



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

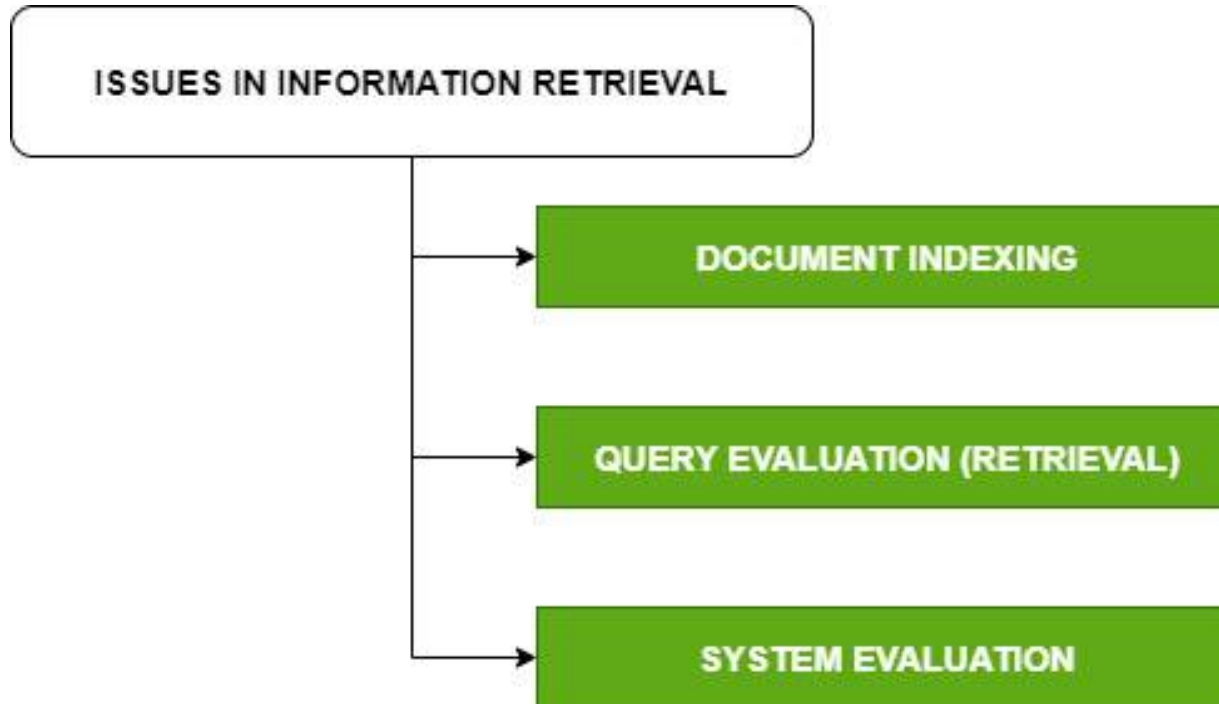
IVYEAR / VIII SEMESTER

Unit 1- INTRODUCTION

Topic 1 : Information Retrieval Techniques



Problem





19CS732 Information Retrieval Techniques



Syllabus

UNIT I INTRODUCTION

9

Information Retrieval – Early Developments – The IR Problem – The User’s Task – Information versus Data Retrieval - The IR System – The Software Architecture of the IR System – The Retrieval and Ranking Processes - The Web – The e-Publishing Era – How the web changed Search – Practical Issues on the Web – How People Search – Search Interfaces Today – Visualization in Search Interfaces

UNIT II MODELING AND RETRIEVAL EVALUATION

9

Basic IR Models - Boolean Model - TF-IDF (Term Frequency/Inverse Document Frequency) Weighting - Vector Model – Probabilistic Model – Latent Semantic Indexing Model – Neural Network Model – Retrieval Evaluation – Retrieval Metrics – Precision and Recall – Reference Collection – User-based Evaluation – Relevance Feedback and Query Expansion – Explicit Relevance Feedback.



19CS732 Information Retrieval Techniques



UNIT III TEXT CLASSIFICATION AND CLUSTERING

9

A Characterization of Text Classification – Unsupervised Algorithms: Clustering – Naive Text Classification – Supervised Algorithms – Decision Tree – k-NN Classifier – SVM Classifier – Feature Selection or Dimensionality Reduction – Evaluation metrics – Accuracy and Error – Organizing the classes – Indexing and Searching – Inverted Indexes – Sequential Searching – Multi-dimensional Indexing

UNIT IV WEB RETRIEVAL AND WEB CRAWLING

9

The Web – Search Engine Architectures – Cluster based Architecture – Distributed Architectures – Search Engine Ranking – Link based Ranking – Simple Ranking Functions – Learning to Rank – Evaluations -- Search Engine Ranking – Search Engine User Interaction – Browsing – Applications of a Web Crawler – Taxonomy – Architecture and Implementation – Scheduling Algorithms – Evaluation



19CS732 Information Retrieval Techniques



UNIT V IR APPLICATIONS

9

IR applications-Information Extraction-Question answering- Opinion summarization- General Applications-Digital libraries-Information filtering-Media search-Search engines—Domain specific applications-Expert search finding-Geographic information retrieval- Case study on information retrieval methods in Text Minings



19CS732 Information Retrieval Techniques



TEXT BOOKS:

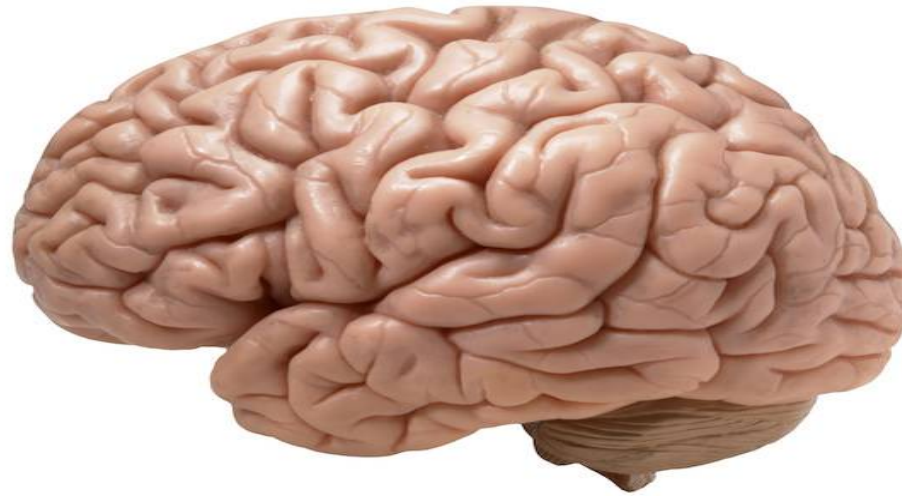
1. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, —Modern Information Retrieval: The Concepts and Technology behind Search, Second Edition, ACM Press Books, 2011.
2. Ricci, F, Rokach, L. Shapira, B.Kantor, —Recommender Systems Handbook||, First Edition, 2011.

REFERENCES:

1. C. Manning, P. Raghavan, and H. Schütze, —Introduction to Information Retrieval, Cambridge University Press, 2008.
2. Stefan Buettcher, Charles L. A. Clarke and Gordon V. Cormack, —Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press, 2010.

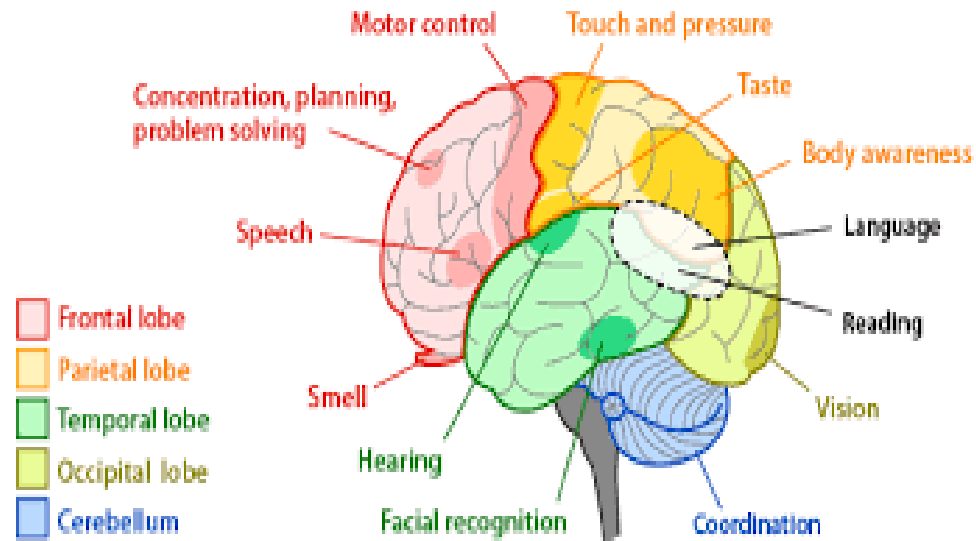


HUMAN BRAIN





Information Retrieval Techniques





Information Retrieval Techniques



Agenda

- ✓ Information Retrieval Techniques
- ✓ Components of IRT
- ✓ The Stages of IRT
- ✓ The Formalized IRT Process
- ✓ Applications



Information Retrieval Techniques



What is Information Retrieval Techniques?

Information retrieval (IR) is the activity of obtaining [information](#) resources relevant to an information need from a collection of information resources.

(OR)

The process of actively seeking out information relevant to a topic of interest.

(OR)

Information Retrieval is finding material of an unstructured nature that satisfies an information need from within large collections.



Information Retrieval Techniques –Cont..



- It is thus concerned with the collection, representation, storage,
 - Organization, accessing, manipulation and display of the information items necessary to satisfying those needs



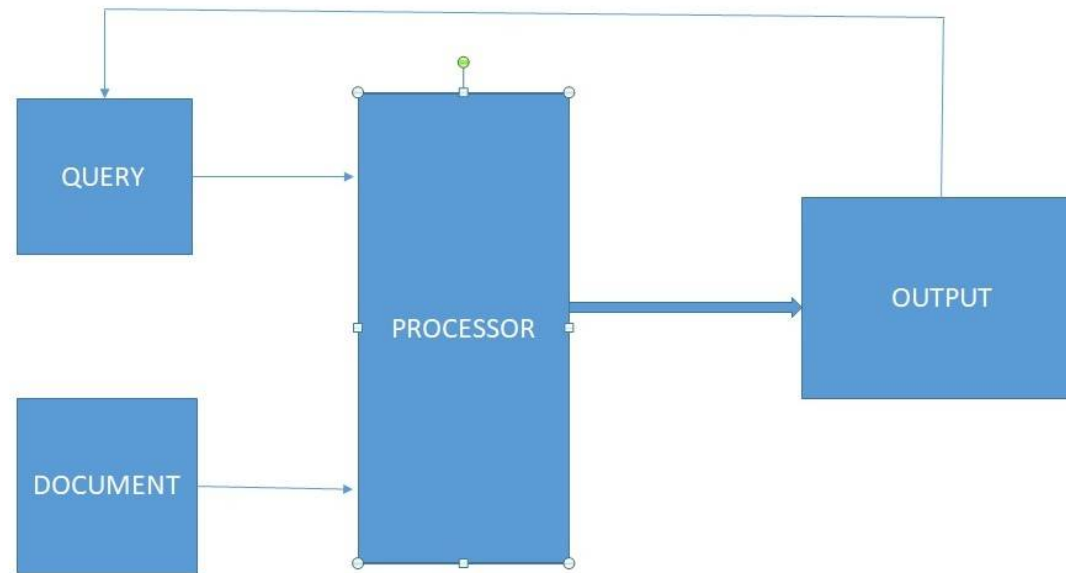
Information Retrieval Techniques –Cont..



- 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



COMPONENTS OF IR





COMPONENTS OF IR



Input – Store Only a representation of the document

A document representative – Could be list of extracted words considered to be significant.

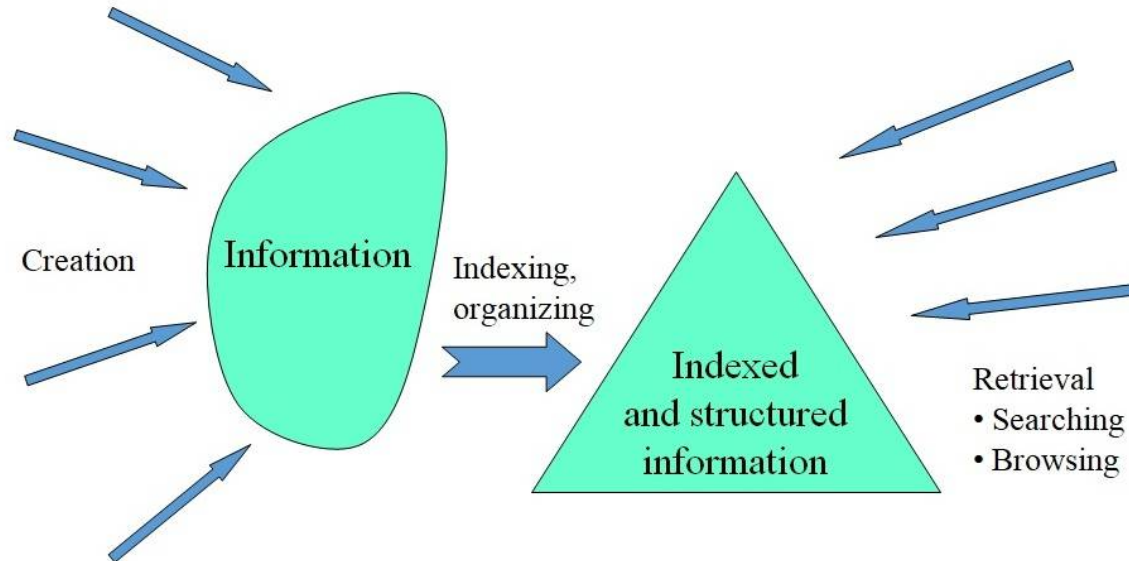
Processor – Involve in performance of actual retrieval function

Feedback – Improve

Output – A set document numbers

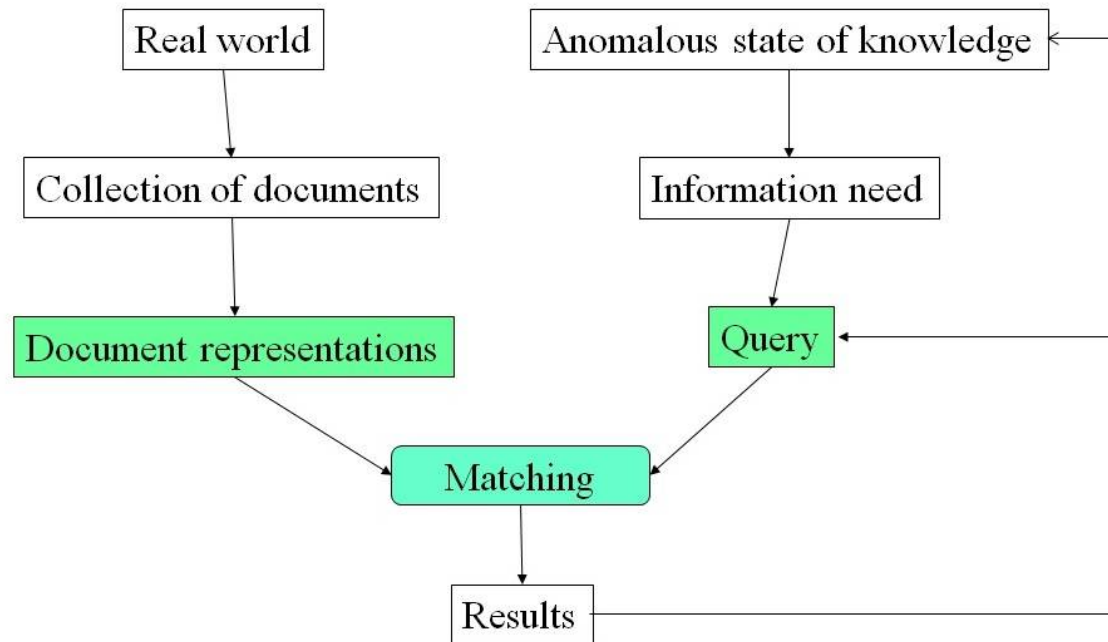


THE STAGES OF IR





THE FORMALIZED IR PROCESS





APPLICATIONS



- Text Search
- Ad search
- Image/Video search
- Email Search
- Question Answering systems
- Recommender systems
- Desktop Search
- Expert Finding
- Jobs
- Prizes
- Products
- News
- Source code
- Videogames
- Maps
- Partners



Activity



Disadvantages



- Primary cost
- Lack of Budget
- IT knowledge
- Lack of training facility
- Electric problem



Advantages



- Time saves
- Easy to understand
- Current information
- Database
- Multi database search
- Multiple concepts



Assessment 1



1. List out the Advantages of IRT

- a) _____
- b) _____
- c) _____
- d) _____

2. Identify the Applications of IRT

- a) _____
- b) _____
- c) _____
- d) _____





TEXT BOOKS:

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.

REFERENCES:

1. C. Manning, P. Raghavan, and H. Schütze, –Introduction to Information Retrieval, Cambridge University Press, 2008.
2. Stefan Buettcher, Charles L. A. Clarke and Gordon V. Cormack, –Information Retrieval: Implementing and Evaluating Search Engines, The MIT Press, 2010.

THANK YOU