

KOM341 Temu Kembali Informasi

- KULIAH #1
- Kontrak Perkuliahan
 - Pendahuluan

JULIO ADISANTOSO - ILKOM IPB

Matakuliah

- Nama Matakuliah : Temu Kembali Informasi
- Kode Matakuliah : KOM431
- Beban Kredit : 3(3-0)
- Semester : Gasal, 2009/2010
- Koordinator : Julio Adisantoso
- Pengajar : 1. Julio Adisantoso (JAS)
2. Sony Hartono Wijaya (SHW)
3. Yeni Herdiyeni (YHY)

JULIO ADISANTOSO - ILKOM IPB

Manfaat dan Tujuan

- Matakuliah ini akan memberi manfaat bagi mahasiswa dalam menerapkan konsep temu kembali informasi untuk membuat sistem aplikasi temu kembali informasi teks.
- Setelah mengikuti matakuliah ini, mahasiswa diharapkan mampu menjelaskan konsep dalam temu kembali informasi, serta menerapkannya untuk membuat sistem aplikasi temu kembali informasi teks.

JULIO ADISANTOSO - ILKOM IPB

Deskripsi

Matakuliah ini menjelaskan pengantar temu kembali informasi, dasar-dasar temu kembali informasi: pemodelan, evaluasi, query, operasi teks dan multimedia, indexing and searching. Topik dalam temu kembali informasi: relevance feedback, query expansion, text classification, text clustering, summarization, cross-language, question answering, web search.

JULIO ADISANTOSO - ILKOM IPB

Strategi

- Mahasiswa S1 Mayor Ilmu Komputer IPB, sebagai matakuliah pilihan.
- Perkuliahan dilakukan sebanyak 14 kali pertemuan kuliah tatap muka.
- Metode perkuliahan adalah kombinasi antara ceramah, diskusi, dan diakhiri dengan presentasi proyek akhir.
- Mahasiswa WAJIB mengikuti perkuliahan minimal 80 persen, dan presentasi proyek akhir 100 persen.

JULIO ADISANTOSO - ILKOM IPB

Strategi

- Mahasiswa pengulang matakuliah Temu Kembali Informasi WAJIB mengikuti keseluruhan kegiatan kuliah dan presentasi proyek akhir selama satu semester.
- Untuk membantu mahasiswa memahami materi kuliah, disediakan website matakuliah online pada alamat <http://web.ipb.ac.id/~julio/tki0910>.

JULIO ADISANTOSO - ILKOM IPB

Tugas Matakuliah

Tugas terdiri atas dua jenis:

- Perorangan
- Kelompok (dalam bentuk proyek akhir) berupa tugas pemrograman, dan setiap kelompok terdiri atas 2-3 orang. Topik dipilih bebas, tidak ada yang sama di antara kelompok. Produk berupa program komputer, laporan hasil kajian, dan slide presentasi. Presentasi proyek akhir dilakukan di luar jadwal kuliah yang telah ditetapkan.

JULIO ADISANTOSO - ILMKOM IPB

Referensi

- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze . 2008. Introduction to Information Retrieval. Cambridge University Press.
- C. J. van Rijsbergen. Information Retrieval. Information Retrieval Group, University of Glasgow.
- Richardo Baeza-Yates and Berthier Rieiro-Neto. Modern Information Retrieval.
- PERL Programming.
- Henk Blanken, et.al. 2007. Multimedia Retrieval. Text Summarization. Tutorial ACM SIGIR, Sheffield, UK July 25, 2004
- TREC. Question Answering System and Cross Language Information Retrieval.

JULIO ADISANTOSO - ILMKOM IPB

Kriteria Penilaian

- Nilai akhir (NA) = kumulatif dari
 - UTS (1-6) dan UAS (7-14), ujian tertulis dengan bobot masing-masing 35%.
 - Nilai Tugas Perorangan adalah rata-rata dari semua tugas yang diberikan, dan diberi bobot 10%
 - Nilai Proyek Akhir (program komputer, laporan, dan presentasi), dengan bobot 20%.
- Selang nilai untuk menetapkan huruf mutu A, B, C, D, atau E ditentukan berdasarkan nilai rata-rata sebaran normal, berlaku sama untuk semua mahasiswa baru maupun pengulang.

JULIO ADISANTOSO - ILMKOM IPB

Tata Tertib

- Sesuai dengan ketentuan yang terdapat pada Buku Panduan Sarjana IPB
- Hadir paling lambat 15 menit. Mahasiswa TIDAK DIPERKENANKAN masuk kelas setelah 15 menit kuliah dimulai.
- Berpenampilan dan berbusana sopan serta rapi.
- Tidak menggunakan sandal atau sejenisnya.
- Tidak mengoperasikan handphone, laptop, atau sejenisnya.
- Tidak ada ujian dan penugasan susulan atau perbaikan.

JULIO ADISANTOSO - ILMKOM IPB

Jadwal Kuliah

- Kuliah dilaksanakan pada hari Kamis pukul 08:00-10:30 di ruang A000GMKL (Ruang Kuliah GMSK L-H).
- Mohon dicarikan pengganti waktu karena bentrok dengan jadwal JAS (selain waktu di bawah ini)*
 - Senin, 08:00-11:00 dan 13:00-15:00
 - Selasa, 08:00-11:00
 - Kamis, 08:00-12:00

JULIO ADISANTOSO - ILMKOM IPB

PENDAHULUAN

JULIO ADISANTOSO - ILMKOM IPB

What is this course about?

- Processing
- Indexing
- Retrieving
- ... textual data

- Fits in four lines, but much more complex and interesting than that

JULIO ADISANTOSO - ILKOM IPB

Need for IR

- With the advance of WWW - more than 8 Billion documents indexed on Yahoo, Google

- Various needs for information:
 - Search for documents that fall in a given topic
 - Search for a specific information
 - Search an answer to a question
 - Search for information in a different language
 - ...
 - Search for images
 - Search for music
 - Search for a (candidate) friend

JULIO ADISANTOSO - ILKOM IPB

Some definitions of IR

- Salton (1989):** "Information-retrieval systems process files of records and requests for information, and identify and retrieve from the files certain records in response to the information requests. The retrieval of particular records depends on the similarity between the records and the queries, which in turn is measured by comparing the values of certain attributes to records and information requests."

- Information retrieval mempelajari algoritme dan model untuk memperoleh informasi dari koleksi dokumen*

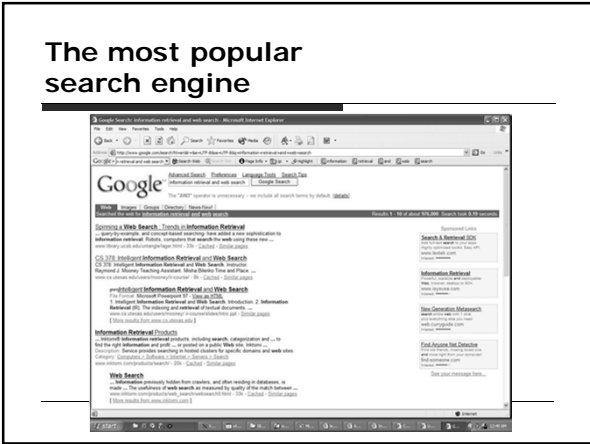
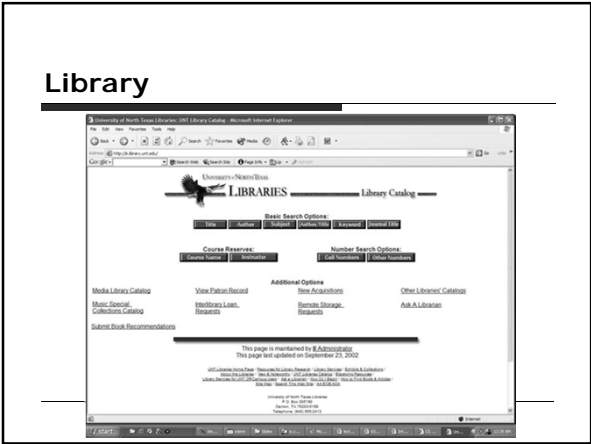
- Information retrieval system : sistem untuk merepresentasikan, menyimpan, mengorganisasikan, dan memproses informasi (Beeza-Yates & Ribeiro-Neto)*

JULIO ADISANTOSO - ILKOM IPB

Examples of IR systems

- Conventional (library catalog)
Search by keyword, title, author, etc. E.g. : You are probably familiar with www.library.unt.edu
- Text-based (Lexis-Nexis, Google, FAST).
Search by keywords. Limited search using queries in natural language.
- Multimedia (QBIC, WebSeek, SaFe)
Search by visual appearance (shapes, colors,...).
- Question answering systems (AskJeeves, Answerbus)
Search in (restricted) natural language
- Other:
cross language information retrieval, music retrieval

JULIO ADISANTOSO - ILKOM IPB



IR systems on the Web

- Search for Web pages <http://www.google.com>
- Search for images <http://www.picsearch.com>
- Search for image content <http://wang.ist.psu.edu/IMAGE/>
- Search for answers to questions <http://www.askjeeves.com>
- Music retrieval <http://www.fxpal.com/people/foote/musicr/>

JULIO ADISANTOSO - ILKOM IPB

IR vs. Data Retrieval

IR

- berkaitan dengan natural language text → unstructured and semantically ambiguous
- spesifikasi set of words untuk menentukan semantics dari information needed

Data Retrieval

- berkaitan dengan data → well defined structure and semantic
- spesifikasi query expression untuk menentukan constrain yang harus dipenuhi untuk obyek yang akan menjadi himpunan jawaban

JULIO ADISANTOSO - ILKOM IPB

IR vs. Databases

- Structured vs unstructured data
- Structured data tends to refer to information in "tables"

Employee	Manager	Salary
Smith	Jones	50000
Chang	Smith	60000
Ivy	Smith	50000

Typically allows numerical range and exact match (for text) queries, e.g., *Salary < 60000 AND Manager = Smith.*

JULIO ADISANTOSO - ILKOM IPB

IR vs. Databases

	Databases	IR
Data	Structured	Unstructured
Fields	Clear semantics (SSN, age)	No fields (other than text)
Queries	Defined (relational algebra, SQL)	Free text ("natural language"), Boolean
Recoverability	Critical (concurrency control, recovery, atomic operations)	Downplayed , though still an issue
Matching	Exact (results are always correct)	Imprecise (need to measure effectiveness)

JULIO ADISANTOSO - ILKOM IPB

IR Principal

- The indexing and retrieval of textual documents.
- Searching for pages on the World Wide Web is the most recent and perhaps most widely used IR application
- Concerned firstly with retrieving relevant documents to a query.
- Concerned secondly with retrieving from large sets of documents efficiently.

➔ *retrieve semua dokumen yang relevan terhadap kueri pengguna & seminimum mungkin retrieve dokumen yang tidak relevan*

JULIO ADISANTOSO - ILKOM IPB

Typical IR Task

Given:

- A corpus of textual natural-language documents.
- A user query in the form of a textual string.

Find:

- A ranked set of documents that are relevant to the query.

JULIO ADISANTOSO - ILKOM IPB

