INFORMATION RETRIEVAL PROJECT

(Wikipedia - Countries and territories search)

Group Details:

Name: DIVVELA SUMANTH

Roll Number: S20210010067

Email: sumanth.d21@iiits.in

The main components used are:

(1)Indexing : Indexing in information retrieval involves organizing and structuring data to expedite efficient search and retrieval processes. It creates indices or data structures that store information about the location of key terms within documents, facilitating quick access. The primary goal is to optimize search speed and accuracy by allowing systems to swiftly identify and retrieve relevant documents based on user queries. Common techniques include inverted indexing, where

terms point to document locations, and forward indexing, where documents point to term occurrences.

(2)Searching : The Searching component in information retrieval executes user queries to locate and retrieve relevant documents from a database. It involves algorithms that match query terms with indexed data to produce accurate and timely results. This component is crucial for delivering effective information retrieval by ensuring the identification and presentation of documents that meet user search criteria. Advanced search techniques, relevance ranking, and real-time processing are often employed to enhance the precision of search outcomes.

(3)Retrieving Top 10 Documents:Refining searches in information retrieval involves applying specific filters to narrow down and customize search results. Users can employ filters based on criteria such as date, document type, or source to tailor information retrieval to their preferences. This process enhances precision by focusing on relevant subsets of data, streamlining the search experience. Filters provide users with a targeted approach to accessing the most pertinent information within a large dataset.

(4)Relavence Feedback: Relevance feedback in information retrieval is a process where users provide feedback on retrieved results to improve subsequent searches. Users mark documents as relevant or non-relevant, and the system adjusts its algorithms based on this feedback. This iterative approach refines search queries, enhancing the system's ability to deliver more accurate and contextually relevant results over time. Relevance feedback is a dynamic mechanism that aims to bridge the gap between user expectations and the information retrieval system's performance.

(5)Assessment Component:

Precision and Recall Calculation:

Precision in information retrieval is the ratio of relevant documents retrieved to the total retrieved, measuring result accuracy. Recall is the ratio of relevant documents retrieved to the total number of relevant documents in the collection, assessing result comprehensiveness. Precision and recall are often used together to evaluate the effectiveness of a search algorithm in balancing accuracy and coverage.

11-Point Interpolated Precision-Recall Curve:

The 11-point interpolated precision-recall curve is a graphical representation of a model's performance in binary classification. It is constructed by calculating precision and recall at 11 equidistant points between 0 and 1 for different classification thresholds. Interpolation is used to connect these points smoothly, providing a more detailed evaluation of a model's trade-off between precision and recall across a range of operating points. This curve is particularly useful for assessing classifier performance when the data is imbalanced.