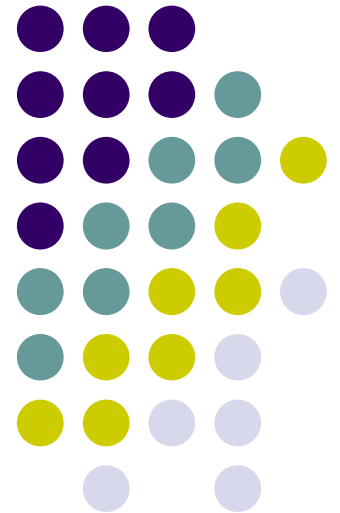
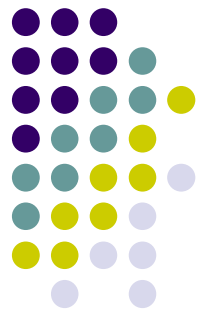


INFORMATION RETRIEVAL (IR)

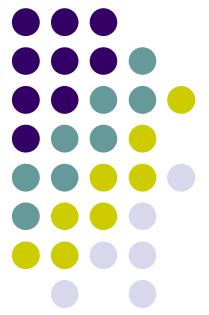
Arif Basofi-Rengga Asmara
©2006





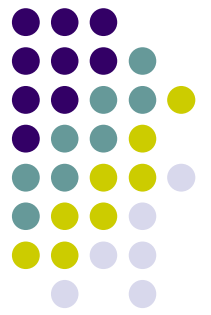
Overview

- **Information Retrieval**
(pencarian/pencapaian informasi)
berhubungan dengan:
 - representasi (gambaran)
 - storage (penyimpanan)
 - organisasi, dan
 - akses ke informasi.



Overview – con't

- **Information Retrieval:**
 - Fokus pada kebutuhan informasi user.
 - Menekankan pada retrieval (pencarian) informasi.
- Dalam representasi dan pengorganisasian, informasi seharusnya memberikan kemudahan bagi user dalam mengakses informasi.

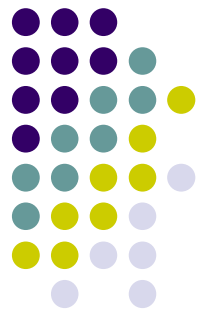


Problem Information Retrieval

- **Problem:**
 - mengetahui karakteristik kebutuhan informasi user.
 - How?
 - Like what?

Problem Information Retrieval

– con't



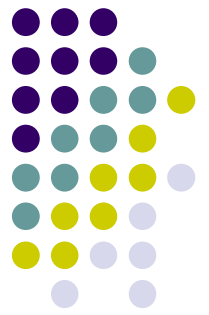
- **Contoh:** dalam WWW, user melakukan searching informasi tentang tim robot kampus ITS, dimana:
 - Dibiayai oleh PENS-ITS.
 - Ikut serta dalam turnamen internasional kontes robot di Jepang.

- **Search:**

“tim robot ITS yang dibiayai PENS-ITS dan pernah ikut turnamen internasional di Jepang”

Problem Information Retrieval

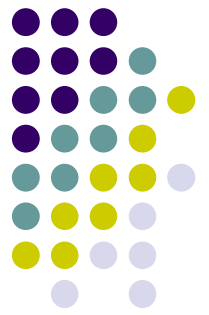
– con't



- Deskripsi kebutuhan informasi user demikian **tidak dapat** langsung didapatkan meski melalui web search engine.
- Sebaliknya, user harus **menterjemahkan terlebih dahulu** kebutuhan informasi yang diharapkan melalui suatu **query** yang kemudian akan diproses oleh search engine (**IR System**).

Problem Information Retrieval

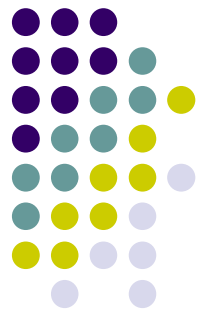
– con't



- Query : pencarian melalui pertanyaan
- Retrieve : pencarian
- Hasil dari penerjemahan oleh user tadi berupa keyword (index) atau kata kunci yang merupakan ringkasan dari deskripsi tadi.

Problem Information Retrieval

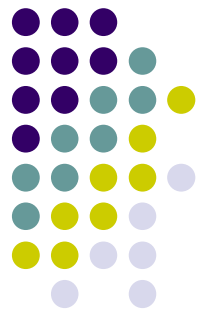
– con't



- **Sehingga dari keyword yang dimasukkan user untuk pencarian tadi, IR system akan retrieve (mendapatkan/mencari) informasi, yang mana informasi yang didapatkan mengandung relevansi / keterkaitan dengan yang diharapkan user tadi.**

Problem Information Retrieval

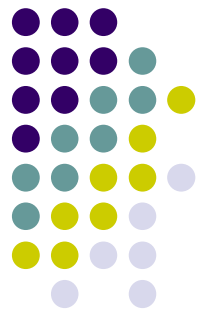
– con't



- Jadi, misal:
- **Search:**

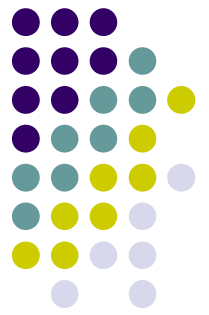
**“turnamen internasional robot PENS-ITS”
atau “ tim robot PENS-ITS”**

Information Retrieval vs Data Retrieval



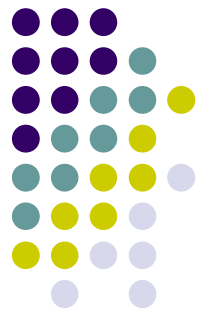
- Information Retrieval berlawanan dengan Data Retrieval.
- **Data Retrieval** (dalam konteks sistem IR) yaitu umumnya dalam menentukan dokumen yang **tepat** dari suatu koleksi data, yang isi dokumen tersebut mengandung **keyword** di dalam query user.
- Umumnya sering digunakan dan tidak akan pernah cukup untuk memenuhi kebutuhan informasi user.

Information Retrieval vs Data Retrieval – con't



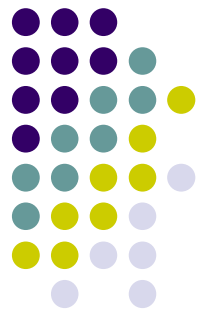
- **Kenyataannya:** user dari sistem IR lebih memperhatikan dalam mendapatkan (retrieve) informasi melalui **subyek**, daripada retrieve data berdasarkan **query** yang diberikan.
- Karena user tidak mau tahu bagaimana proses yang sedang berlangsung.

Information Retrieval vs Data Retrieval – con't

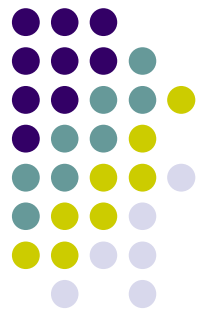


- Dalam bahasa **Data Retrieval**, bertujuan untuk mendapatkan semua objek berdasarkan kondisi yang telah ditentukan/definisikan, seperti melalui **Regular Expression (RE)** (dalam automata) atau melalui **ekspresi aljabar relational** (dalam database).

Information Retrieval vs Data Retrieval – con't



- Sehingga, dalam **Data Retrieval**, suatu kesalahan tunggal dalam retrieve / mendapat sebuah object dari ribuan object yang ter-retrieve sudah merupakan **kesalahan fatal**.
- Sedangkan, dalam **Information Retrieval**, object-object yang ter-retrieve ada kemungkinan tidak akurat dan kesalahan-kesalahan kecil yang biasanya tidak diketahui.

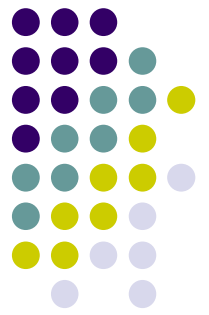


Summary

Jadi, dapat disimpulkan bahwa dalam :

- **Information Retrieval**

- Berhubungan dengan **text** bahasa umum yang tidak selalu terstruktur dan ada kemungkinan memiliki kerancuan arti.
- Informasi mengenai **subyek** atau **topik**.
- **Semantik** sering kali hilang
- Kesalahan kecil masih bisa ditoleransi

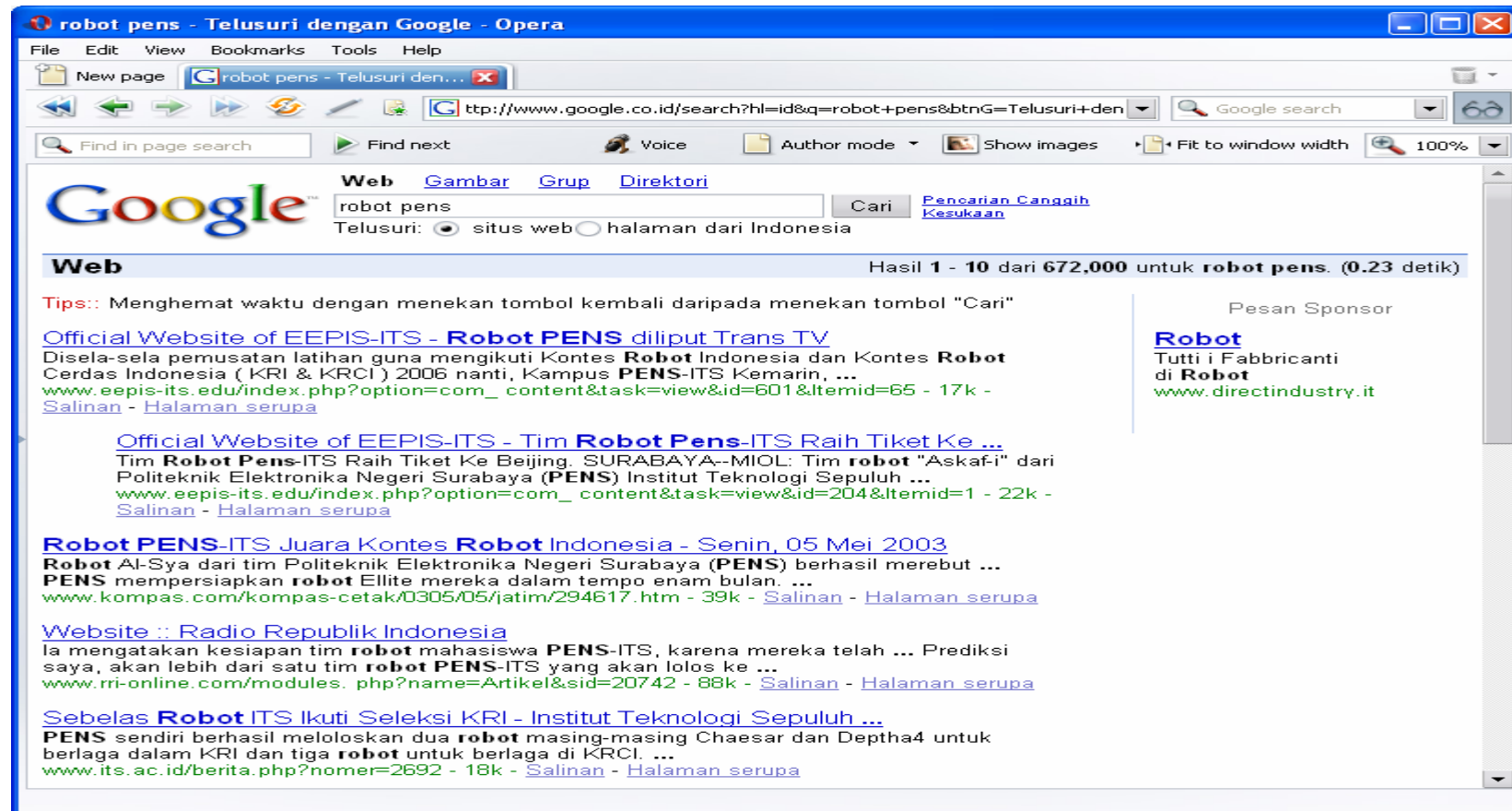
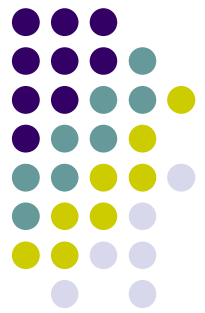


Summary – con't

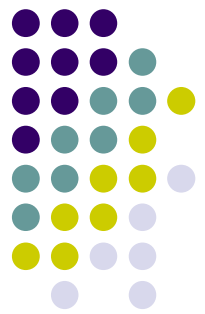
- **Data Retrieval**

- Seperti dalam **database relasional**: berhubungan dengan **data**, yang mana semantik (arti kata) dan strukturnya sudah terdefinisikan.
- Isi dokumen/data mengandung bagian dari **keyword** (kata kunci).
- Semantik terdefinisi dengan baik.
- Kesalahan kecil/tunggal dari suatu obyek menunjukkan kegagalan.

Bentuk contoh Information Retrieval



Contoh Searching Text melalui Web Search Engine



Bentuk contoh Data Retrieval

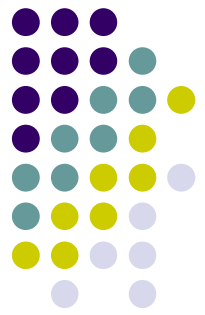
The screenshot shows a database application window titled "Addresses". The main window has a tabbed interface with "General Information", "Members", and "Notes". The "General Information" tab is active, showing fields for "First Name", "Last Name", "Nickname", "Relationship", "Birth Date", "E-mail Address", "Last Contact", "Updated", and "Send Card". The "Company/Family" field is set to "Davolio".

Overlaid on the main window is a smaller window titled "Find Matching Records". This window has a "Find by:" section with four radio buttons: "Company/Family", "First name", "Last name", and "Nickname". The "First name" radio button is selected. To the right of this section is a "Select Name:" dropdown menu with "Pau" selected. At the bottom right of the "Find Matching Records" window are "OK" and "Cancel" buttons.

At the bottom of the "Addresses" window, there are buttons for "Show All", "Datasheet View", and "Close". The status bar at the very bottom indicates "Record: 14 of 1".

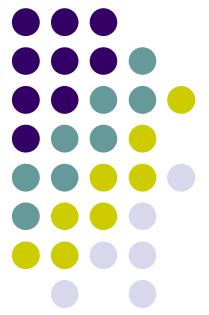
Contoh Data Retrieval dalam Database

Bentuk contoh Data Retrieval – con't

A screenshot of a SQL query window titled "AddressList : Select Query". The window has a dark red title bar with a small icon on the left and standard minimize, maximize, and close buttons on the right. The main area is a white text box containing a SQL query. The query selects various fields from two tables, Household and Members, using a LEFT JOIN. The fields selected are HouseholdName, FirstName, LastName, Address, City, StateOrProvince, PostalCode, Country, and SendCard. The join condition is on AddressID.

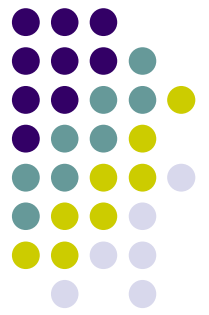
```
SELECT Household.HouseholdName, Members.FirstName,  
Members.LastName, Household.Address, Household.City,  
Household.StateOrProvince, Household.PostalCode, Household.Country,  
Members.SendCard  
FROM Household LEFT JOIN Members ON Household.AddressID =  
Members.AddressID;  
|
```

Contoh SQL (Structured Query Language) dalam Database Relasional



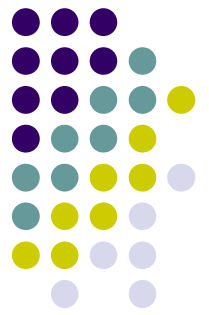
Konsep Dasar IR

- Dalam sistem IR, kebutuhan item-item informasi user (dokumen) harus di terjemahkan terlebih dahulu dalam suatu koleksi/kumpulan keyword dan kemudian dilakukan urutan ranking relevansi dari query user.

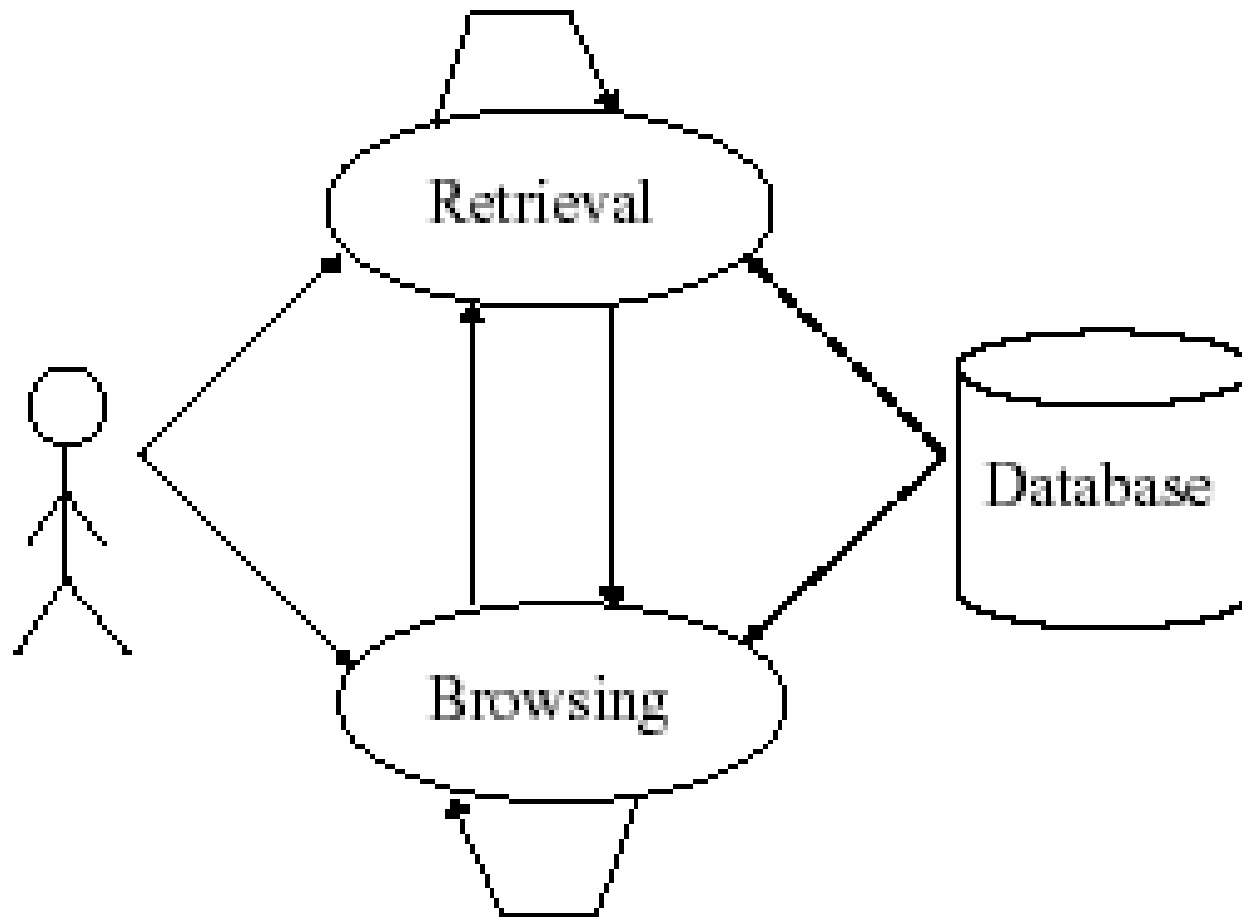


Konsep Dasar IR – con't

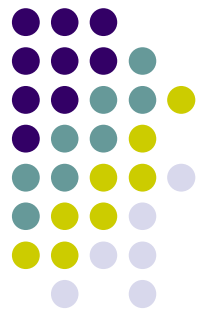
- Dalam sistem retrieval efektif, terjalin hubungan antara ***user task*** (tugas user) dengan ***logical view*** dari dokumen seperti pada gambar berikut.



Konsep Dasar IR – con't

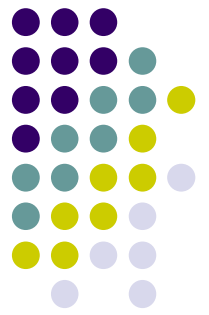


Interaksi user dengan sistem retrieval melalui tugas yang berbeda



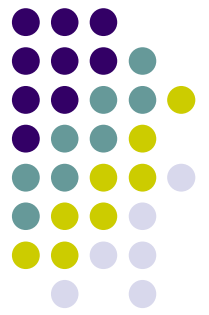
Konsep Dasar IR – con't

- Gambar diatas menggambarkan interaksi **user** dengan dua task (tugas) yang berbeda,
 - yang pertama user melakukan **browsing** dokumen informasi yang tersedia, dan
 - yang kedua user melakukan **retrieval** informasi dan data (searching) yang ada di dalam database (storage).



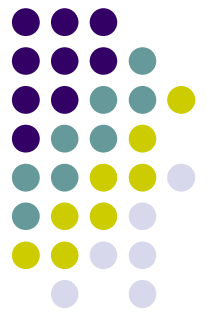
Summary - Konsep Dasar IR

- **Retrieval:**
 - Informasi atau data
 - Dengan maksud tertentu
- **Browsing:**
 - Melihat-lihat (sementara)
 - Misal. Browse ttg F1: mobil balap, trancis, turis, aksesoris, dll.



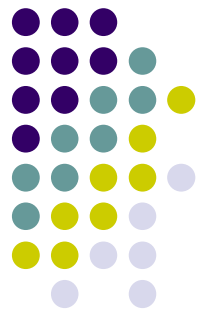
User Task

- User harus menterjemahkan kebutuhan informasinya kedalam bentuk bahasa query yang disediakan oleh sistem.
- Dengan sistem **Information Retrieval**, secara normal akan menentukan serangkaian kata-kata yang memberikan arti/maksud dari kebutuhan informasi user.



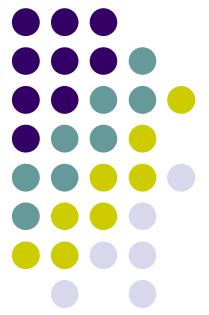
User Task – con't

- Dengan sistem **Data Retrieval**, ekspresi query (RE) digunakan dalam menyampaikan objek-objek yang saling berhubungan.
- Contoh:
 - user mencari informasi mengenai balap mobil secara umum, namun dengan User Interface yang tersedia, user akan melihat-lihat tentang informasi balap mobil formula 1 melalui link informasi yang ada.



User Task – con't

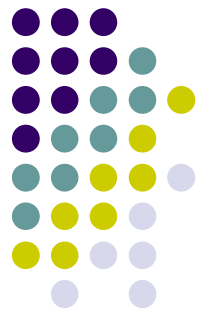
- Proses yang dilakukan user diatas adalah browsing melalui kumpulan dokumen yang ber-relasi, bukan melakukan searching.
- Tapi masih dalam proses information retrieve, hanya object atau sasaran yang dicari tidak di definisikan terlebih dahulu di awal permulaan bahkan kemungkinan adanya perubahan tujuan yang dilakukan user.



Logical View Pada Dokumen

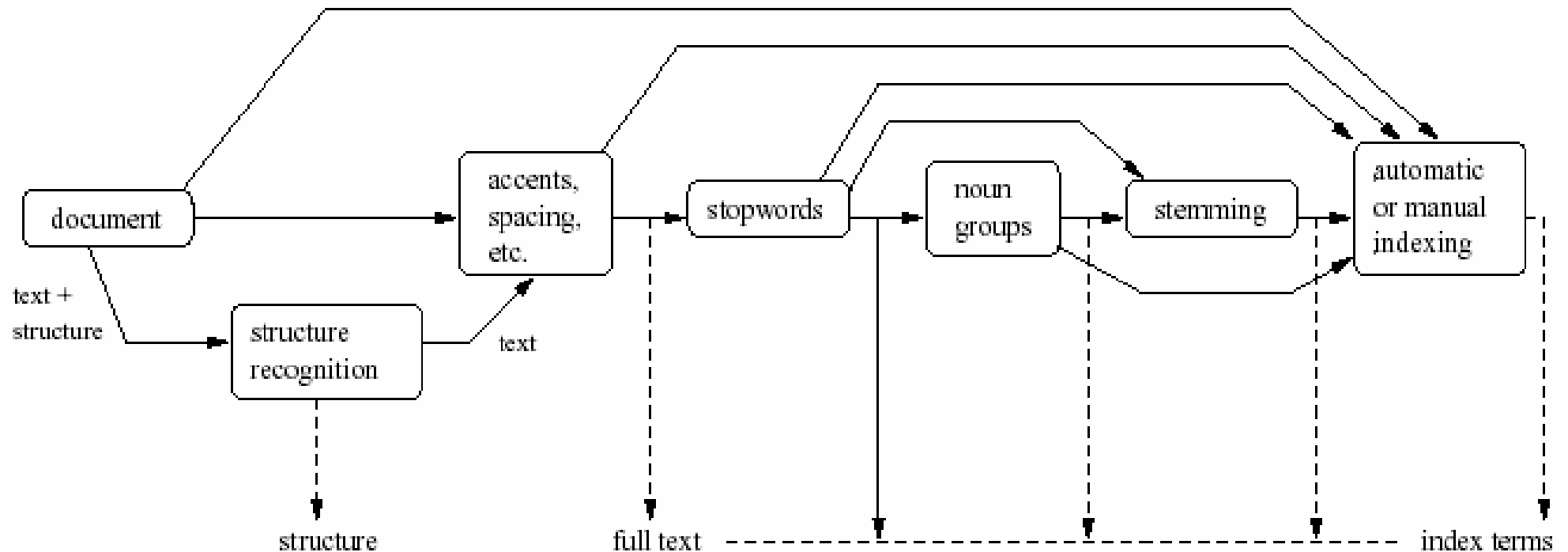
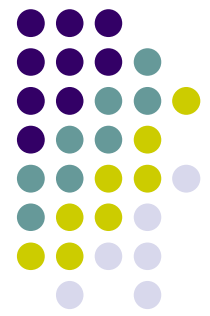
- Dokumen yang ditampilkan dari database, dapat melalui **index** atau **keyword**.
- Dalam menentukan index maupun keyword dari suatu dokumen dapat di ekstrak secara langsung dari text dalam dokumen, atau menspesifikasikan subject-subject dalam dokumen oleh user.

Logical View Pada Dokumen – con't



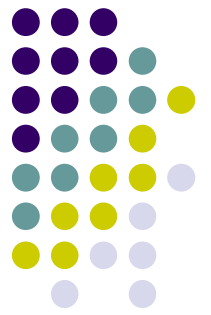
- Baik ditampilkan secara langsung yang didapatkan secara otomatis atau digenerate oleh spesialis, yang mana telah menyediakan suatu logical view dari dokumen.
- Logical view dokumen disini merupakan gambaran proses dalam membangun suatu rangkaian text dari dokumen.

Logical View Pada Dokumen – con't



Logical view suatu dokumen, dari full text hingga bentuk index

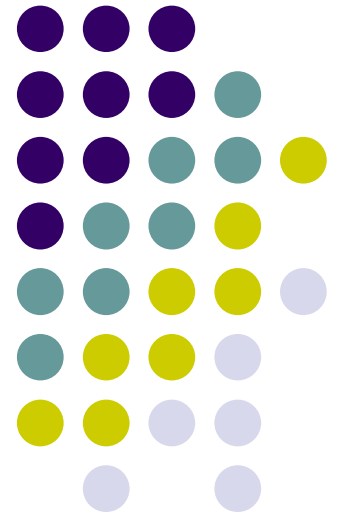
Logical View Pada Dokumen – con't

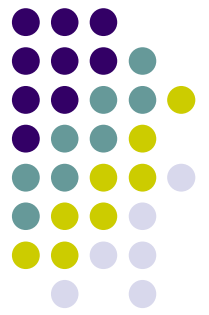


- Gambar diatas menjelaskan bagaimana suatu dokumen dengan struktur full text dapat menghasilkan rangkaian kata-kata (keyword, index) yang nantinya akan di gunakan dalam proses pencarian/retrieve.

Question ?

End of Session





Tugas

- Carilah 5 contoh bentuk **Information Retrieval**
- Carilah 5 contoh bentuk **Data Retrieval**
- **Capture tugas anda dalam sebuah dokumen laporan dan kumpulkan minggu depan.**
- **Cantumkan source alamat tiap contoh (web, pembuat, instansi/organisasi).**