

Temu-Kembali Informasi 2019

02A: Arsitektur Sistem Temu-Kembali Informasi

Husni

Husni.trunojoyo.ac.id

Outline

- Arsitektur Search Engine

Rekap: Mengapa *Information Retrieval*

- *Information overload*
 - “Ini merujuk pada kesulitan seseorang dapat memahami permasalahan dan membuat keputusan yang dapat disebabkan oleh adanya terlalu banyak informasi.” - wiki



Rekap: IR v.s. DB

- Temu-Kembali Informasi (IR):

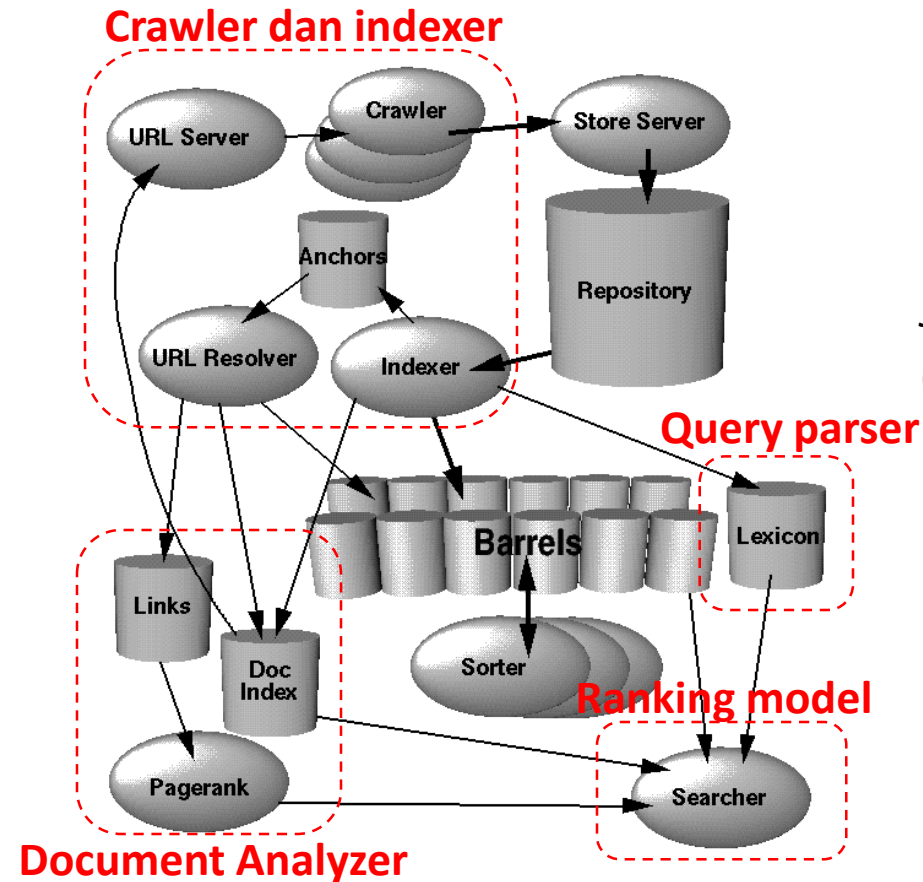
- Data Tak-terstruktur
- Semantik dari obyek bersifat subyektif
- Query berbentuk keyword sederhana
- Temu-Kembali Digerakkan Relevansi
- Efektivitas adalah masalah utama, meskipun efisiensi juga penting.

- Sistem Basis Data (DB):

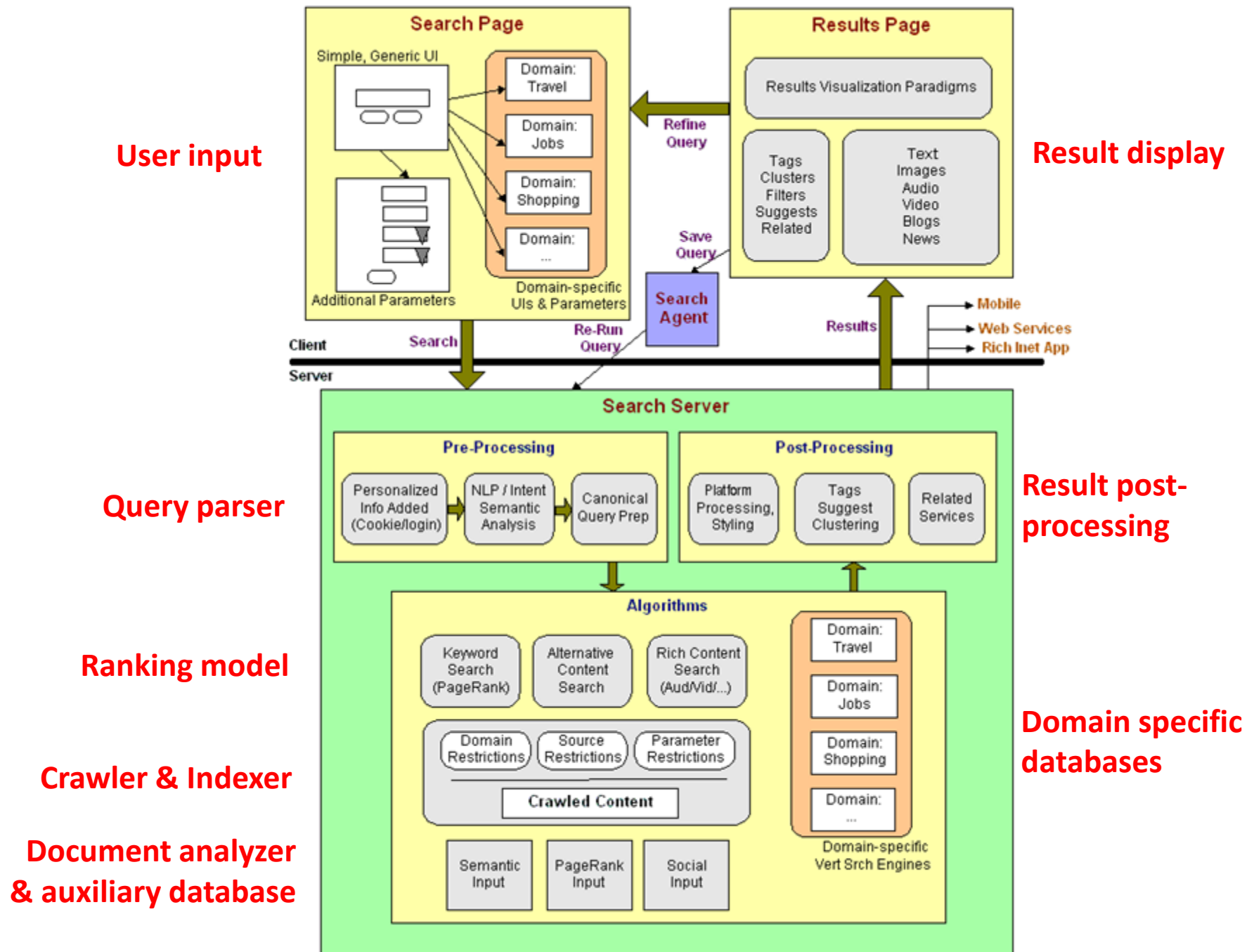
- Data Terstruktur
- Semantik setiap obyek terdefinisi baik
- Bahasa Query Terstruktur (mis. SQL)
- Temu-Kembali pasti
- Penekanan pada efisiensi.

Arsitektur Search Engine Klasik

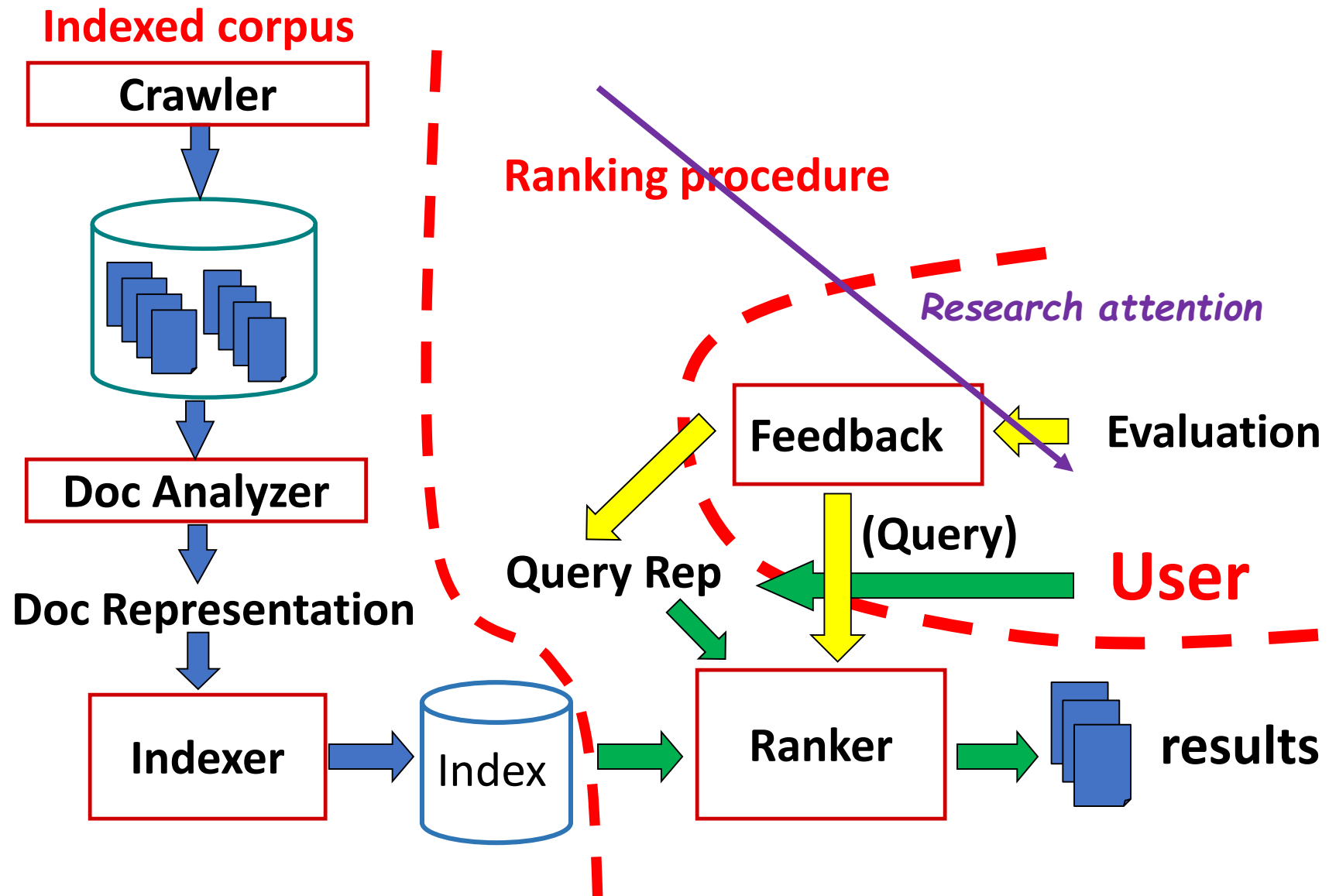
- *“The Anatomy of a Large-Scale Hypertextual Web Search Engine” - Sergey Brin and Lawrence Page, Computer networks and ISDN systems 30.1 (1998): 107-117.*



Jumlah sitasi: 17,463
(pada September 6, 2017)



Abstraksi dari Arsitektur Search Engine



Konsep IR Inti

- Kebutuhan Informasi (*information need*)
 - “*keinginan individu atau kelompok untuk menemukan dan memperoleh informasi untuk memenuhi kebutuhan sadar atau tidak sadar*” – wiki
 - Suatu sistem IR adalah untuk memenuhi kebutuhan informasi pengguna
- Permintaan (Query)
 - Representasi (*designed*) dari kebutuhan informasi pengguna
 - Dalam bahasa alami, atau **suatu form yang dikelola.**

Konsep IR Inti

- Dokumen (*document*)
 - Suatu representasi informasi yang berpotensi memenuhi/memuaskan kebutuhan informasi pengguna
 - Text, image
 - Relevansi (*relevance*)
 - Keterkaitan pengguna
 - Banyak pengguna
- Satu kalimat tentang IR - “ranking-kan dokumen-dokumen berdasarkan relevansinya terhadap kebutuhan informasi pengguna”**

Komponen Kunci dalam *Search Engine*

- Web crawler
 - Suatu program otomatis yang secara sistematis menjelajah web untuk tujuan pengindeksan (*indexing*) dan updating konten Web
- Document analyzer & indexer
 - Mengelola konten web yang telah dicrawl dan menyediakan akses yang efisien ke dokumen-dokumen web
- Web crawler mengambil dokumen dari Internet, Document Analyzer melakukan preprocessing (pembersihan teks), indexer membangun *inverted index*.

Komponen Kunci dalam *Search Engine*

- Query parser
 - Menyusun query kata kunci masukan (input) pengguna ke dalam representasi sistem terkelola (*managed*)
- Ranking model
 - Mengurutkan dokumen-dokumen kandidat sesuai dengan relevansinya terhadap query yang diberikan
- Result display
 - Menyajikan hasil yang diambil (retrieve) kepada pengguna untuk memenuhi kebutuhan informasi mereka.
- Query parser melakukan preprocessing terhadap Query, Ranking model menentukan peringkat dokumen terpilih terhadap Query, dan Result Display menampilkan daftar dokumen yang telah diranking tersebut.

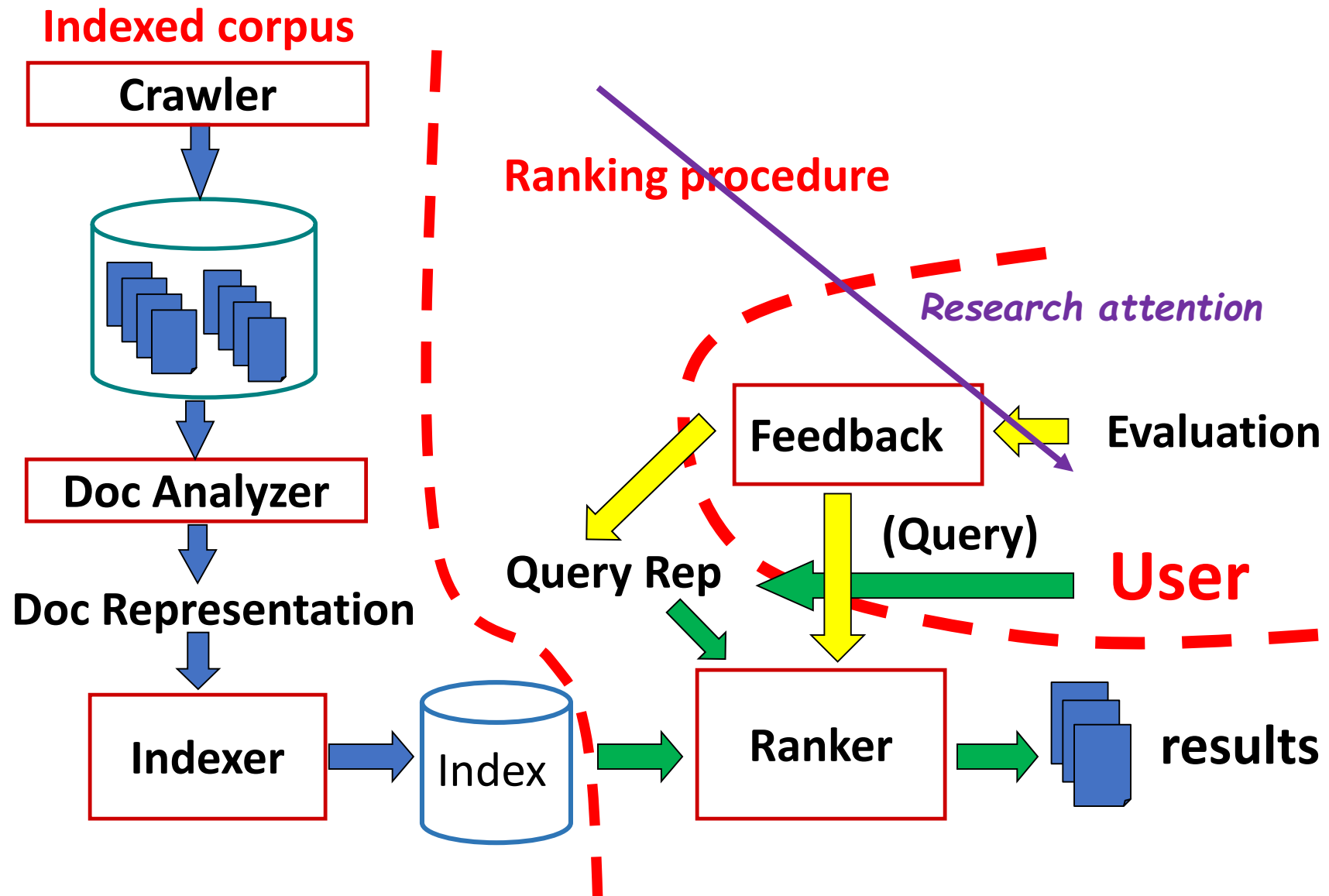
Komponen Kunci dalam *Search Engine*

- Retrieval evaluation
 - Menilai kualitas dari hasil yang dikembalikan
- Relevance feedback
 - Mengirimkan penilaian (penghakiman) kualitas kembali ke sistem untuk memperbaiki (menyaring) hasil pencarian.
- Evaluasi digunakan untuk menilai kualitas dokumen yang diberikan oleh sistem kepada pengguna, Feedback merupakan jalur pengguna untuk menyarankan kualitas bagi sistem IR.

Komponen Kunci dalam Search Engine

- Search query logs
 - Merekam sejarah interaksi pengguna dengan *search engine*
- User modeling
 - Memahami kebutuhan informasi tepat dari pengguna
 - Menilai kepuasan pengguna terhadap hasil mesin pencarian.

Rekap: Abstraksi Arsitektur Search Engine



Diskusi: Browsing v.s. Querying

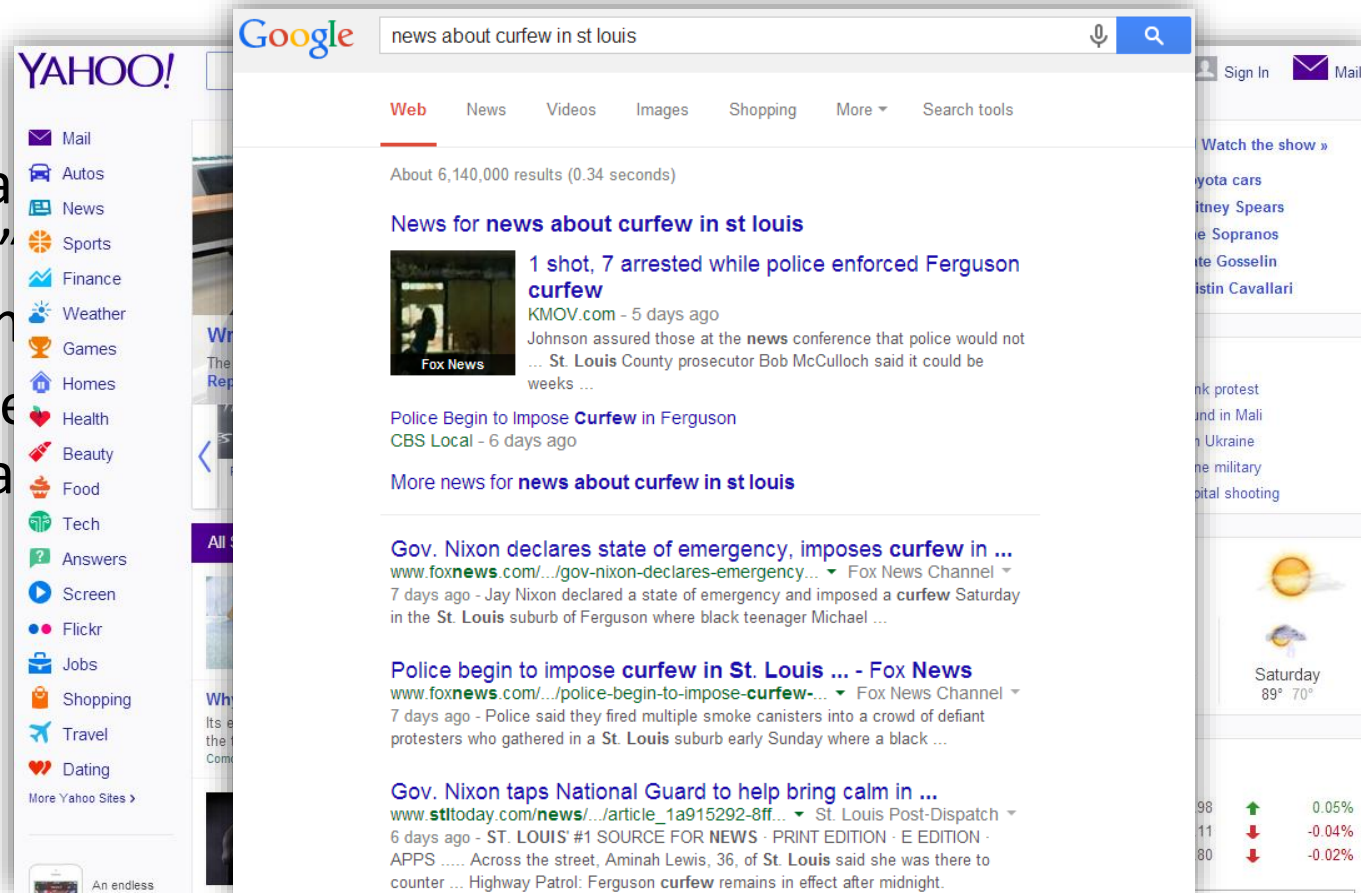
- Browsing – Terjadi di Yahoo dahulu
 - Sistem mengatur informasi dengan struktur, dan pengguna menavigasi ke informasi yang mengikuti jalur struktur itu
 - Berfungsi dengan pengguna ingin atau tidak tahu digunakan, atau mudah memasukan dengan smartp
- Querying – Apa yang dilakukan Google
 - Pengguna memasukkan Query (kata mengembalikan men yang relevan baik ketika rsis Query apa yang mengekspresikan asinya



Pull vs. Push dalam Temu-Kembali Informasi

- Mode Tarik

- Pengguna “menarik” informasi dari sistem
- Bekerja dengan permintaan pengguna informasi



a Query

inisiatif dan
masi yang
ngguna

baik ketika
kebutuhan
oil atau sistem
uan yang baik
pengguna.

Diskusi: Apakah Yelp Suatu Search Engine?

The screenshot displays the Yelp homepage for Charlottesville, VA. The search bar at the top shows 'Find pizza' and 'Near Charlottesville, VA, United States'. Below the search bar, there are navigation links for Restaurants, Nightlife, Home Services, Write a Review, Events, and Talk. The main heading is 'Best pizza in Charlottesville, VA', with a note 'Showing 1-10 of 96'. There are filters for price (\$ to \$\$\$\$) and 'Open Now'. The results list includes:

- Ad Marco's Pizza**: 930 Olympia Dr, Charlottesville, VA 22911, (434) 465-6800. 3 reviews. \$\$. Sandwiches, Pizza, Chicken Wings. Description: 'We are officially open for business! Try our Ahlthentic® Italian pizza made with fresh ingredients - dough made in store daily, a 3-cheese blend that is fresh, never frozen, and...' [read more](#)
- Ad The Carving Board Cafe**: 624 Albemarle Sq, Charlottesville, VA 22901, (434) 974-9004. 40 reviews. \$\$. Delis, Caterers, Sandwiches. Description: 'I just moved here and found this place on Yelp when I needed to satisfy a Reuben craving. Unfortunately, the corned beef had just come out of the oven wasn't totally ready, so I had...' [read more](#)
- 1. Lampo**: 205 Monticello Rd, Charlottesville, VA 22902, (434) 244-3226. 188 reviews. \$\$\$. Italian, Pizza.
- 2. Uncle Maddio's Pizza**: 3912 Lenox Ave, Charlottesville, VA 22901, (434) 234-3717. 11 reviews. \$. Pizza, Salad. Description: 'Hard to locate a decent gluten free crust...well Maddios NAILED the delivery!!! Perfect crunch with great sauce to boot!!! Will be returning!' [read more](#)

On the right side, there is a map showing the location of the search results in Charlottesville, VA. Below the map, there is an advertisement for the Rockefeller Foundation with the text: 'Do you agree? - Our core values—are threatened. See curated opportunities from the Rockefeller Foundation to make a difference.'

Apa yang Sebaiknya Diketahui

- Alur kerja dasar dan komponen dalam sistem IR
- Konsep inti dalam IR
- Menjelajah (Browsing) vs. Meminta (Querying)
- Menarik vs. Mendorong informasi

Bacaan Hari Ini

- Introduction to Information Retrieval
 - Bab 19: Web search basics

Pertanyaan?

Terimakasih!