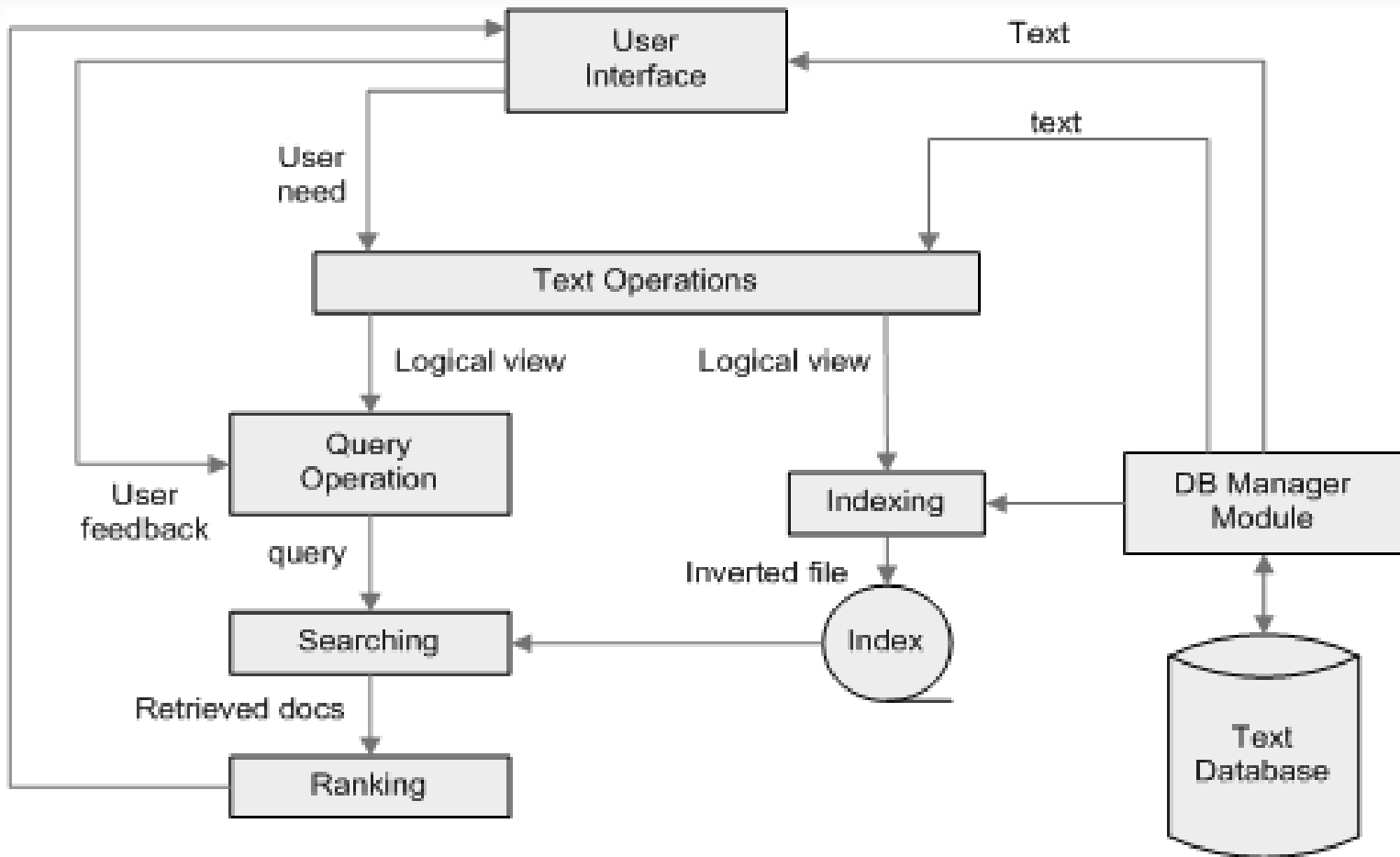


Sistem IR





Sistem IR





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)



Bukan Sekedar Keyword

- Kita akan mendiskusikan dasar-dasar IR berbasis *keyword*, tetapi...
 - Fokus pada perluasan dan pengembangan terakhir untuk mendapatkan hasil terbaik.
- Kita akan membahas dasar-dasar pembangunan sistem IR yang efisien, tetapi...
 - Fokus pada algoritma dan kemampuan dasar, bukan masalah sistem yang memungkinkan pengembangan ke database ukuran industri.



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



Perkembangan IR

1. Klasifikasi Dokumen
2. Clustering Dokumen
3. Peringkasan Teks
4. Question Answering System

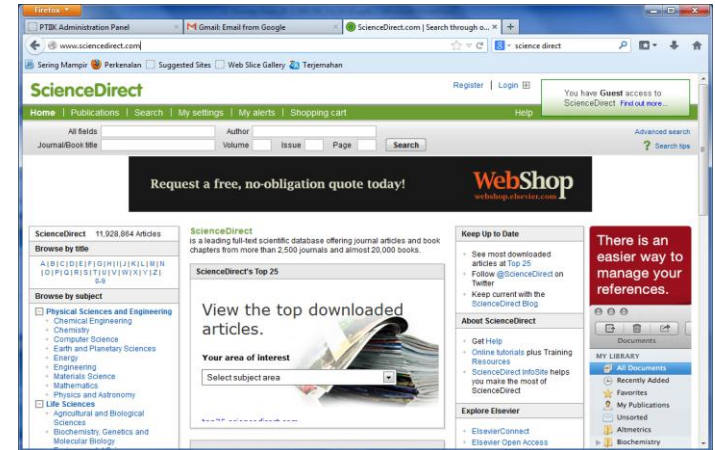


Portal Jurnal

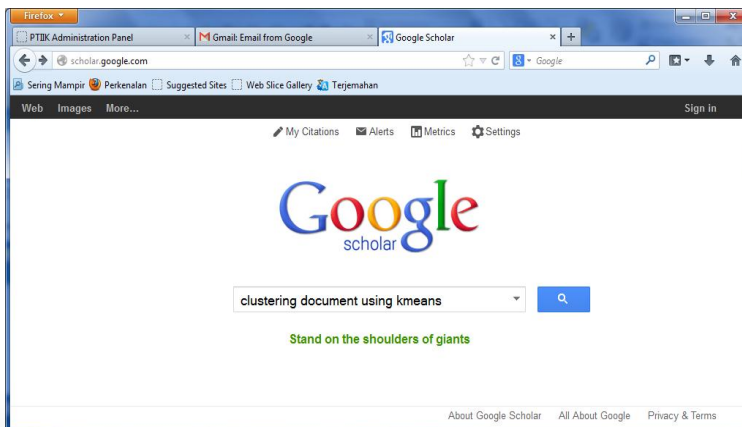
- Computer.org (gunakan proxy UB) :



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



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



- Other

Selesai

