Name : Dhiraj Bodake

Roll No : 18141216

TUTORIAL NO : 07

Q.1 Explain in brief : Recall and Pression.

→ In pattern recognition, information retrieval and classification precision and recall are perform once matrics that apply to data retrieval from collections a sample space.

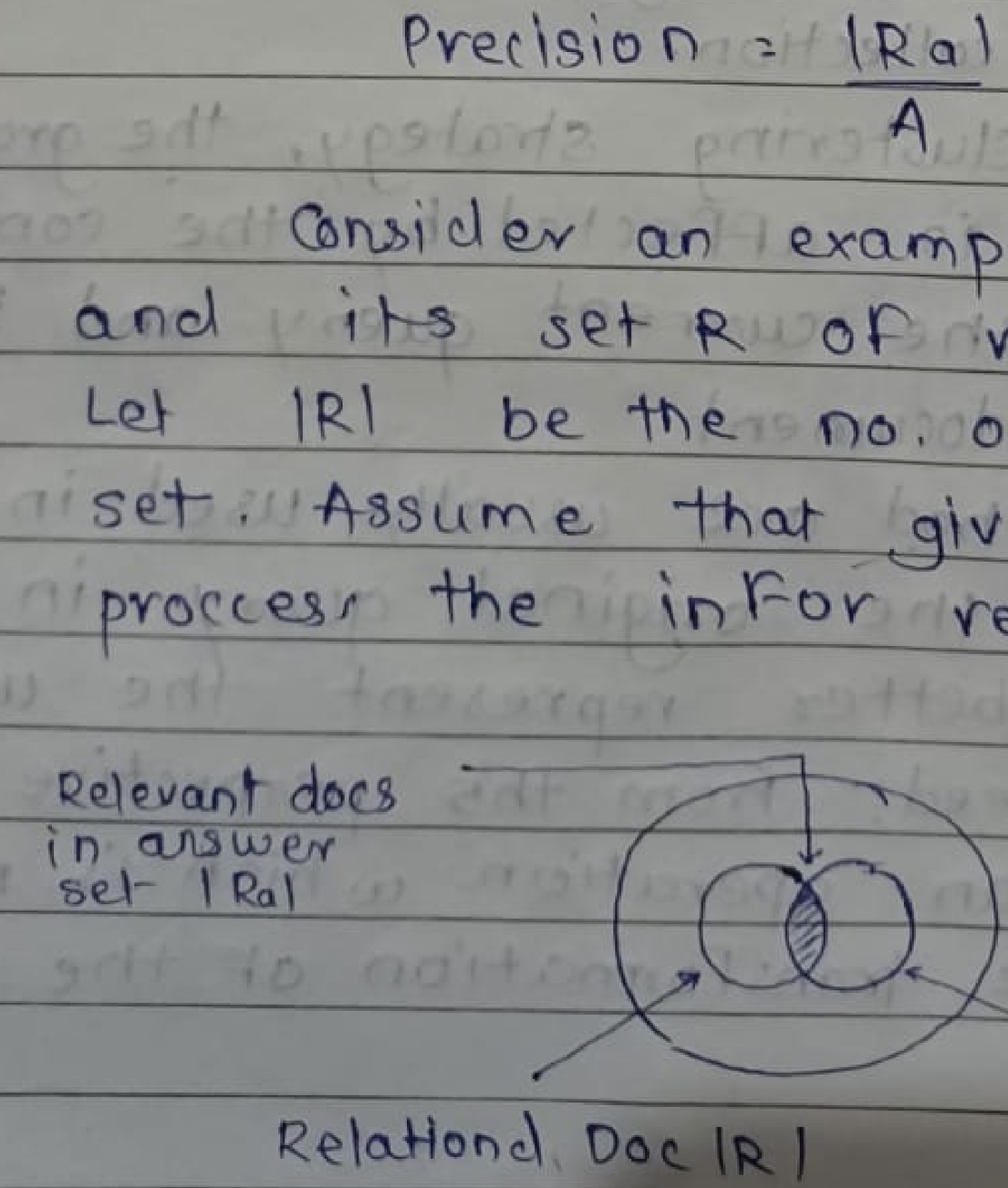
- Recall. - It is the fraction of refrant doc which has been retrieved.

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

- Rrecision - It is the fraction of retrieved document which is relevant.

$$\text{Precision} = \frac{|Ra|}{A}$$

Consider an example information request and its set R of relevant documents. Let |R| be the no. of documents in this set. Assume that given relevant strategy process the infor request.



Q 2

Explain in details: five text operations of document pre-processing.

→ Document pre-processing is a procedure which can be divided mainly into five text operations.

① Lexical analysis of text with the objective of treating digits, hyphens, punctuation marks and case of letters.

② Elimination of stopwords with the objective of filtering out words with very low discrimination values for retrieval purposes.

③ Stemming of remaining words with the objective of removing affixes (prefixes and suffixes) and allowing the retrival of documents containing syntatic variations of query terms (e.g. connect, connecting, etc)

④ Selection of index terms to determine which words / stem will be used as an indexing elements. Usually decision on whether a particular word will be used as index term is related to the syntatic nature of the word, its

⑤ Construction of term categorization structures such as the sources or extraction of structure directly represented in the text, for allowing expansion of original query with related terms.

Q 3 Explain in detail: Logical view of document preprocessing with help of suitable diagram.

→ The retrival system adopts full text logical view of the documents. with very large collection however, even modern computers might have to reduce the set of representative keywords. This can be accomplished through elimination of stopwords, the use of stemming and the identification of noun groups.

- Further compression might be employed. These operations are called text operations.
- Text operations reduce complexity of doc representation and allow moving the logical view from that of full text to that of set of index terms.