

Sistem Temu Kembali Informasi

“Pengenalalan Sistem Temu Kembali Informasi (STKI)”



Tim Dosen STKI



Profil Dosen Pengampu

Abu Salam, M.Kom

Research Interest

- Software Engineering, Data Mining, Information Retrieval, AI dan BIG DATA.

Activity

- Dosen TI Fasilkom UDINUS
- Manager Departement Software PT DINUSTEK



Kontak Dosen Pengampu

Email :

abu.salam@dsn.dinus.ac.id

Telegram/WA:

0817244958



Kontrak Perkuliahan

- Wajib gabung pada Grup Telegram (Cek SIADIN / KULINO)
- STKI dilaksanakan 14 Kali Pertemuan
- Materi Perkuliahan disediakan tiap pertemuan melalui media kulino.dinus.ac.id
- Kehadiran **minimal 75% (Tidak ada Dispensasi)**
- Prosentase Nilai : Tugas : 40, UTS: 30, UAS: 30
- Nilai Tugas adalah akumulasi nilai (Kehadiran, Sikap, Keaktifan, Kuis, Tugas dan lain-lain)



Perkenalan Makul

- ☉ Nama Mata Kuliah : Information Retrieval (STKI)
- ☉ Jumlah SKS : 3 SKS Teori
- ☉ Tim Dosen :
 - Dr. Catur Supriyanto, M.Cs
 - **Abu Salam M.Kom**
 - Junta Zeniarja M.Kom
 - Ardytha Luthfiarta M.Kom



Outline

- Pengenalan sistem temu-kembali informasi
- Dokumen Preprocessing
- Model sistem temu kembali informasi; Boolean model
- Model sistem temu kembali informasi; vector space model
- Skema pembobotan; tf, log tf, idf, tf.idf
- Search Engine Concept
- Evaluasi dalam STKI
- Klasifikasi Dokumen; naive bayes
- Klasifikasi Dokumen; KNN
- Kuster Dokumen; K-Means
- Peringkasan Dokumen Teks
- Feature Selection
- Opinion Mining (Sentiment Analysis)
- Web Mining



Buku Penunjang & Literatur





Dokumen



Penyimpanan Yang Terorganisasi



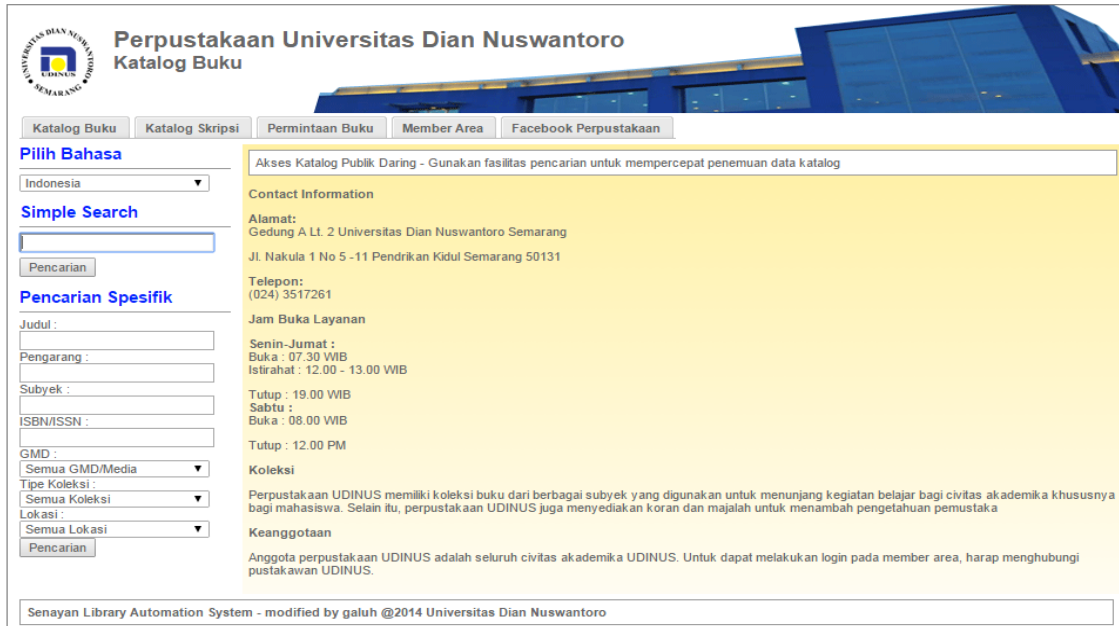


Database

Database Buku

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

Perpustakaan Udinus (E- Library)



The screenshot displays the UDINUS library website. At the top, the header includes the university logo, the name 'Perpustakaan Universitas Dian Nuswantoro', and 'Katalog Buku'. Navigation tabs for 'Katalog Buku', 'Katalog Skripsi', 'Permintaan Buku', 'Member Area', and 'Facebook Perpustakaan' are visible. The left sidebar contains a 'Pilih Bahasa' dropdown set to 'Indonesia', a 'Simple Search' section with a search bar and 'Pencarian' button, and a 'Pencarian Spesifik' section with dropdowns for 'Judul', 'Pengarang', 'Subyek', 'ISBN/ISSN', 'GMD' (set to 'Semua GMD/Media'), 'Tipe Koleksi' (set to 'Semua Koleksi'), and 'Lokasi' (set to 'Semua Lokasi'), each followed by a 'Pencarian' button. The main content area features a yellow background with a header 'Akses Katalog Publik Daring - Gunakan fasilitas pencarian untuk mempercepat penemuan data katalog'. It includes 'Contact Information' with address, phone, and hours, and a 'Koleksi' section describing the book collection and membership. A footer note mentions the 'Senayan Library Automation System'.

Perpustakaan Universitas Dian Nuswantoro
Katalog Buku

Katalog Buku Katalog Skripsi Permintaan Buku Member Area Facebook Perpustakaan

Pilih Bahasa
Indonesia

Simple Search
Pencarian

Pencarian Spesifik
Judul :
Pengarang :
Subyek :
ISBN/ISSN :
GMD :
Semua GMD/Media
Tipe Koleksi :
Semua Koleksi
Lokasi :
Semua Lokasi
Pencarian

Akses Katalog Publik Daring - Gunakan fasilitas pencarian untuk mempercepat penemuan data katalog

Contact Information
Alamat:
Gedung A Lt. 2 Universitas Dian Nuswantoro Semarang
Jl. Nakula 1 No 5 -11 Pendrikan Kidul Semarang 50131
Telepon:
(024) 3517261
Jam Buka Layanan
Senin-Jumat :
Buka : 07.30 WIB
Istirahat : 12.00 - 13.00 WIB
Tutup : 19.00 WIB
Sabtu :
Buka : 08.00 WIB
Tutup : 12.00 PM
Koleksi
Perpustakaan UDINUS memiliki koleksi buku dari berbagai subyek yang digunakan untuk menunjang kegiatan belajar bagi civitas akademika khususnya bagi mahasiswa. Selain itu, perpustakaan UDINUS juga menyediakan koran dan majalah untuk menambah pengetahuan pemustaka
Keanggotaan
Anggota perpustakaan UDINUS adalah seluruh civitas akademika UDINUS. Untuk dapat melakukan login pada member area, harap menghubungi pustakawan UDINUS.

Senayan Library Automation System - modified by galuh @2014 Universitas Dian Nuswantoro



Database Retrieval vs Sistem Temu Kembali Informasi

- Database Retrieval
 - Dokumen mana yang mengandung himpunan keyword?
 - Semantik didefinisikan dengan baik
 - Error dari suatu obyek mengakibatkan kegagalan!
- Sistem Temu Kembali Informasi
 - Informasi mengenai suatu subyek atau topik
 - Semantik dapat bersifat lepas (longgar)
 - Error kecil ditoleransi



Pengertian

- 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.



Sistem Temu Kembali Informasi

- 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.

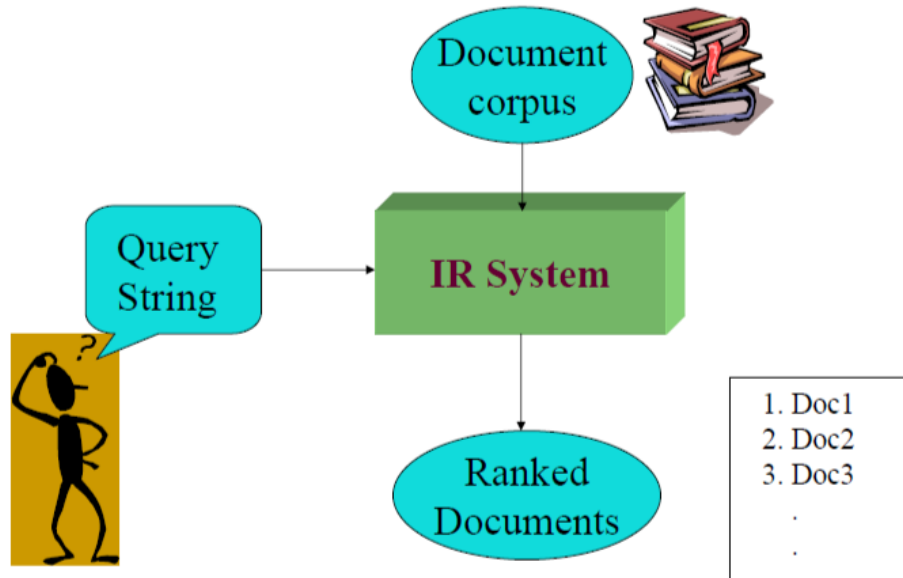


Relevan

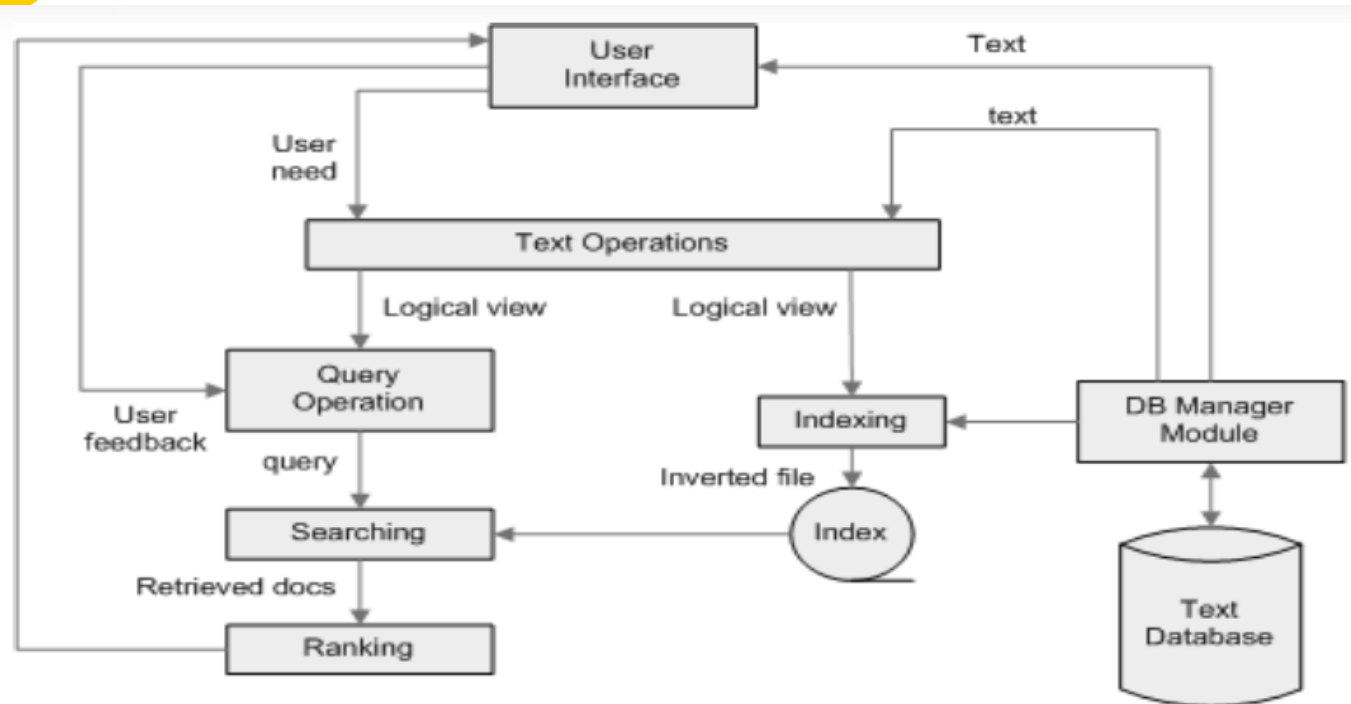
- Relevansi merupakan suatu *judgment* (keputusan) subyektif dan dapat didasarkan pada:
 - Topik yang tepat.
 - Waktu (informasi terbaru).
 - Otoritatif (dari suatu sumber terpercaya).
 - Kebutuhan informasi dari pengguna.
- Kriteria relevansi utama: suatu system IR sebaiknya (harus) memenuhi kebutuhan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



Sistem Sistem Temu Kembali Informasi (IR System)



Sistem Temu Kembali Informasi





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 system IR yang efisien, *tetapi...*
 - Fokus pada algoritma dan kemampuan dasar, bukan masalah sistem yang memungkinkan pengembangan ke database ukuran industri.



STKI Cerdas

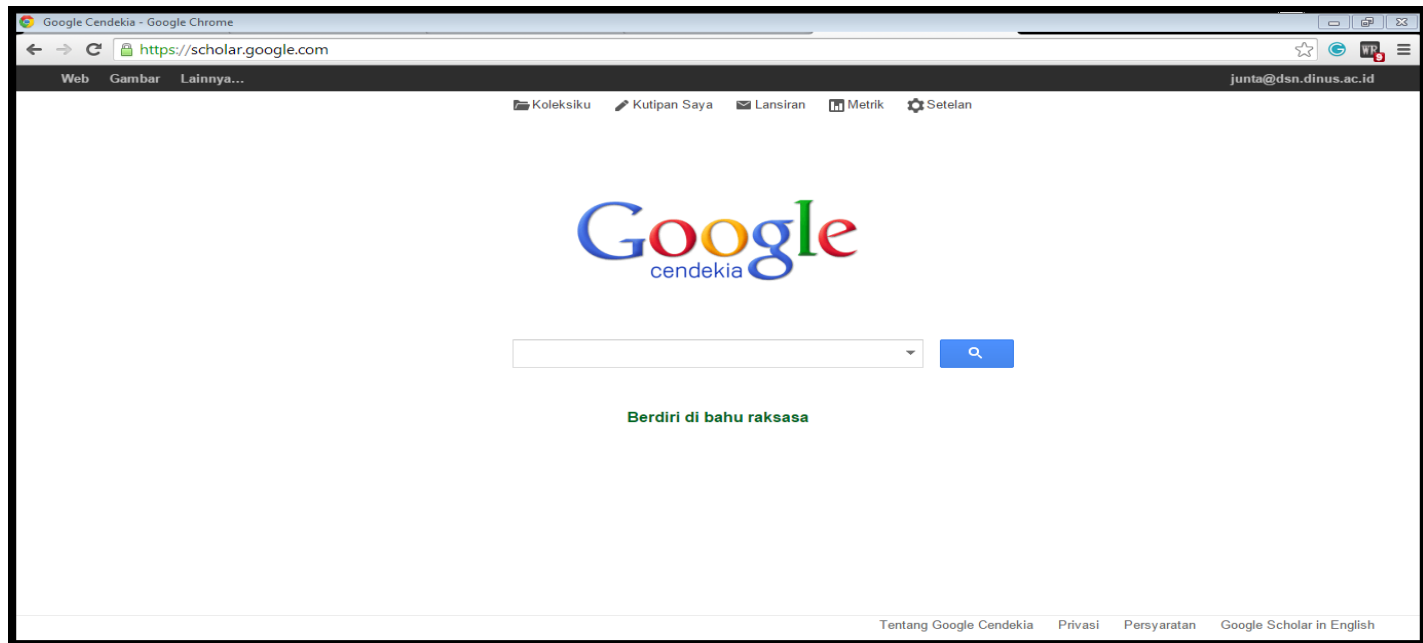
- Memanfaatkan pengertian atau makna dari kata yang digunakan.
- Melibatkan urutan kata di dalam query.
- Beradaptasi dengan pengguna berdasarkan pada *feedback, langsung atau tidak langsung*.
- Memperluas pencarian dengan term terkait.
- Mengerjakan pemeriksaan ejaan / perbaikan tanda pengenal otomatis.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



Perkembangan Sistem Temu Kembali Informasi

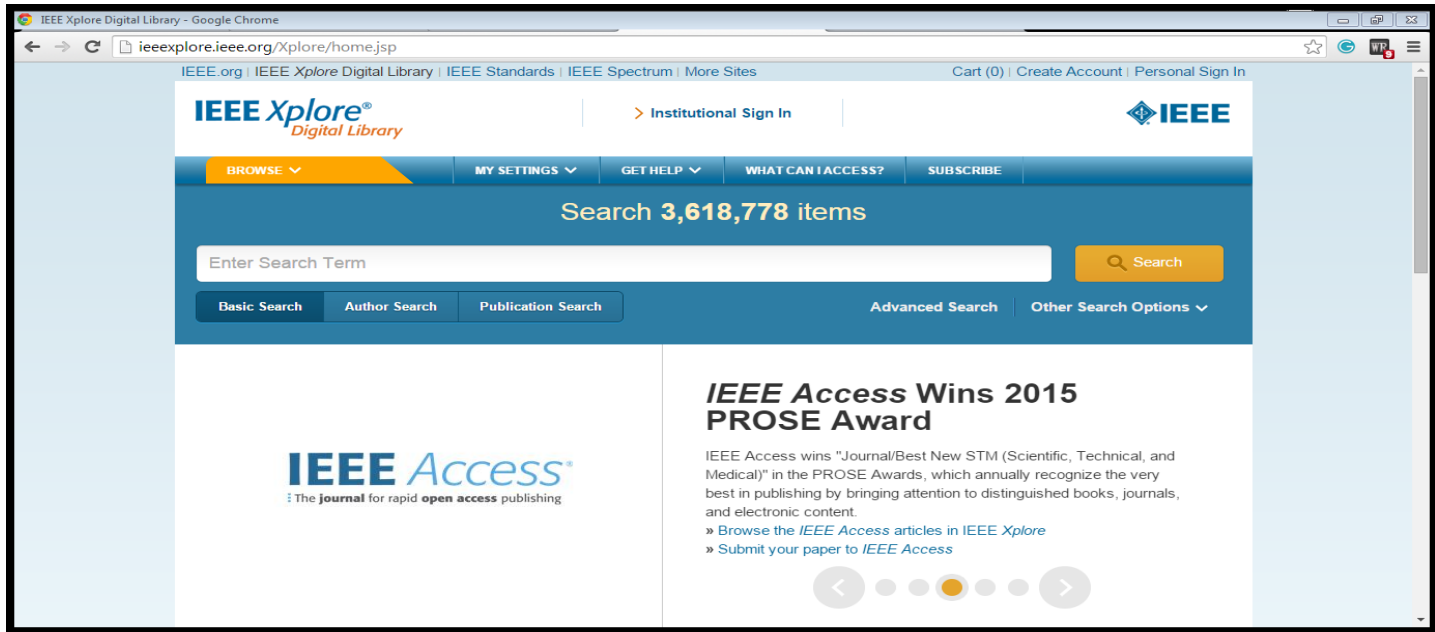
1. Klasifikasi Dokumen
2. Clustering Dokumen
3. Peringkasan Teks
4. Question Answering System

Portal Jurnal (<http://scholar.google.com/>)



Portal Jurnal

(<http://ieeexplore.ieee.org/Xplore/home.jsp>)





Portal Jurnal

(<http://www.sciencedirect.com>)

The screenshot shows the ScienceDirect website in a Google Chrome browser. The page features a green header with the ScienceDirect logo and navigation links for Journals, Books, Shopping cart, Sign in, and Help. A search bar is prominently displayed. Below the header, there's a section for 'ScienceDirect' with a brief description and a link to 'Browse publications by subject'. This section is divided into four columns: Physical Sciences and Engineering, Life Sciences, Health Sciences, and Social Sciences and Humanities. Each column lists various sub-fields. To the right, there's an 'Elsevier's Language Editing service' advertisement. At the bottom, there's a section for 'Open Access' and 'Latest articles'.

ScienceDirect.com | Science, health and medical journals, full text articles and books. - Google Chrome

www.sciencedirect.com

ADVERTISEMENT

The app covers 21,000 journals from All publishers

Research Highlights

ScienceDirect Journals Books Shopping cart Sign in Help

Search all fields Author name Journal or book title Volume Issue Page Advanced search

You have Guest access to ScienceDirect Find out more...

ScienceDirect is a leading full-text scientific database offering journal articles and book chapters from nearly 2,500 journals and 26,000 books.

13,363,844 articles on ScienceDirect

ADVERTISEMENT

ELSEVIER WebShop

Elsevier's Language Editing service

Make sure your manuscript is written in perfect English before submission

Learn more

Open Access ScienceDirect's Top 25 Latest articles

Articles published in our Open Access journals are made permanently free for everyone to access immediately upon

View the top downloaded

* Integrating groundwater into land planning: A risk assessment methodology Journal of Environmental Management



Thanks!

Any questions ?