

Information Retrieval

D. Thenmozhi
Associate Professor
SSNCE

Information Retrieval (IR)

- IR deals with the representation, storage, organization of, and access to information items
 - Types of information items: documents, Web pages, online catalogs, structured records, multimedia objects
- Early goals of the IR area: indexing text and searching for useful documents in a collection
- Nowadays, research in IR includes:
 - Modeling, Web search, text classification, systems architecture, user interfaces, data visualization, filtering and languages

The IR Problem

- Users of modern IR systems, such as search engine users, have information needs of varying complexity
- An example of complex information need is as follows:

Find all documents that address the role of the Federal Government in financing the operation of the National Railroad Transportation Corporation (AMTRAK)

The IR Problem

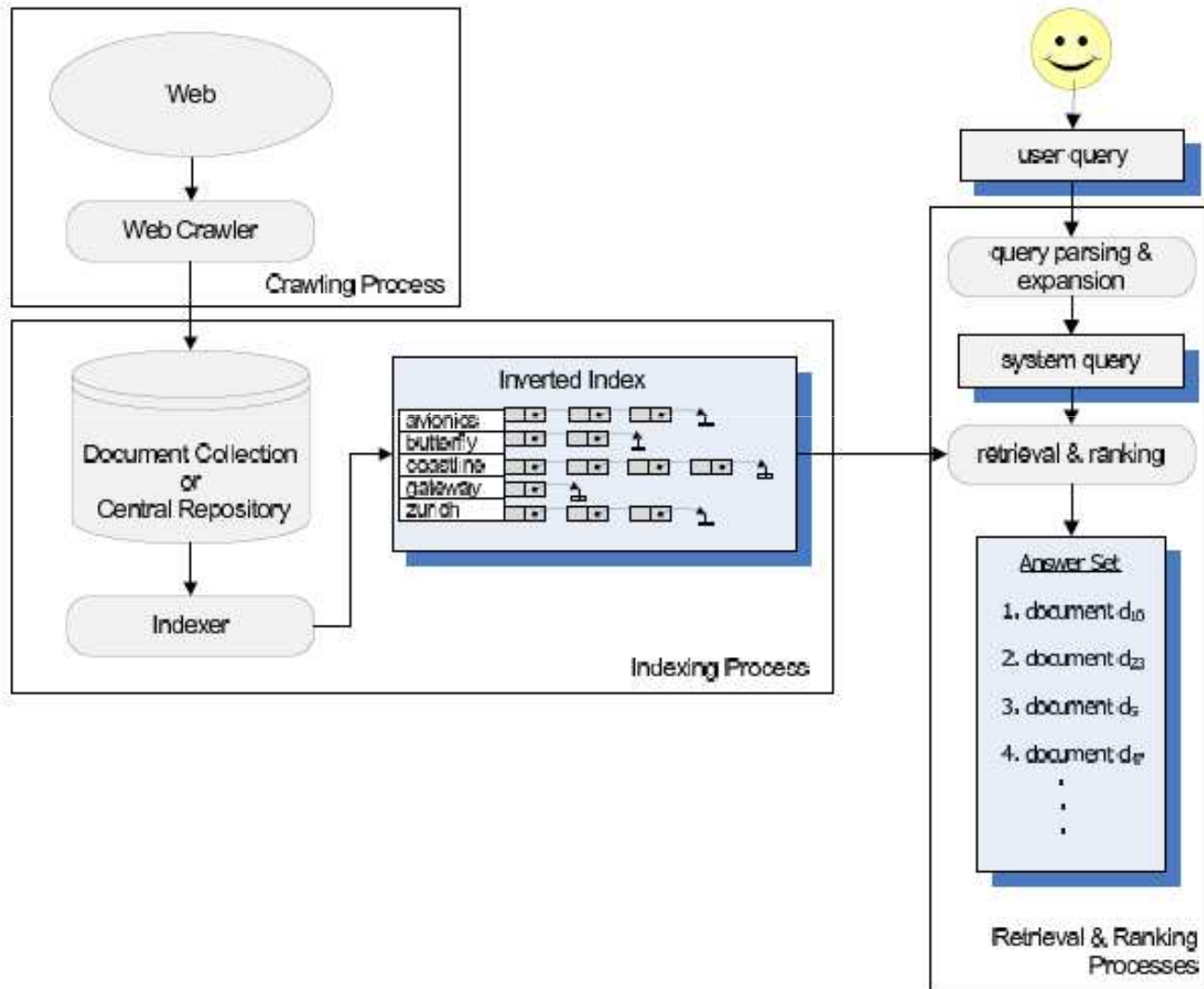
- This full description of the user information need is not necessarily a good query to be submitted to the IR system
- Instead, the user might want to first translate this information need into a query
- This translation process yields a set of *keywords, or index terms, which summarize the user information need*
- Given the user query, the key goal of the IR system is to retrieve information that is useful or relevant to the user

The IR Problem

- That is, the IR system must rank the information items according to a degree of relevance to the user query
- The IR Problem

The key goal of an IR system is to retrieve all the items that are relevant to a user query, while retrieving as few non-relevant items as possible
- The notion of relevance is of central importance in IR

Architecture of the IR System



Retrieval and Ranking Processes

