CCS/00009/020
CCS 416: INTRODUCTION TO INFORMATION RETRIEVAL

Definition

Information Retrieval is the process of finding and retrieving information resources that are relevant to a user's query or need. A classic example is a search engine, which provides links to web pages based on natural language queries. These resources can be structured, semi-structured, or unstructured data. Structured data adheres to defined syntax, like SQL databases, while unstructured data lacks a specific skeletal definition (e.g., eBooks, webpages, videos). Semi-structured data, like JSON and XML, are flexible without a rigid structure.

Distinction from Data Retrieval (DR)

IR systems specialize in unstructured or semi-structured data, such as text documents, web pages, images, videos, and audio files, **while data retrieval** focuses on structured data stored in databases and tables.

IR systems employ probabilistic models to estimate the relevance of documents to a user's query, **while data retrieval** utilizes deterministic models based on relational algebra and calculus to provide exact matches.

IR systems allow users to interact with them using natural language queries, making them accessible to non-technical users, **while data retrieval** typically requires users to know SQL to formulate precise queries.

IR systems aim to find relevant results, even if they don't match the query terms, **while data retrieval** systems focus on finding exact matches for the specified query.

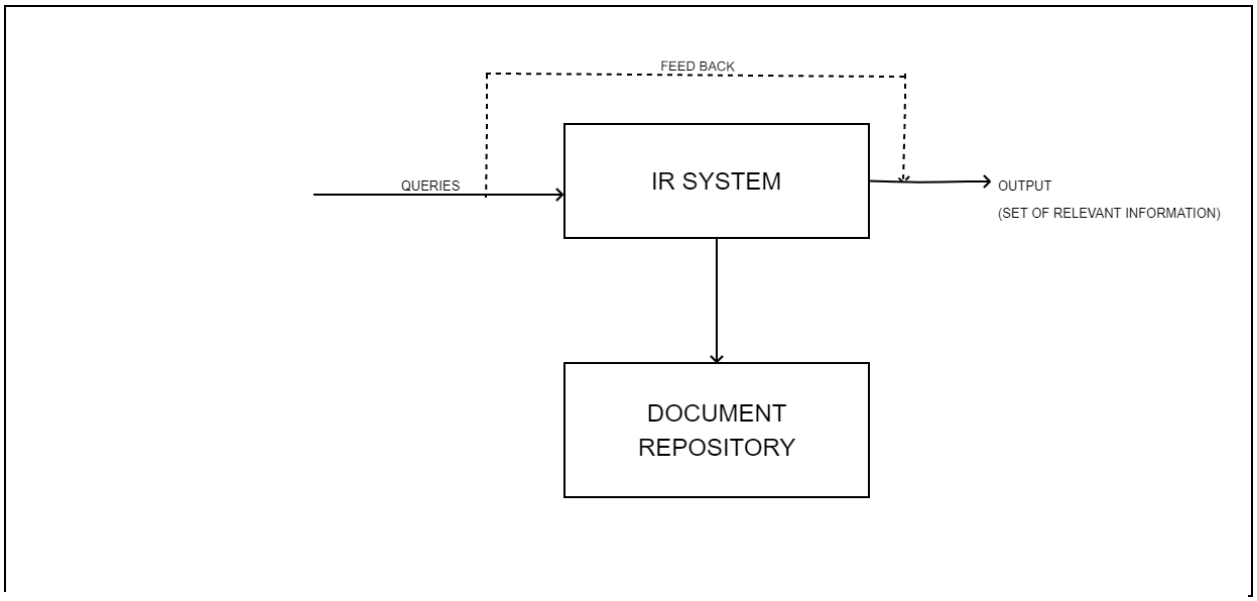

Figure 1: Block Diagram for IR system