SEARCH ENGINE

THE MECHANISM BEHIND WEB SEARCH

MADE BY PRATHAM MEHTA

INTRODUCTION TO SEARCH ENGINES

01 WHAT IS A SEARCH ENGINE?

- A system that crawls, indexes, and ranks web pages
- Examples: Google, Bing, Yahoo, DuckDuckGo

02 WHY ARE THEY IMPORTANT?

- Help users find relevant information
- Handle billions of queries per day

03 MAIN COMPONENTS

- CRAWLING – COLLECTING WEB PAGES
- INDEXING – STORING AND STRUCTURING DATA
- RANKING – SORTING RESULTS BASED ON RELEVANCE

STEP 1
CRAWLING THE WEB

- **Process of scanning web pages**
- **Uses bots (web crawlers/spiders)**

01

HOW CRAWLERS WORK?

- **Start with a seed URL (e.g., Google.com)**
- **Follow links on the page**
- **Add new pages to queue for crawling**
- **Repeat until millions of pages are collected**

02

CHALLENGES IN CRAWLING:

- **Keeping pages updated**
- **Handling duplicate & spam pages**
- **Managing billions of websites**

Googlebot → Visits CNN.com → Follows links → Adds new pages to crawl

STEP 2
INDEXING THE WEB

- Organizing collected web pages in a database
- Converts raw HTML into structured data

HOW INDEXING WORKS?

- Extracts text, metadata, images
- Breaks text into tokens (words)
- Builds inverted index for fast searching

WHY IS INDEXING IMPORTANT?

- Fast retrieval of information
- Removes duplicate content

DIFFERENCE BETWEEN INDEXING & INVERTED INDEXING

Defination

- Indexing: A technique to improve search speed by organizing and storing data efficiently.
- Inverted Indexing: A specialized form of indexing used in search engines to map words to the documents they appear in.

Basic Indexing (Forward Indexing)

Stores documents sequentially with their corresponding words.

- Doc 1 → [apple, banana, orange]
- Doc 2 → [banana, grape, mango]

Efficient for retrieving full documents but slow for searching specific words across documents.

Inverted Indexing

Instead of mapping documents to words, it maps words to documents

- apple → [Doc 1]
- banana → [Doc 1, Doc 2]
- grape → [Doc 2]

Allows quick lookup of documents containing a specific word.

STEP 3
RANKING SEARCH RESULTS

HOW SEARCH ENGINES RANK RESULTS?

Uses ranking algorithms to sort relevant pages

Ranking Factors:

- **Keyword Matching – Checks if query words appear in content**
- **Page Authority – Trustworthiness of the website**
- **Backlinks – Links from other websites**
- **User Experience – Mobile-friendliness, page speed**

Google’s Algorithm (PageRank, BERT, RankBrain)

- **Uses machine learning & NLP to improve search results**

📌 **Example:**
🔍 **Query: “Best smartphones 2025”**

- **Page A (Ranking 1st) – Trusted tech blog, many backlinks**
- **Page B (Ranking 5th) – New website, fewer backlinks**

RANKING WITH TF-IDF

(TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY)

TF (TERM FREQUENCY) – MEASURES HOW OFTEN A WORD APPEARS IN A DOCUMENT.

$$TF = \frac{\text{Number of times the word appears in a document}}{\text{Total words in the document}}$$

IDF (INVERSE DOCUMENT FREQUENCY) – MEASURES HOW IMPORTANT A WORD IS ACROSS ALL DOCUMENTS.

$$IDF = \log\left(\frac{\text{Total number of documents}}{\text{Number of documents containing the word}}\right)$$

CALCULATING TF-IDF SCORE

$$TF - IDF = TF \times IDF$$

STEP 4
DISPLAYING SEARCH RESULTS

How results are Shown?
Search Engine Results Page (SERP)
Snippet, Title, URL, and Meta Description

Types of Results

ORGANIC RESULTS

Based on ranking algorithms

PAID ADS (GOOGLE ADS)

Businesses pay for visibility

FEATURED SNIPPETS

Instant answers for common queries

Why Search Engine UI Matters?

- Click-Through Rate (CTR) – Higher rankings get more clicks
- Personalization – Results tailored to user preferences

THE ROLE OF AI IN SEARCH ENGINES

How AI is Changing Search?

Machine Learning improves ranking
Natural Language Processing (NLP) understands queries better

01

Google RankBrain

- RankBrain is an AI-driven ranking algorithm introduced by Google in 2015.
- It helps Google process and understand search queries that are new or ambiguous by identifying patterns in past queries.

02

BERT (Bidirectional Encoder Representations from Transformers)

- Introduced by Google in 2019, BERT is a Natural Language Processing (NLP) model that improves how search engines understand the meaning of words in context.
- Unlike older algorithms that process words individually, BERT reads the entire sentence at once, helping Google understand nuanced queries.

03

Voice Search & Chatbots

- With the rise of virtual assistants like Google Assistant, Alexa, Siri, and Cortana, voice search is becoming more important.
- AI models like BERT and Google's LaMDA (Language Model for Dialogue Applications) are improving conversational understanding in search.

CONCLUSION

Recap of How Search Works:

01 **Crawling**
Collects web pages

02 **Indexing**
Organizes content

03 **Ranking**
Sorts relevant results

04 **Displaying Results**
Shows best matches


THANK
YOU