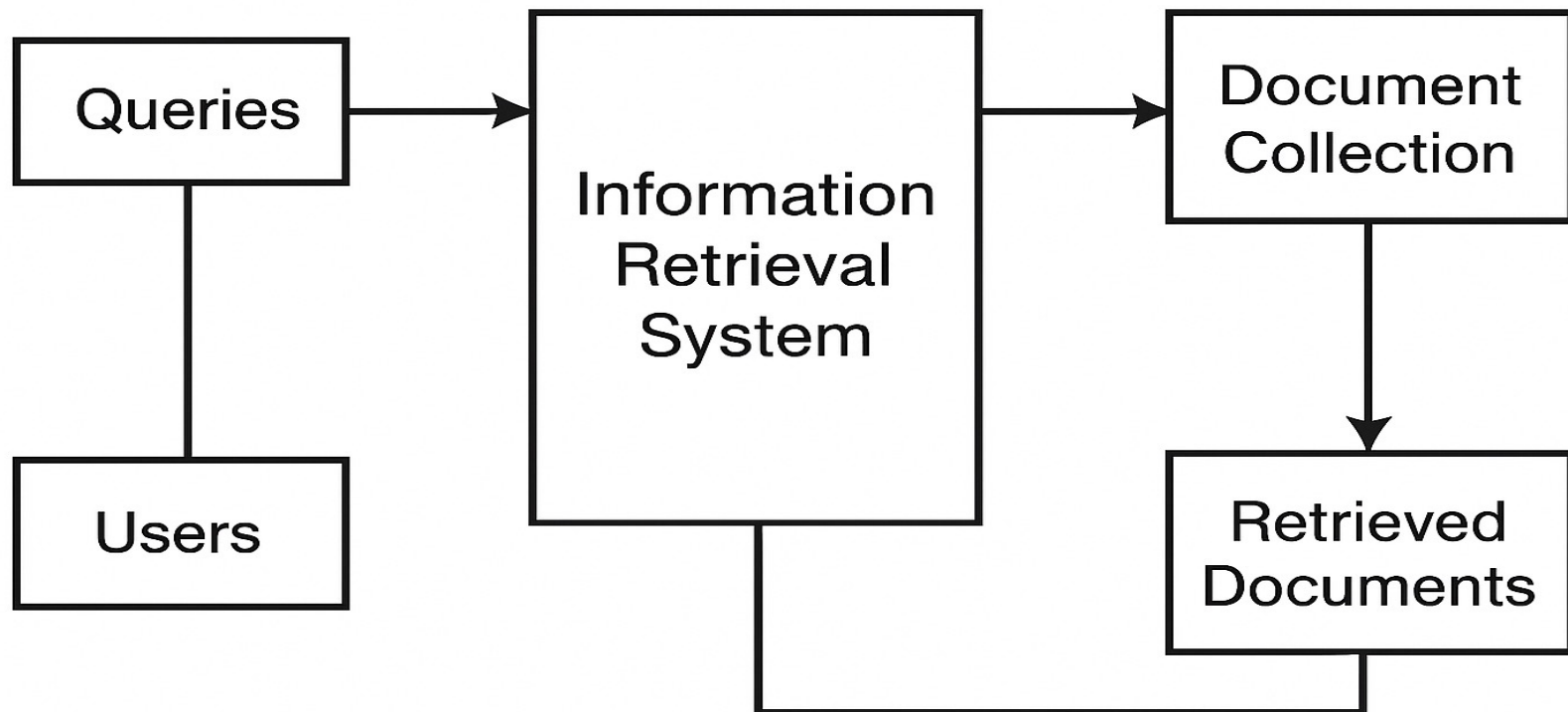# Information Retrieval (IR) Systems

Information Retrieval

- Information Retrieval (IR) = finding relevant documents from a large collection based on a user query.
  Example: Google Search.

IR Architecture

Document Collection / Corpus

- Large set of documents (web pages, PDFs, emails, etc.)

- Forms the **database** from which results are retrieved.

Document Processing / Pre-processing

- Before indexing, every document goes through NLP steps:
- Tokenization
- Stop-word removal
- Stemming / Lemmatization
- Normalization
- POS tagging (sometimes)
- Goal → convert raw text into clean tokens.

Indexing

- IR builds an **Inverted Index**:
- Each **word** → mapped to list of documents where it appears
- Speeds up searching massively
- Similar to index at the back of a book

```
cat → doc1, doc7, doc19
camera → doc3, doc5
```

Index = the **heart** of any IR system.

Query Processing

- User enters a query → IR system processes it:

- Tokenize the query

- Remove stop words

- Apply stemming

- Expand query (synonyms, spell correction)

- Example:
  User: "affordable mobile camera"
  Query expansion → "cheap phone camera"

Matching & Ranking

- The query is compared with documents using:
- TF-IDF
- Vector Space Model
- BM25
- Language Models
- BERT / Semantic embeddings (modern systems)
- Each document receives a **similarity score**, then the system:
- **Ranks documents** from most relevant to least relevant.

User Interface & Result Presentation

Finally, IR displays results:

- Titles

- Snippets

- URLs

- Highlighted query terms

- Filters

- Example: Google search result page.

Evaluation Metrics in Information Retrieval

- IR evaluation measures **how good** the system is at returning relevant documents.

- There are two main categories:

A) Basic Metrics

B)Ranking-Based Metrics (used for Google-like systems)

Basic Metrics & Ranking-Based Metrics

- 1) Precision
- 2) Recall
- 3) F-Measure (F1-score)

- 1) Mean Average Precision (MAP)
- 2)Discounted Cumulative Gain (DCG)
- 3) Mean Reciprocal Rank (MRR)

⭐ 1) Precision

Precision = **Out of the documents retrieved, how many were relevant?**

✔️ **Formula:**

$$Precision = \frac{\text{Relevant Retrieved}}{\text{Total Retrieved}}$$

✔️ **Example:**

System retrieved 10 documents.

- Relevant = 6
- Irrelevant = 4

$$Precision = \frac{6}{10} = 0.6$$

👉 Means: 60% of the retrieved results were correct.

⭐ 2) Recall

Recall = **Out of all the relevant documents in the system, how many did we retrieve?**

✔️ **Formula:**

$$Recall = \frac{\text{Relevant Retrieved}}{\text{Total Relevant in Database}}$$

✔️ **Example:**

In the whole database: 20 documents are relevant.
System retrieved only 6 of them.

$$Recall = \frac{6}{20} = 0.3$$

👉 Means: system found only 30% of the relevant material.

What is F1-Score?

F1-Score is the harmonic mean of precision and recall.

It gives one single value that balances both precision and recall.

Why do we need F1-Score? (Use / Purpose)

Sometimes:

- Precision is high
- Recall is low

or

- Recall is high
- Precision is low

Which metric is better?
It becomes difficult to judge performance.

F1-Score solves this problem by combining both into one metric.

Where is F1-Score used?

We use F1-Score when:

1. We need a balance between precision and recall
   Especially when both are important.

Example: medical diagnosis system

- Precision helps avoid false positives
- Recall helps avoid false negatives
- Both are very important
  Therefore, F1-Score is used.

⭐ 3) F1-Score (Harmonic Mean of Precision & Recall)

Balances precision and recall.

✔️ Formula:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Using previous values:

Precision = 0.6

Recall = 0.3

$$F1 = 2 \cdot \frac{0.6 \times 0.3}{0.6 + 0.3}$$

$$F1 = 2 \cdot \frac{0.18}{0.9} = 0.4$$

👉 F1 = 0.4 (balanced performance measure)

⭐ 4) Average Precision (AP)

Used when results are **ranked** (like Google).

We calculate precision **every time** we retrieve a relevant document.

✔️ Example (very simple):

Ranked list returned:

Rank	Document	Relevant?
1	D1	Yes
2	D2	No
3	D3	Yes
4	D4	Yes
5	D5	No

Calculate:

- At rank 1 → Precision = 1/1 = **1.0**
- At rank 3 → Precision = 2/3 = **0.66**
- At rank 4 → Precision = 3/4 = **0.75**

AP = average of these:

$$AP = \frac{1.0 + 0.66 + 0.75}{3} = 0.80$$

👉 Higher AP = more relevant documents appears early.

⭐ 5) Mean Average Precision (MAP)

Just mean (average) of AP values over **multiple queries**.

Used for evaluating entire search engines.

⭐ 2) MAP = Mean Average Precision

MAP is simply the **average AP over multiple queries.**

Example:

Query 1 → AP = 0.80

Query 2 → AP = 0.50

Query 3 → AP = 1.00

$$MAP = (0.80 + 0.50 + 1.00)/3 = 0.76$$

✔️ MAP measures **whole search engine performance**, not one query

✔️ Higher MAP = better ranking quality

⭐ 3) DCG (Discounted Cumulative Gain)

Used when documents have **graded relevance**, like:

- 3 = highly relevant
- 2 = relevant
- 1 = somewhat relevant
- 0 = not relevant

DCG gives **more weight** to documents at the top.

✔️ Simple DCG Example

Ranked list:

Rank	Relevance
1	3
2	2
3	1

Formula:

$$DCG = rel_1 + \frac{rel_2}{\log_2(2)} + \frac{rel_3}{\log_2(3)}$$

Now plug numbers:

- $rel_1 = 3$
- $rel_2 = 2$
- $rel_3 = 1$

Calculate:

$$DCG = 3 + \frac{2}{1} + \frac{1}{1.58}$$

= 3 + 2 + 0.63

= 5.63

DCG tells: *how good the ranking is considering relevance + position*