

Name: Dhiraj Ravindra Bodake

R.No: 18/4/216

## TUTORIAL NO: 01

Q. 1. What is information retrieval? Explain Data Retrieval & information retrieval in details.

→ Information Retrieval (IR) deals with the representation, storage, organization of data & access to info. item. The representation & organization of info item should provide the user with easy access to info in which he is interested.

- Data retrieval: Data retrieval in context of IR system consists mainly of determining which documents of the collection contain keyword in user query which most frequently is not enough to satisfy user information need.

A data retrieval language, aims at retrieving all object which satisfy clearly defined conditions such as those in regular expression or relational algebra expression.

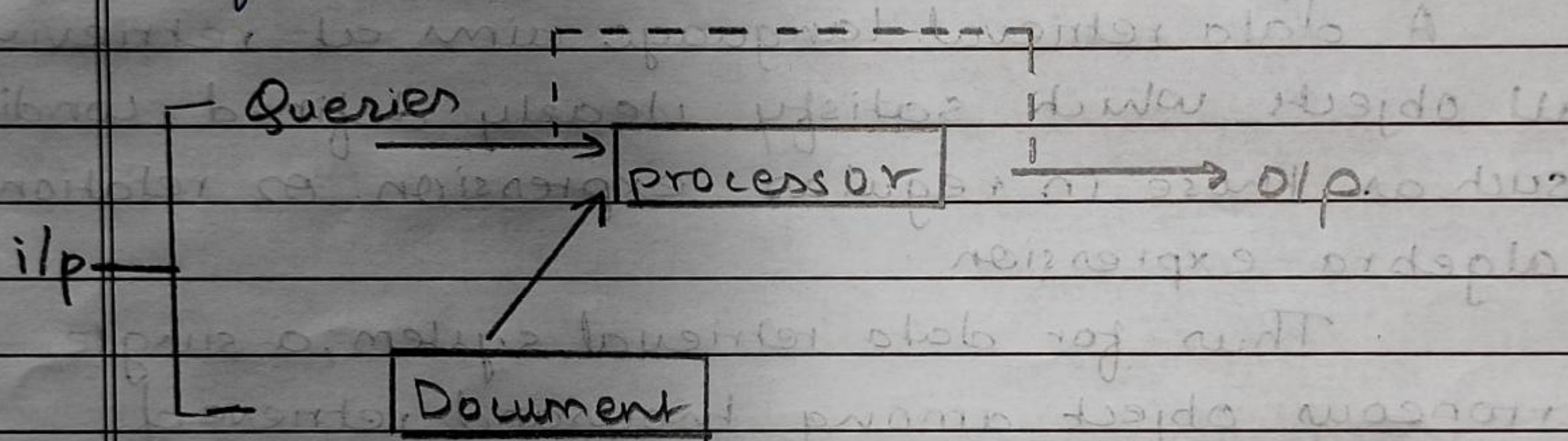
- Information retrieval: For information retrieval system, the retrieval object might be inaccurate & small errors are likely to go unnoticed. The main reason for this difference is that info retrieval usually deals with natural language text which is not always well structured.



To be effective in its attempt to satisfy the user info. need, IR system must somehow interpret contents of info. items in a collection & rank them according to degree of relevance to the user query.

Q.2. Explain typical IR system in details with its block diagram.

→ Info. retrieval system is set of algo. that facilitate relevance of displayed documents to search queries. In simple words, it works to sort & rank documents based on queries of user.



Starting with ILP things the main problem is here to obtain representation of each document & query suitable for a computer to use. A document representative method could be list of extracted words considered to be significant rather than have the computer process natural language.



Alternative approach is to have an artificial approach i.e. to have an artificial language within which all queries & documents can be formulated when retrieval system is online. It's possible for user to change this request during on search session in light of sample retrieval. Such procedure commonly referred to feedback.

Secondly, processor is that part of retrieval system concerned with retrieval process. In diag. document have been separated to emphasize the fact that they are not just  $ilp$  but can be used during retrieving process.

Q.3. What is automatic text analysis? Explain concept of Luhn's idea & null hypothesis with the help of diagram.  
→ Consider the diagram.

Let  $f$  be frequency of occurrence of various word types in given position of text.

$r$  be the rank order, order of frequency of occurrence plot relating  $f$  &  $r$  yields a curve that is similar to hyperbolic curve. The curve also demonstrates Zipf's law.

According to statement by Luhn - It is here proposed that frequency of word occurrence in an article furnishes a useful measurement



for determining significance of word.

It is further proposed that relative position within sentence of words having given values of significance furnish useful measurement for determining significance of sentences.

