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 3 : Information versus Data Retrieval ,The IR System and
The Software Architecture of the IR System

Query Evaluation

In the retrieval model how can a document be represented with the selected keywords and how are documents and query representations compared to calculate a score. Information Retrieval (IR) deals with issues like uncertainty and vagueness in information systems.

Uncertainty :

The available representation does not typically reflect true semantics of objects such as images, videos etc.

Vagueness   :

The information that the user requires lacks clarity, is only vaguely expressed in a query, feedback or user action.
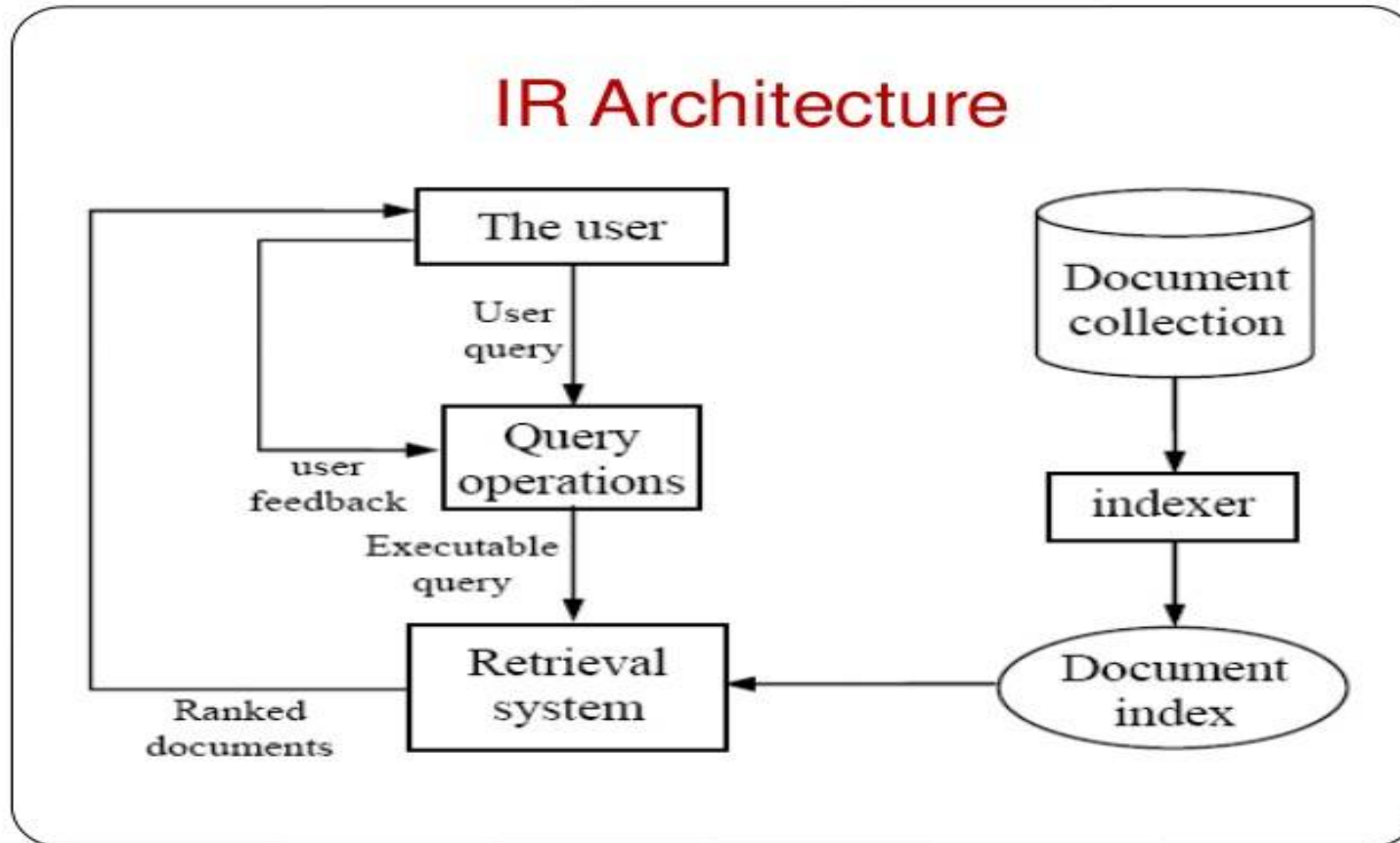
Information versus Data Retrieval

IR Vs. DR

	Data Retrieval (DR)	Information Retrieval (IR)
Matching	Exact match	Partial match, best match
Inference	Deduction	Induction
Model	Deterministic	Probabilistic
Classification	Monothetic	Polythetic
Data	Database tables, structured	Free text, unstructured
Query language	Artificial, SQL, relational algebras.	Natural, Keywords, free text
Query specification	Complete	Incomplete
Items wanted	Matching	Relevant

The Software Architecture of the IR System

IR Architecture

**Unit-1/Introduction/19CS732 Information Retrieval Techniques
/K.Karthikeyan/CSE/SNSCE**

The Software Architecture of the IR System –Cont..

- IR systems include two types of terms

- Objective term

- Nonobjective term

- Objective term

    Objective term are extrinsic to semantic content, and there is generally no

disagreement about how to assign them.

Ex : include author name , document URL and date of publication.

- Nonobjective term

    Nonobjective term are intended to reflect the information manifested in the

documents , and there is no agreement about the choice or degree of applicability of
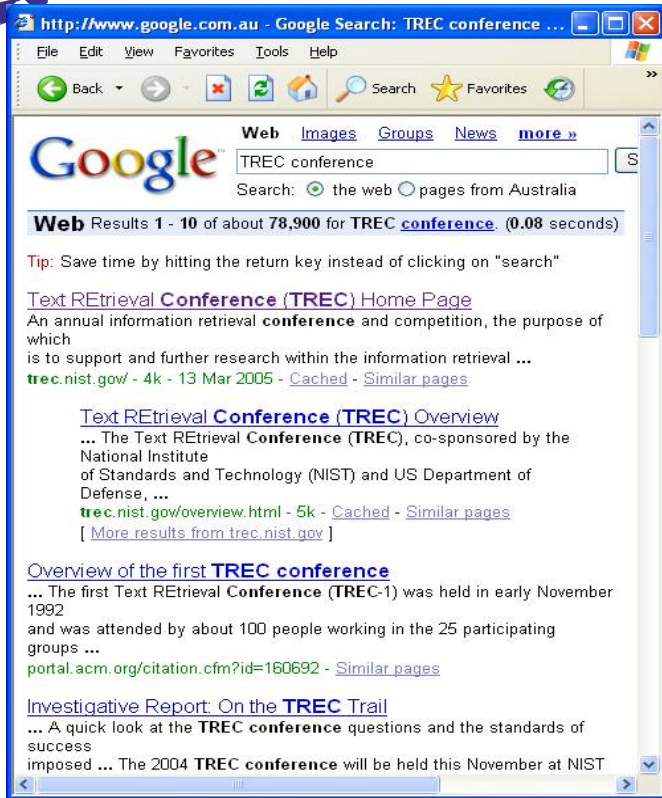
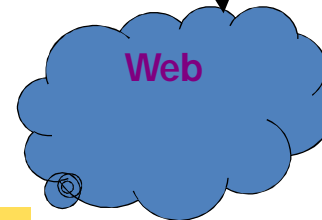these terms.

Early Developments –Cont..

➢ IR queries are as follows

➢ Keyword queries

➢ Boolean queries

➢ Phrase queries

➢ Full document queries
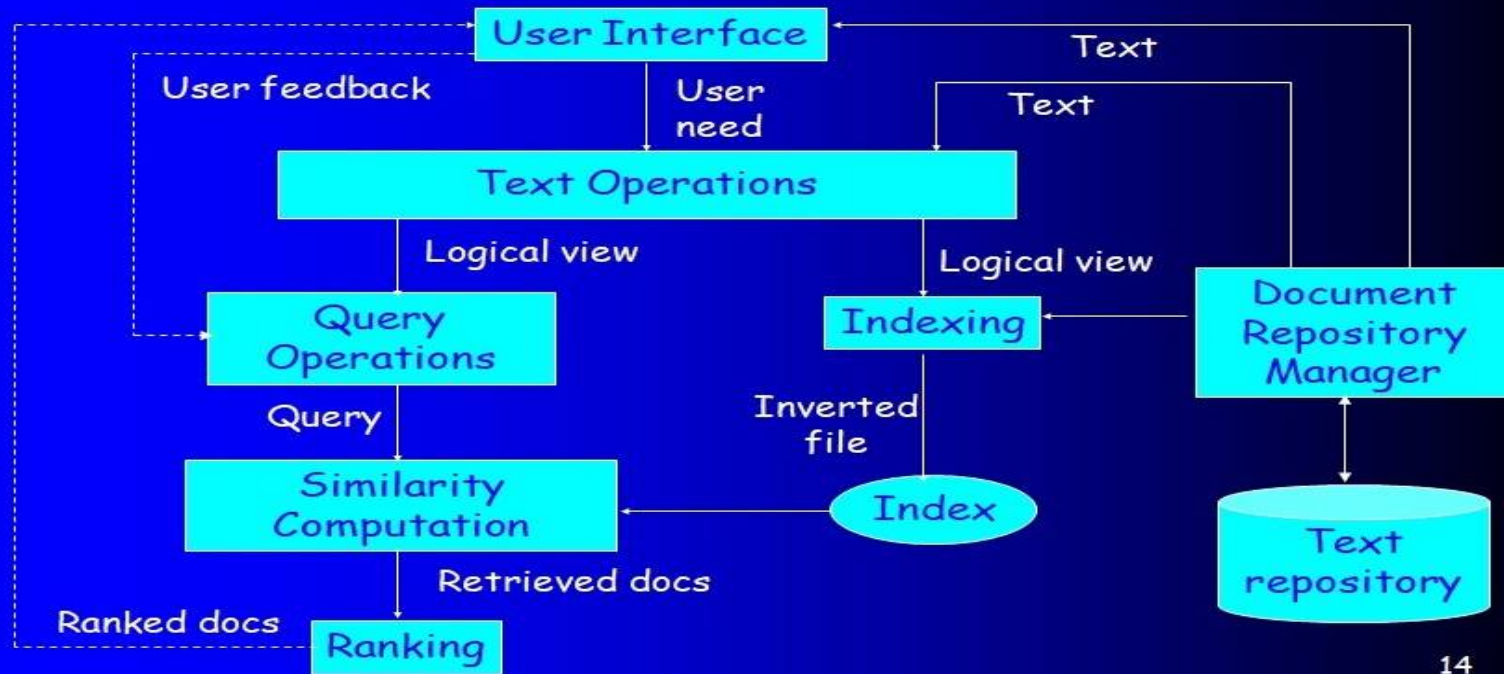
➢ Natural language questions

Examples

Unit-1/Introduction/19CS732 Information Retrieval Techniques /K.Karthikeyan/CSE/SNSCE

Process of retrieving info

Unit-1/Introduction/19CS732 Information Retrieval Techniques
/K.Karthikeyan/CSE/SNSCE

Activity

Advantages and Disadvantages

ADVANTAGES & DISADVANTAGES

- Advantages
  - Results are predictable, relatively easy to explain
  - Many different features can be incorporated
  - Efficient processing since many documents can be eliminated from search
- Disadvantages
  - Effectiveness depends entirely on user
  - Simple queries usually don't work well
  - Complex queries are difficult.

Assessment 1

1. List out the Advantages of architecture in IRT

a)_____

b)_____

c)_____

d)_____

2. Identify the Applications of architecture in 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