



SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE NAME : 19CS732 Information Retrieval Techniques

IV YEAR /VII SEMESTER

Unit 1- Introduction

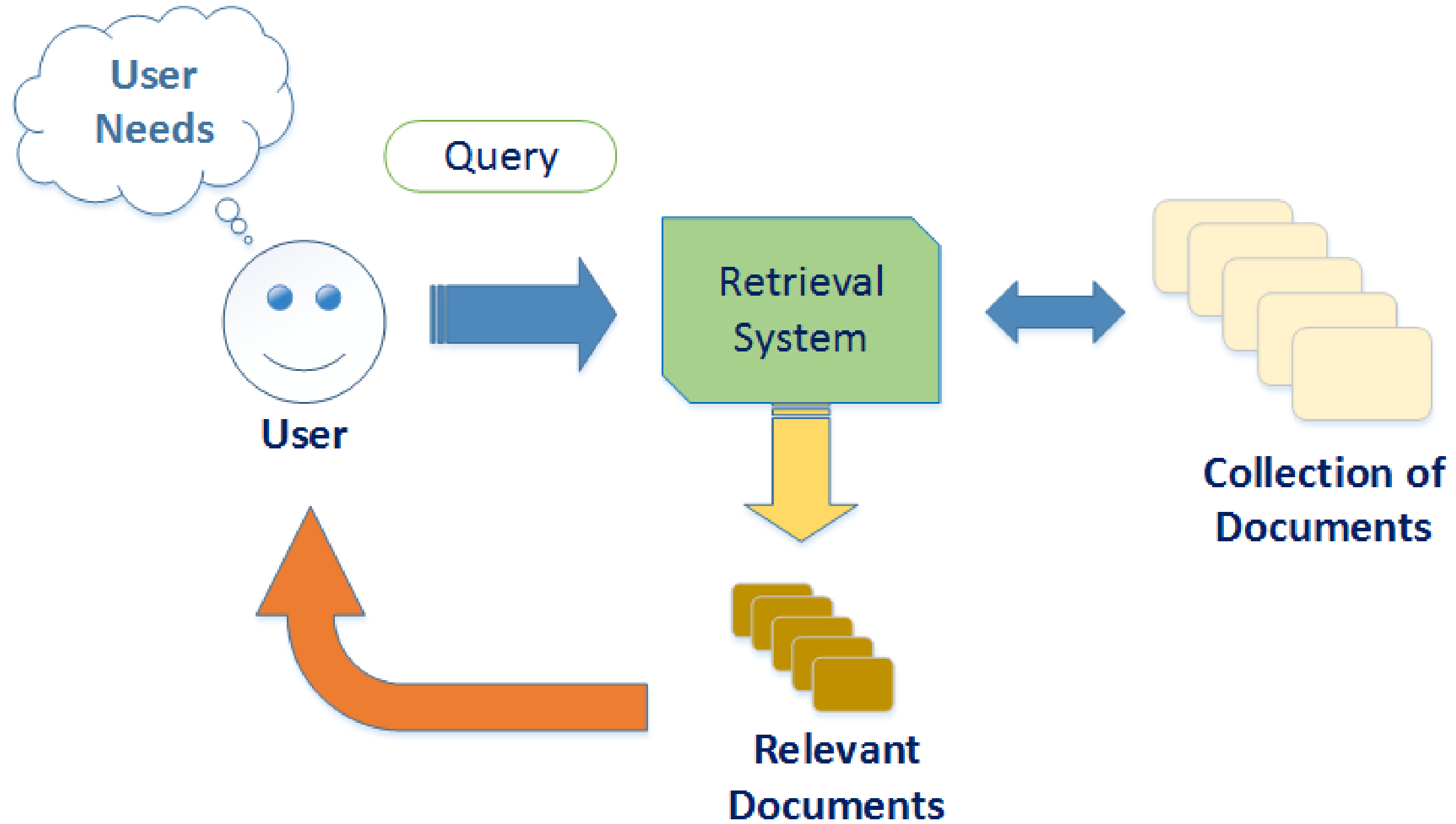
Topic 01 : Introduction





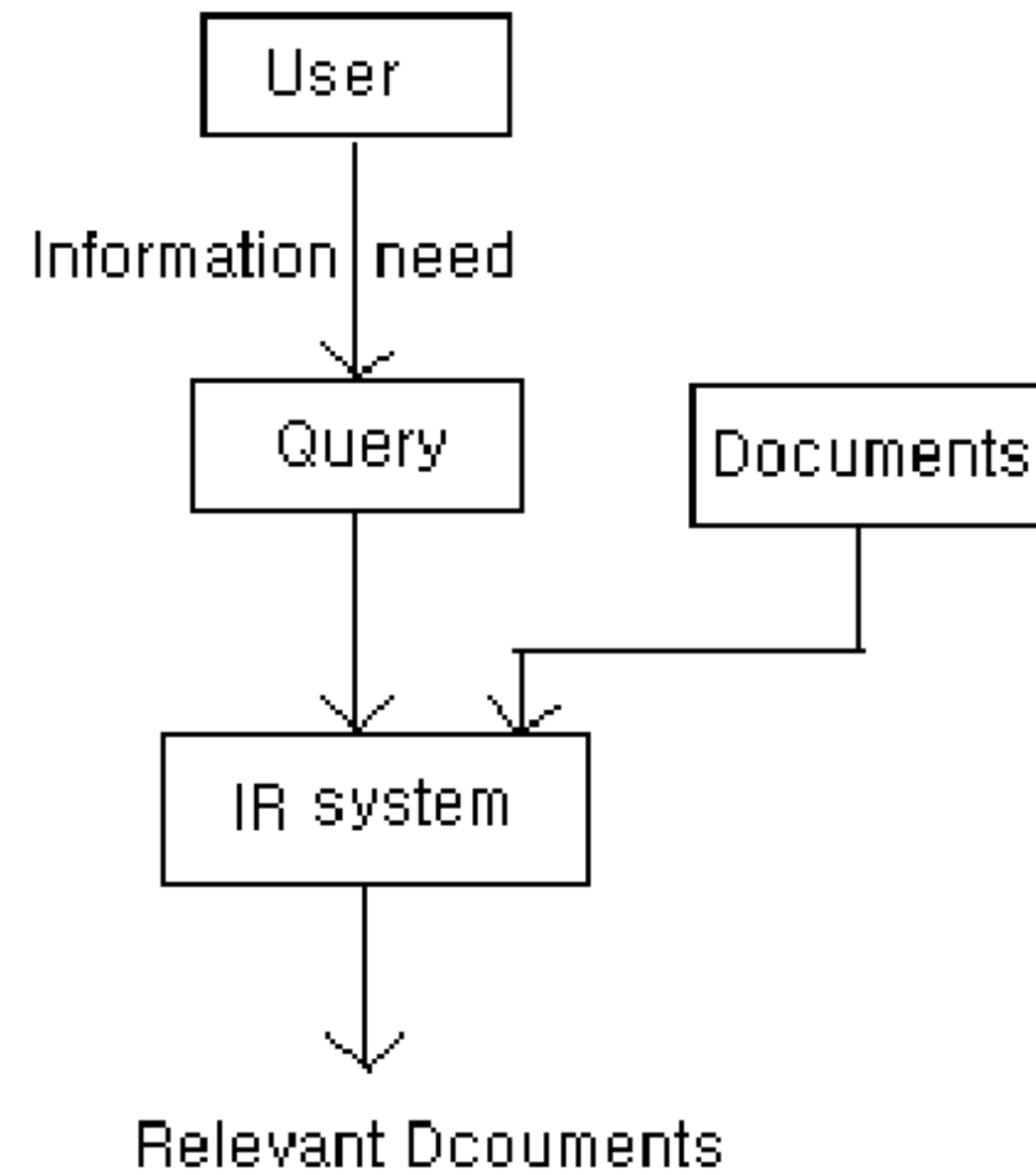
Information Retrieval







Do you remember this?





Meaning of Information Retrieval



- The ever-increasing amount of digitally available documents also brings with it the demand for a quick, targeted search. In the classical sense, this refers to the search for text documents.
- However, information must always be recovered for all multimedia documents.
- In addition to the main application of search engines, the information retrieval process includes also relevant for digital libraries, image databases or multimedia archives.



Type of search influences

- Database in which the search is made: big differences between self-managed database and database on the Internet
- Request for information: concrete vs. rather vague idea when searching Document type: Texts in various formats (e.g. doc, pdf, html file), videos, images, audio files



Type of search influences

- Another problem with the selection of the appropriate information is the uncertain knowledge of the information retrieval system, i.e. it has no knowledge of the document contents.
- The retrieval system can only use certain methods, e.g. Text statistics or term weighting, but has problems with certain word uses, e.g. with synonyms or homonyms.



Google Algorithm Updates and Terms for Analyzing the Search Engine's working principles:



- [Florida Update Effects for SEO](#)
- [Hilltop Update Effects for SEO](#)
- [Possum Update Effects for SEO](#)
- [Penguin Update Effects for SEO](#)
- [TrustRank Algorithm Effects on SEO](#)
- [RankBrain Algorithm and Working Principles](#)
- [Google Caffeine and Faster Indexing](#)
- [Google Humminbird Update and Better Search Intent Analysis](#)
- [Google Phantom Updates and Their Effects](#)
- [Google Dance Definition and Examples](#)



Origin of the Information Retrieval Term



- The term “information retrieval” was first used in 1950 by Calvin N. Mooers. Vannevar Bush described in the essay *As We May Think* in *Atlantic Monthly* in 1945 how the use of existing knowledge could be revolutionized through the use of knowledge stores.
- His vision was called Memex. This system should store all types of knowledge carriers and enable targeted searches and browsing for documents using links. Bush was already thinking about using search engines and retrieval tools.



Origin of the Information Retrieval Term



- The term “information retrieval” was first used in 1950 by Calvin N. Mooers. Vannevar Bush described in the essay *As We May Think* in *Atlantic Monthly* in 1945 how the use of existing knowledge could be revolutionized through the use of knowledge stores.
- His vision was called Memex. This system should store all types of knowledge carriers and enable targeted searches and browsing for documents using links. Bush was already thinking about using search engines and retrieval tools.



Information Retrieval Models



➤ Boolean Model

- Oldest information retrieval model based on Boolean logic from 1854
- Contents can only be found using the operators “and”, “or”, “not”
- The content is not sorted – there is no ranking of the results.



Information Retrieval Models



➤ **Ontological Model**

- It is not based on the evaluation of the document content, but on the evaluation of the link structure between documents – this results in a ranking of the documents
- The structure allows a statement on the authority of documents
- This includes, for example, the PageRank from Google, developed by Larry Page and Sergey Brin



Information Retrieval Models



➤ Text Statistics

- Examining terms within a document:
- Weighting is done via WDF and IDF
- WDF: Within Document Frequency – relative frequency of a term within a document
- IDF: Inverse Document Frequency – Frequency with which a document occurs in a database with a specific term
- The vector model is also part of the text statistics model: each text corresponds to a point in space, the angles of the vectors indicate the similarity of the words to each other.



Information Retrieval Models



➤ Cluster model

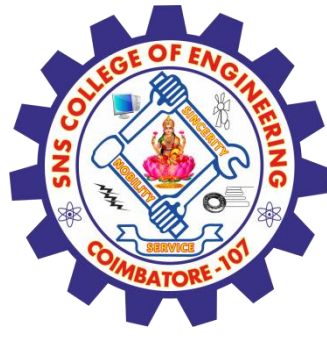
- Summary of documents by Similarity can speed up the search process since only access to a document pool is required
- Problems can arise if the clusters are incomplete or very large



How Do Search Engines Use Information Retrieval?



- Every internet search engine uses information retrieval to process search queries. With search engines, it is important to evaluate the “determined” information and sort it according to importance/relevance – this results in the ranking. As soon as you enter a search term in the search field, the search engine returns relevant information about your search term from the stored data (SERP).
- Accordingly, SEO tries to improve the recovery of information from the optimized page – one of the measures is, for example, the $WDF * IDF$ optimization of websites.



An Example of Information Retrieval System Process



➤ <https://www.holisticseo.digital/theoretical-seo/information-retrieval/>



REFERENCES



1. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, –Modern Information Retrieval: The Concepts and Technology behind Search, Second Edition, ACM Press Books, 2011.
2. Ricci, F, Rokach, L. Shapira, B.Kantor, –Recommender Systems Handbook||, First Edition, 2011.

THANK YOU