

Temu-balik/Temu-Kembali, menemukan data dari kumpulan data yang banyak

Kayak 10m data, bisa diretrieve dokumen yg relevan.

# Introduction

CSCE604135 Information Retrieval  
Fakultas Ilmu Komputer, UI

Alfan Farizki Wicaksono

# Search

"Search" is one of our primary needs.

"Search" is a mechanism to fulfill our information needs!

- Tiba-tiba ingin makan "es krim goreng". **Search!**
- Anda penasaran dengan bagaimana alam semesta ini diciptakan? **Search!**
- Anda seorang dokter yang sedang menangani kasus yang belum pernah dihadapi sebelumnya. **Search!**
- Anda seorang *librarian* (pustakawan) yang ditugaskan menemukan buku-buku yang membahas tentang *Einstein's general relativity*. **Search!**
- Anda seorang programmer yang *stuck* selama 1 jam memikirkan apa yang salah dari program Anda. **Search!**

Sejarah Information Retrieval (IR) bermula dari **Perpustakaan**.  
Bagaimana mengatur buku-buku yang begitu banyak, dan bagaimana membantu proses search? --> **Catalogue Scheme**

Photo credit: Alistair Moffat

Bagaimana cara mencari buku2 nya?

Biasanya bisa pakai catalog, dimana kita bisa pakai katalog untuk melihat dimana buku disimpan



Sejarah Information Retrieval (IR) bermula dari **Perpustakaan**.  
Bagaimana mengatur buku-buku yang begitu banyak, dan bagaimana membantu proses search? --> **Catalogue Scheme**

Photo credit: Alistair Moffat



Sejarah Information Retrieval (IR) bermula dari **Perpustakaan**.  
Bagaimana mengatur buku-buku yang begitu banyak, dan bagaimana  
membantu proses search? --> **Catalogue Scheme**

M 21498

**ECKEL, Bruce**

Thinking in Java / Bruce Eckel. - 2. izd. -  
Upper Saddle River : Prentice-Hall Ptr, 2000. -  
1127 str. : CD Rom ; 24 cm.

ISBN 0-13-027363-5

519.681

ID=10982

000021498

Sebuah Katalog Buku



# 1920 - 1930

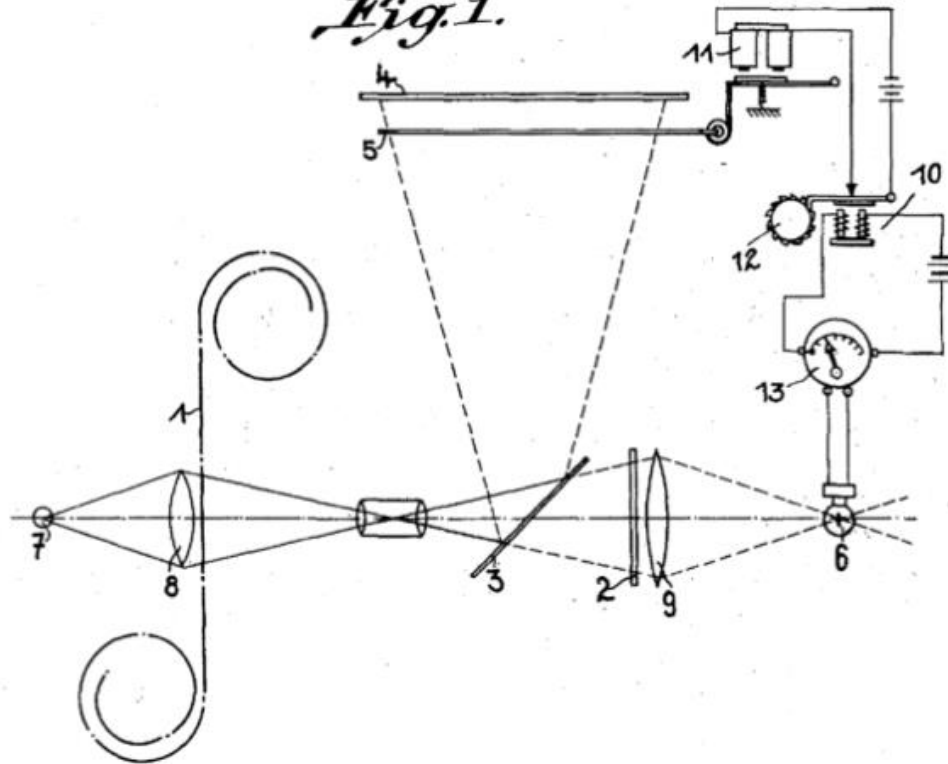
Mechanical devices that searched a catalogue were devised by **Emanuel Goldberg**.

Dec. 29, 1931.

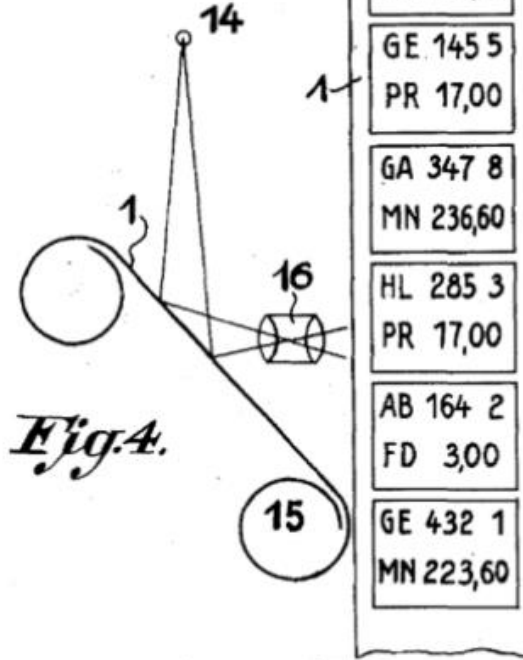
E. GOLDBERG  
STATISTICAL MACHINE  
Filed April 5, 1928

1,838,389

*Fig. 1.*



*Fig. 4.*



*Fig. 2.*

*Fig. 3.*



Inventor:

*Emanuel Goldberg*

Silakan baca "The History of Information Retrieval Research" oleh Mark Sanderson; ada di SCELE

# 1940 - 1950

- Early use of computers for search
- Univac Machine
  - Searching for text references associated with subject code
  - The code and text were stored on a magnetic steep tape
  - This could process at the rate of 120 words per minute.
- A computer developed by General Electric that search over 300,000 document abstracts.

# 1960 - 1990

- Teknologi dan penelitian di bidang IR berkembang pesat.
  - Konsep "term frequency" & "inverse document frequency"
  - Probability Ranking Principle
- Teknologi Komputer dan ukuran storage juga berkembang.
- Sebagian besar IR system untuk mendukung:
  - Library
  - Legal search, WESTLAW's WIN hukum, kita perlu retrieve pasal-pasal yang relevan. Kayak compliance GPT
  - Company-Specific -> search system for NASA



# 1990 - 2000

- Teknologi Web berkembang sangat pesat!
- Banyaknya Dokumen Web meningkat drastis
  - Perlu teknologi **Web Search**



# 2000 – 201x



anak Fasilkom UI



Semua

Gambar

Berita

Shopping

Video

Lainnya

Alat

Sekitar 99.700 hasil (0,33 detik)



youtube.com

<https://m.youtube.com/watch>

## OHVision #2: 5 Anak Fasilkom UI Jujur-jujuran? - YouTube



Hello Dreamers! Open House **Fasilkom UI** hadir kembali dengan video yang seru, nih! Kali ini ceritanya langsung dari 5 orang **mahasiswa** ...

YouTube · Fasilkom UI · 13 Nov 2021



ui.ac.id

<https://cs.ui.ac.id>

## Fakultas Ilmu Komputer Universitas Indonesia – Fasilkom UI ...

Pastikan Anda telah membaca latest updates **Fasilkom UI** mengenai COVID-19 terkait berbagai pengumuman, pembelajaran jarak jauh, pencegahan dan prosedur pelaporan ...

<https://cs.ui.ac.id> › 2014/09/22 › beasiswa-iwk-ui-dan-...

## Beasiswa IWK UI dan IWK Fasilkom untuk Anak-Anak Bangsa

22 Sep 2014 — Kamis (11/9), **Universitas Indonesia** melalui IWK (Ikatan Wanita Keluarga) UI menyelenggarakan kegiatan Penyerahan Beasiswa untuk **Anak-Anak** ...



anak Fasilkom UI



All

Images

Videos

News

Maps

Settings

Indonesia (en)

Safe search: moderate

Any time

<https://cs.ui.ac.id> › about

## Tentang Fasilkom UI - Fakultas Ilmu Komputer Universitas Indonesia

Tentang Fasilkom UI Manajemen Staf Pengajar Tetap (IND Version) Profil Fakultas Ilmu Komputer Universitas Indonesia 2021 2.800+ mahasiswa Terbaik dari seluruh penjuru nusantara 80+ dosen Lulusan perguruan tinggi top dunia 6.000+ alumni Dengan jaringan yang tersebar di tujuh benua Sejarah 1972

<https://cs.ui.ac.id>

## Fakultas Ilmu Komputer Universitas Indonesia - Fasilkom UI. Terbaik d...

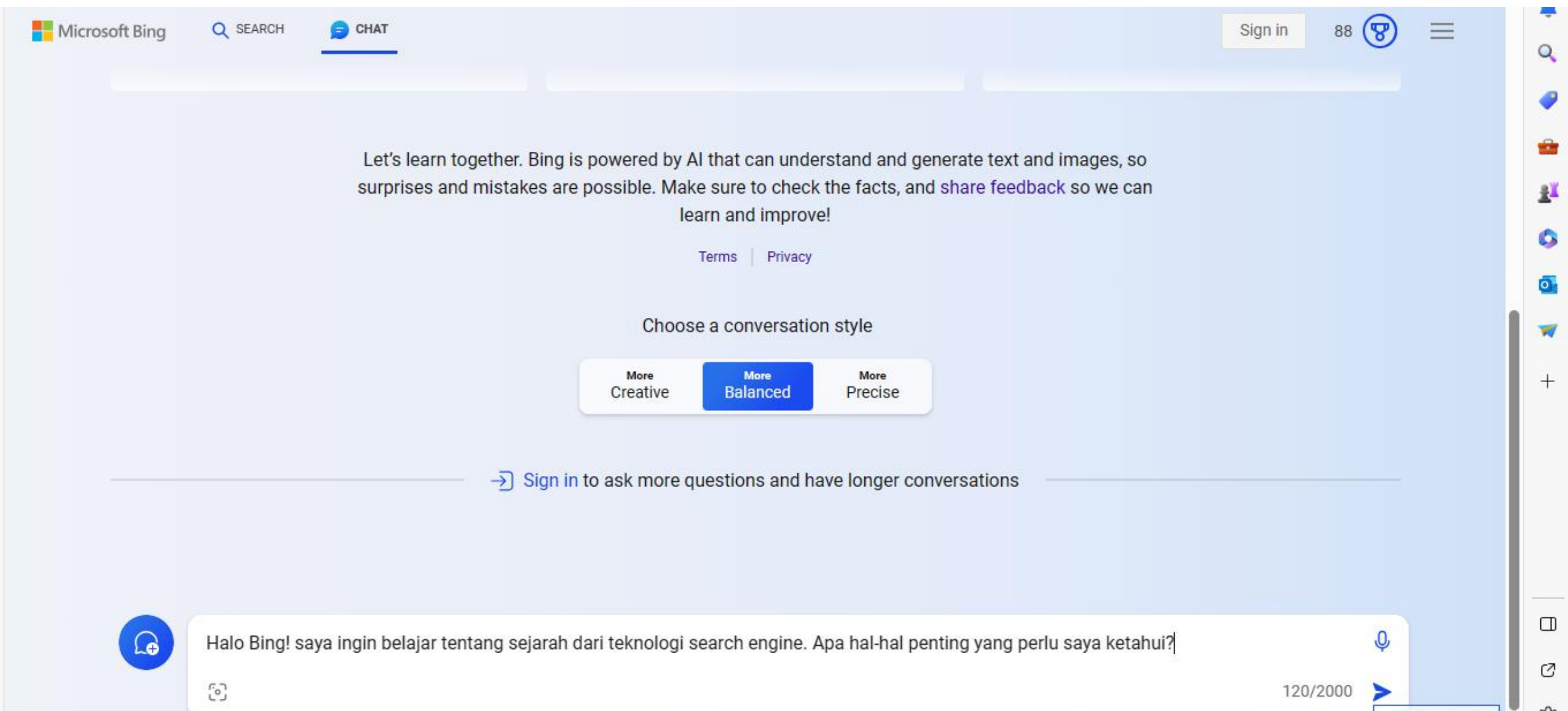
Fasilkom UI Sambut Mahasiswa Baru Program Sarjana Tahun Akademik 2023/2024

<https://www.anakui.com> › ya-begini-nasib-jadi-anak-fasilkom-ui

## Ya Begini Nasib Jadi Anak Fasilkom UI - ANAK UI

FASILKOM UI adalah singkatan dari Fakultas Ilmu Komputer Universitas Indonesia merupakan salah satu fakultas di bawah Universitas Indonesia yang didirikan untuk menghasilkan lulusan sarjana komputer yang mampu bersaing di pasar tenaga kerja global, menciptakan dan mengembangkan pengetahuan dalam bidang ilmu komputer dan teknologi informasi, berp...

# Now: Web Search + Chat Systems



The screenshot shows the Microsoft Bing homepage with a light blue background. At the top, there is a navigation bar with the Microsoft Bing logo, a search bar, and a chat icon. The chat icon is highlighted with a blue underline. To the right of the chat icon, there is a 'Sign in' button and a user profile icon. Below the navigation bar, there is a large text area with the following text: "Let's learn together. Bing is powered by AI that can understand and generate text and images, so surprises and mistakes are possible. Make sure to check the facts, and [share feedback](#) so we can learn and improve!". Below this text, there are links for "Terms" and "Privacy". Further down, there is a section titled "Choose a conversation style" with three buttons: "More Creative", "More Balanced" (which is highlighted in blue), and "More Precise". At the bottom of the page, there is a chat input area with a blue speech bubble icon on the left. The input text reads: "Halo Bing! saya ingin belajar tentang sejarah dari teknologi search engine. Apa hal-hal penting yang perlu saya ketahui?". To the right of the input text, there is a microphone icon. At the bottom right of the input area, there is a character count "120/2000" and a blue arrow icon. On the far right side of the page, there is a vertical sidebar with various icons for different Bing features.

Microsoft Bing

SEARCH

CHAT

Sign in 88

Let's learn together. Bing is powered by AI that can understand and generate text and images, so surprises and mistakes are possible. Make sure to check the facts, and [share feedback](#) so we can learn and improve!

[Terms](#) | [Privacy](#)

Choose a conversation style

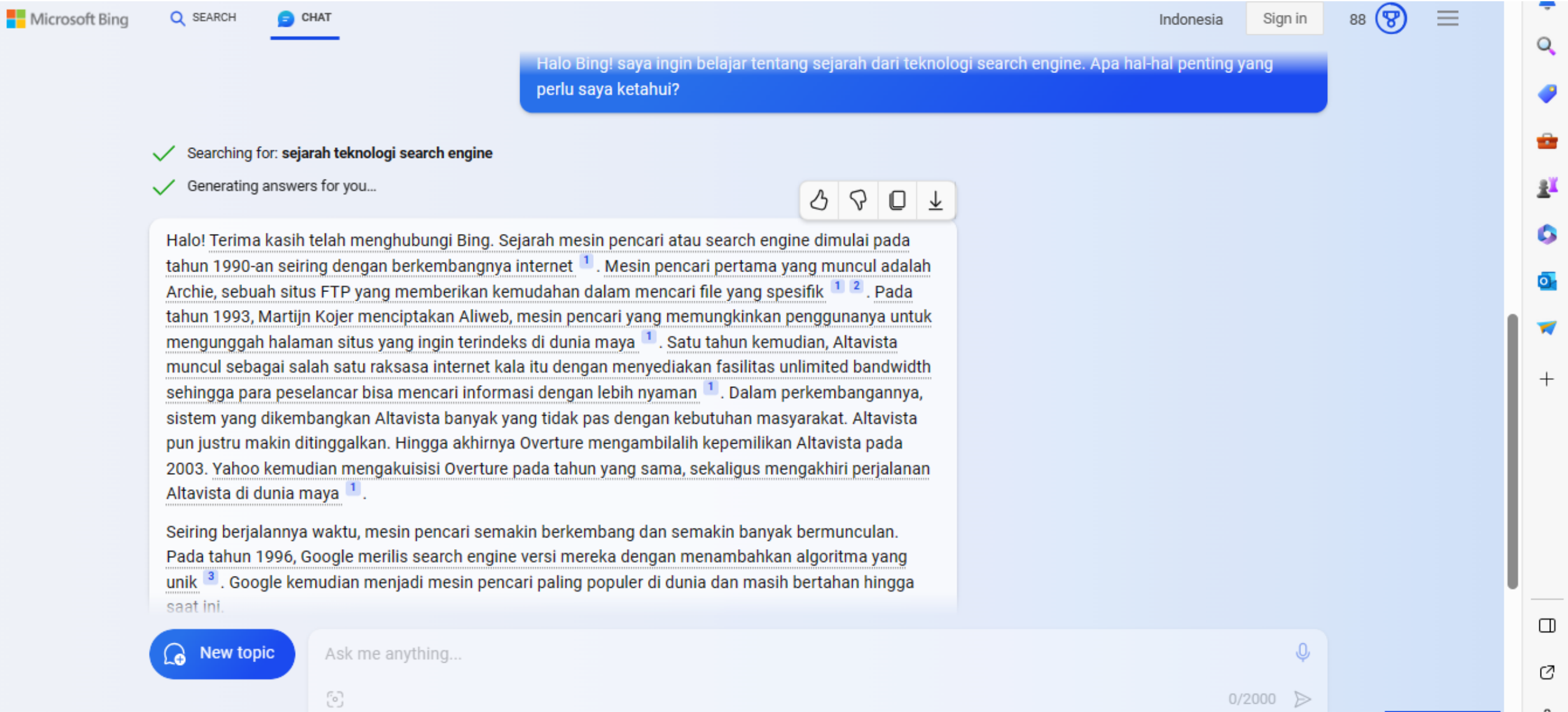
More Creative **More Balanced** More Precise

[Sign in](#) to ask more questions and have longer conversations

Halo Bing! saya ingin belajar tentang sejarah dari teknologi search engine. Apa hal-hal penting yang perlu saya ketahui?

120/2000

# Now: Web Search + Chat Systems



The screenshot shows the Microsoft Bing homepage with the chat interface active. The chat history shows a search for "sejarah teknologi search engine" and a generated answer. The answer discusses the history of search engines, starting from Archie in 1990, Aliweb in 1993, and Altavista, which was eventually acquired by Yahoo in 2003. It also mentions Google's search engine in 1996.

Microsoft Bing SEARCH CHAT Indonesia Sign in 88

Halo Bing! saya ingin belajar tentang sejarah dari teknologi search engine. Apa hal-hal penting yang perlu saya ketahui?

✓ Searching for: **sejarah teknologi search engine**

✓ Generating answers for you...

Halo! Terima kasih telah menghubungi Bing. Sejarah mesin pencari atau search engine dimulai pada tahun 1990-an seiring dengan berkembangnya internet <sup>1</sup>. Mesin pencari pertama yang muncul adalah Archie, sebuah situs FTP yang memberikan kemudahan dalam mencari file yang spesifik <sup>1 2</sup>. Pada tahun 1993, Martijn Koster menciptakan Aliweb, mesin pencari yang memungkinkan pengguna untuk mengunggah halaman situs yang ingin terindeks di dunia maya <sup>1</sup>. Satu tahun kemudian, Altavista muncul sebagai salah satu raksasa internet kala itu dengan menyediakan fasilitas unlimited bandwidth sehingga para peselancar bisa mencari informasi dengan lebih nyaman <sup>1</sup>. Dalam perkembangannya, sistem yang dikembangkan Altavista banyak yang tidak pas dengan kebutuhan masyarakat. Altavista pun justru makin ditinggalkan. Hingga akhirnya Overture mengambil alih kepemilikan Altavista pada 2003. Yahoo kemudian mengakuisisi Overture pada tahun yang sama, sekaligus mengakhiri perjalanan Altavista di dunia maya <sup>1</sup>.

Seiring berjalannya waktu, mesin pencari semakin berkembang dan semakin banyak bermunculan. Pada tahun 1996, Google merilis search engine versi mereka dengan menambahkan algoritma yang unik <sup>3</sup>. Google kemudian menjadi mesin pencari paling populer di dunia dan masih bertahan hingga saat ini.

New topic Ask me anything... 0/2000

# Information Retrieval

**Goal:** finding material (usually documents) of an **unstructured nature** (usually text) that satisfies an **information need** from within **large collections** and helps the user complete a **task**.

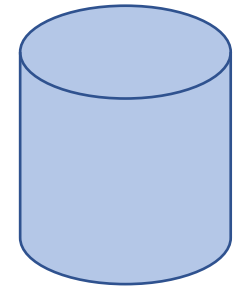
User's Head:  
Information Need



Query



Search Engine



Large Collections

Hasil search diharapkan "relevan"  
terhadap information need.

**Bukan relevan terhadap query!**

Results

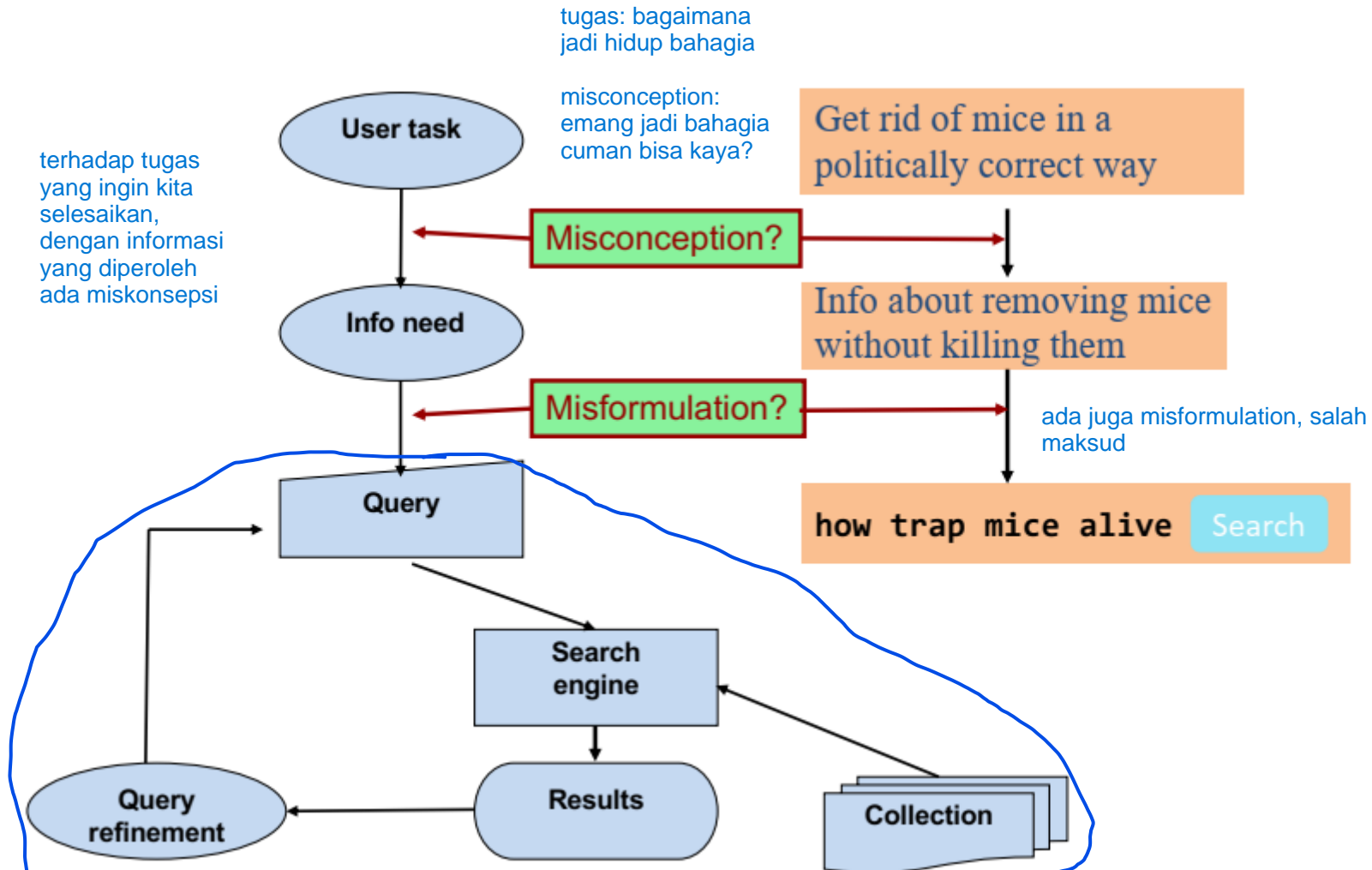
1. Doc #43
2. Doc #87
3. ...

ada ranking

harus relevan terhadap information neednya. Jangan relevan terhadap query.



# The Classic Search Model



Contoh lain

Lakukan kajian bagaimana hidup Anda bisa Bahagia.

Butuh informasi tentang membangun bisnis dan cepat kaya

"bisnis untung besar" Search



# IR vs Databases

- IR Systems

- **Koleksi:** tidak terstruktur atau semi-terstruktur
- **Queries:** boleh berupa *natural language*; query bisa ambigu
- **Results:** diharapkan relevan terhadap *information needs*, dan berguna untuk menyelesaikan sebuah *task*; dan bisa diurutkan berdasarkan skor.

bahasa nya udah gak terstruktur (misal buaya vs crocodile kan sama aja, tapi bisa beda-beda)

hoping to be relevant

- Database Management Systems

- **Koleksi:** terstruktur (misal, dalam bentuk tabel dan relasi antar tabel)
- **Queries:** terstruktur, formal dan tidak ambigu, seperti SQL pada *relational databases*
- **Results:** exact matches dengan query

Exact match

# Google -> hundreds of billions of webpages!!


The image shows a Google search interface. The search bar contains 'dr who'. Below the search bar, the 'All' tab is selected and circled in red. Below the tabs, it says 'About 7,290,000,000 results (0.79 seconds)'. The search results show a card for 'Doctor Who' with a small image of the TARDIS, the title 'Doctor Who', and the text '1963 · Sci-fi · 26 seasons'. To the right of the card are buttons for 'Overview', 'Cast', 'Episodes', 'Trailers & clips', 'Reviews', and 'Merchandise'. Below the 'Cast' button, there is a section titled 'Cast >' with a row of six actor portraits: Mandip Gill (Yazmin Khan), Matt Smith, Tom Baker (The Doctor, ...), David Tennant, William Hartnell (The Doctor), and Jodie Whittaker (Thirteenth D...). Below the cast section, there is a link to 'https://en.wikipedia.org > wiki > Doctor\_Who' and a heading 'Doctor Who - Wikipedia'. Below the heading, it says 'Doctor Who is a British science-fiction television programme broadcast by the BBC since 1963. The programme depicts the adventures of a Time Lord called the ...'. Below this, there is a heading 'Doctor\_(Doctor\_Who)' and the start of a paragraph 'Series leads: William Hartnell (1963–1966); Patrick Troughton'.

Google

dr who


All Images News Videos Shopping More Tools


About 7,290,000,000 results (0.79 seconds)


 **Doctor Who**  
1963 · Sci-fi · 26 seasons


Overview Cast Episodes Trailers & clips Reviews Merchandise


Cast >


  
Mandip Gill  
Yazmin Khan

  
Matt Smith

  
Tom Baker  
The Doctor, ...

  
David Tennant

  
William Hartnell  
The Doctor

  
Jodie Whittaker  
Thirteenth D...

[https://en.wikipedia.org > wiki > Doctor\\_Who](https://en.wikipedia.org/wiki/Doctor_Who)

**Doctor Who - Wikipedia**

Doctor Who is a British science-fiction television programme broadcast by the BBC since 1963. The programme depicts the adventures of a Time Lord called the ...

**Doctor\_(Doctor\_Who)**

Series leads: William Hartnell (1963–1966); Patrick Troughton

Already watched Want to watch

About

8.4/10 · [IMDb](#)

92% liked this TV show  
Google users

Doctor Who is a British science-fiction television programme broadcast by the BBC since 1963. The programme depicts the adventures of a Time Lord called the Doctor, an extraterrestrial being who appears to be human. The Doctor explores the universe in a time-travelling space ship called the TARDIS.

# Big Picture Kuliah IR Selama 1 Semester

dari pilih 1000 data yang diretrieve, untuk ranking top 10, itu pakai ML

