Experiment No. 3 (Group - B)

Problem Statement:

Implement a program to calculate precision & recall for sample input.

Objectives:

1. To understand precision & recall in information retrieval.
2. To study indexing structures for informatn retrieval.

Theory:

Precision & Recall in Information Retrieval.

Informatn Retrival system can be measured with two metrices: precision & recall. It can be divided into 4 categories.

1) Relavant & Retrieved
2) Relavant & not retrieved
3) non-relavant & Retrieved
4) Non - relevant & non retrieved.

Relevant items are those documents that help the user in answering his question. Non relavent items are items that don't provide actually useful information. for each item there are two possibilities. it can be retrieved by the user is query precision is defined as the ration of the no. of relevant & retrieved documents.

- Recall is defined as ratio of the no. of retrieved & relavent documents to the / possible relevant documents.

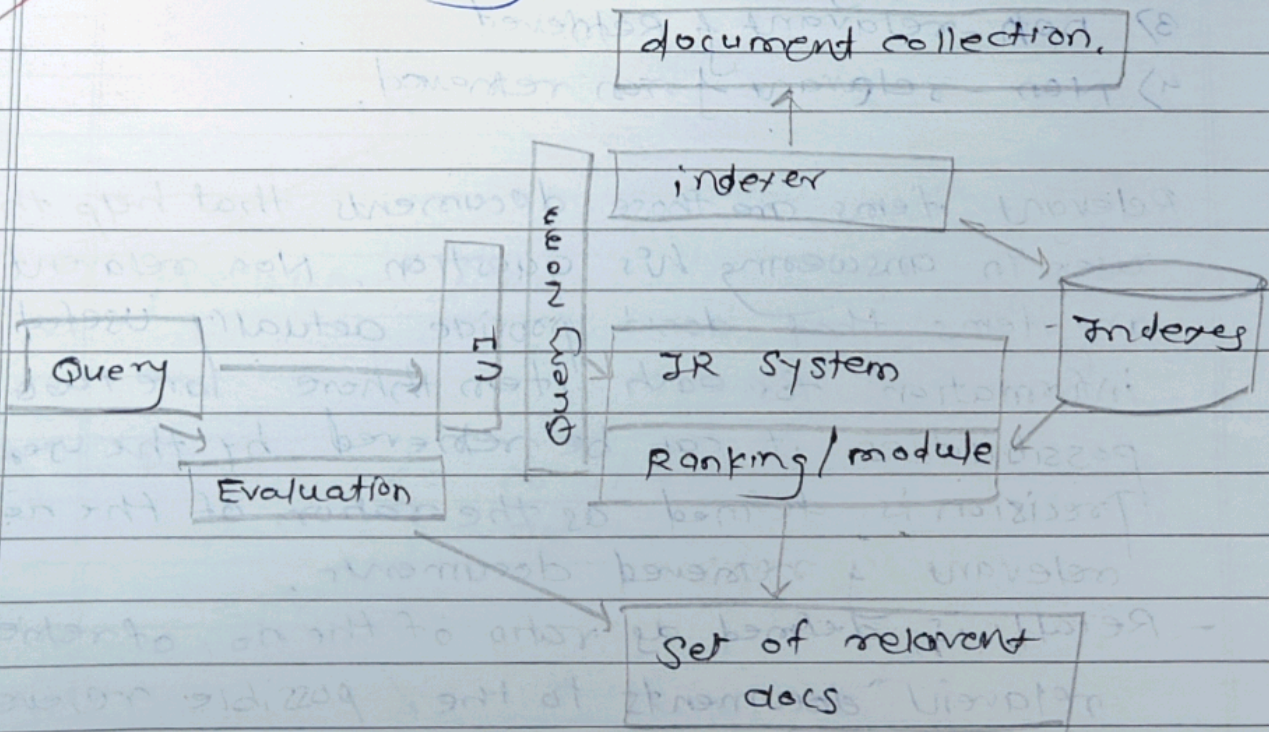Precision:- $\dfrac{\#(\text{relevant item retrieved})}{\#(\text{retrieved items})}$

Recall $= \dfrac{\#(\text{relevant item retrieved})}{\#(\text{relevant items})}$

The truth

What the system thinks		relevant	Non relevant	
	Retrived	TP	FP	e-
	Non retrievd	FN	TN	

Precision / Recall trade - off:

You can increase recall by returning more docs. Recall is a non - decreasing funn of the no. of docs retrieved.

document collection.

↑

indexer

Query

Query Processing

UI

Evaluation

IR System

Ranking / module

Indexes

Set of relevant docs

Precision & Recall explanation:

Consider,

I : an information request

R : the set of relevant documents for I
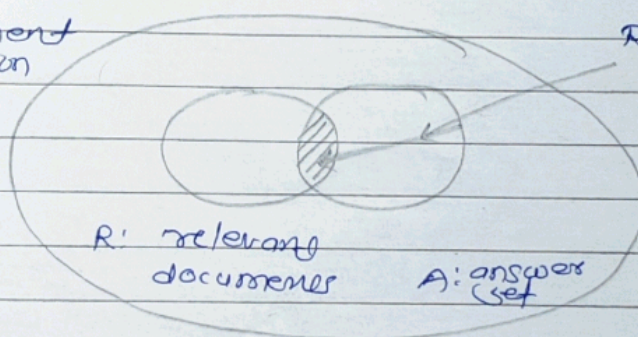
A : the answer for I, generated by IR system

R∩A : the intersection of the sets R & A

|A| - no. of docs in the set A

|Ra| - no. of docs in the intersection of sets R & A

document collection

R∩A : relevant document in the answer set

R : relevant documents   A : answer set

Recall : $\dfrac{|R \cap A|}{|R|}$

Precision = $\dfrac{|R \cap A|}{|A|}$
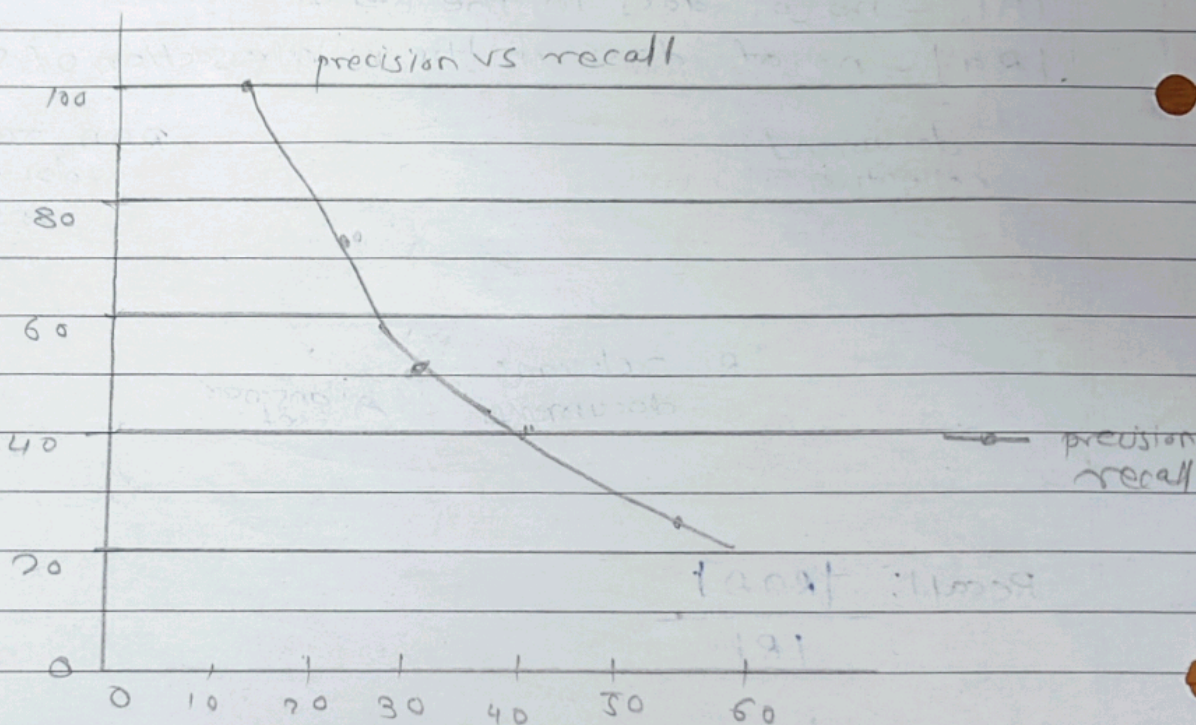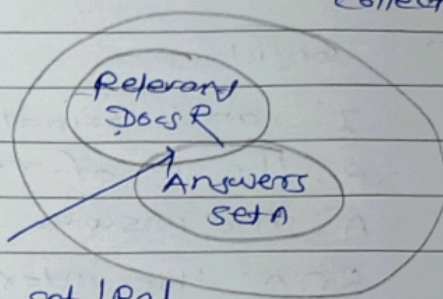
Recall : $R = \dfrac{\text{no. of relevant items retrieved}}{\text{Total no. of relevant items in collection}}$

Precision : $P = \dfrac{\text{no. of relevant items retrieved}}{\text{Total no. of items retrieved}}$

$$\text{Recall} = \frac{|R_q|}{|R|}$$

Collection

$$\text{precision} = \frac{|R_a|}{|A|}$$

relevant docs in any set $|R_a|$

Relevant Docs R

Answers set A

precision vs recall

Conclusion :- Implementation is concluded by execu
a program to calculate precision and recall
for (sample input with relevent document
R_{q1} } for query q_2

21-10-2024