CS60092: Information Retrieval

Text Books

- Introduction to Information Retrieval C D Manning, P Raghavan and H Schutze, Cambridge University Press

- Information Retrieval *Implementing and Evaluating Search Engines* S Buttcher, C L. A. Clarke and G V. Cormack, MIT Press, 2010.
- Mining the Web, Soumen Charabarti, Morgan-Kaufmann, 2002

Course Webpage and Moodle

- https://cse.iitkgp.ac.in/~pabitra/course/ir/ir.html

- https://moodlecse.iitkgp.ac.in/moodle/course/view.php?id=149
- Joining Student Keys: IR@STUD$

Evaluation Scheme

- Endsem: 30
- Midsem: 30
- Programming Assignments: 3 x10
- Attendance: 10

Course Outline

- Introduction
- Boolean Retrieval
- Vector Space Model
- Indexing and Preprocessing
- Relevance Feedback
- Probabilistic Information Retrieval
- Language Models
- Text Classification and Learning to Rank
- Link Analysis
- Neural Information Retrieval

Problem Definition

- Given a user's information need, find documents satisfying that need (often to complete a task).

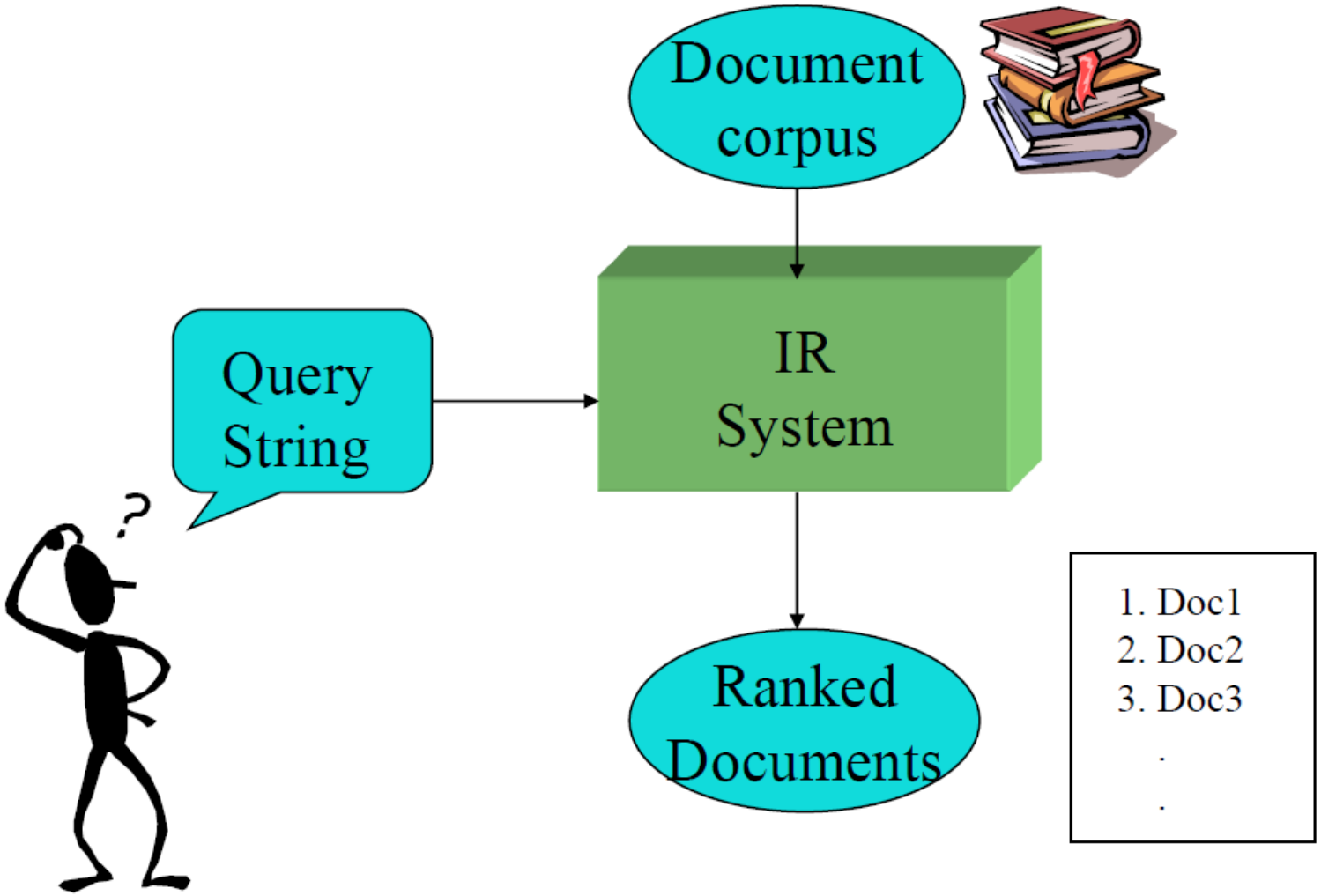
- Information need: what user is looking for
- Query: actual representation of above
- Document: any unit / item that can be retrieved

- For this course, we will only consider textual information
- (no images/graphics, maps, speech, video, etc.).

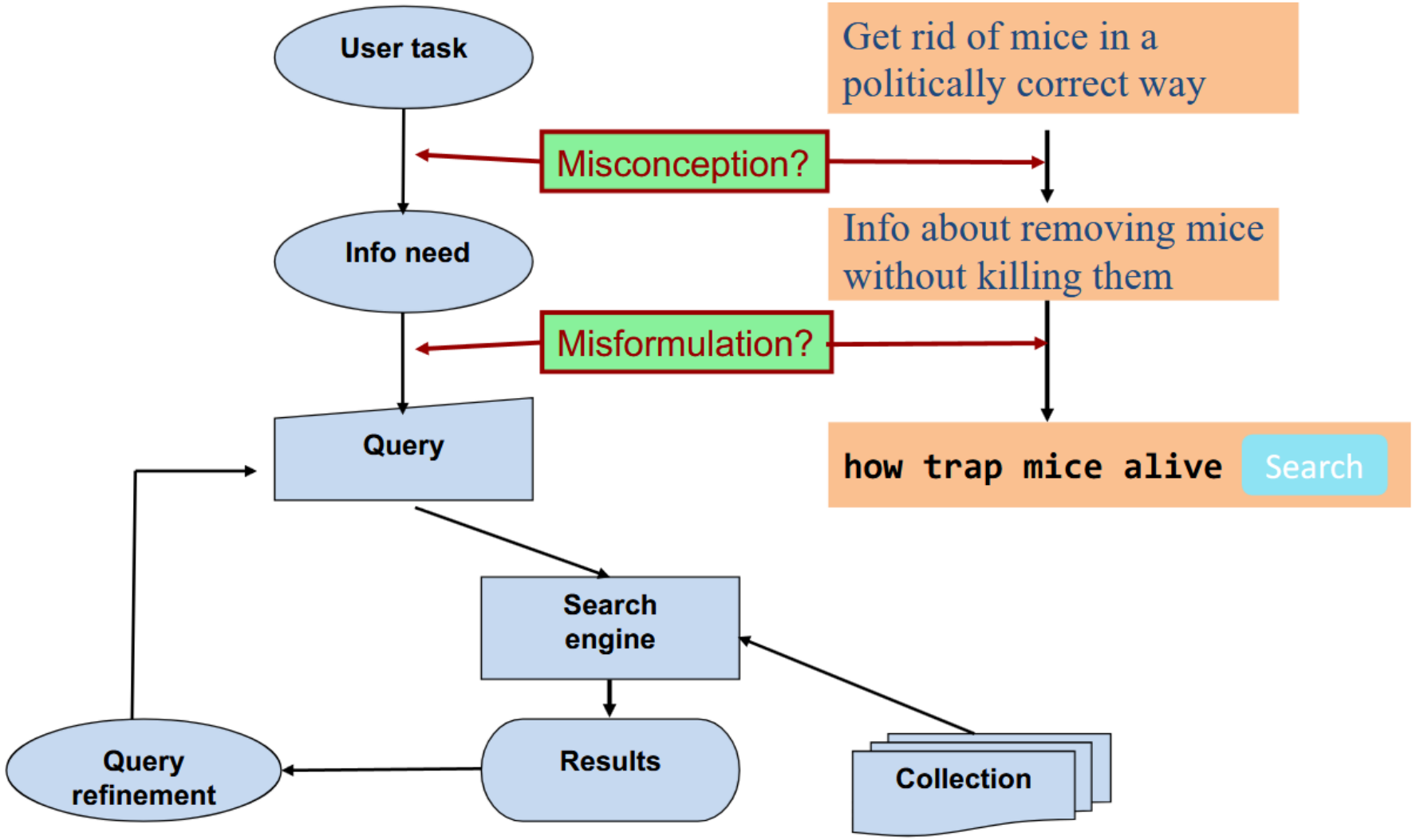
Information Retrieval

- Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).
- Example: Web Search
- But there are many other cases:
  - E-mail search
  - Searching your laptop
  - Corporate knowledge bases
  - Legal information retrieval

Information Retrieval System

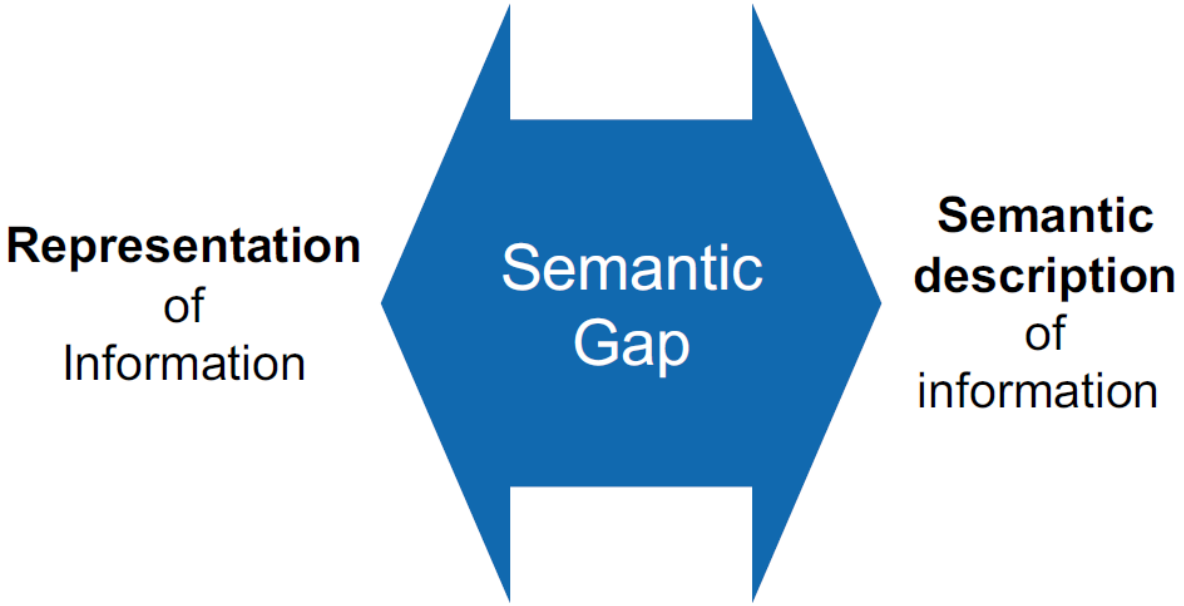


Search Model

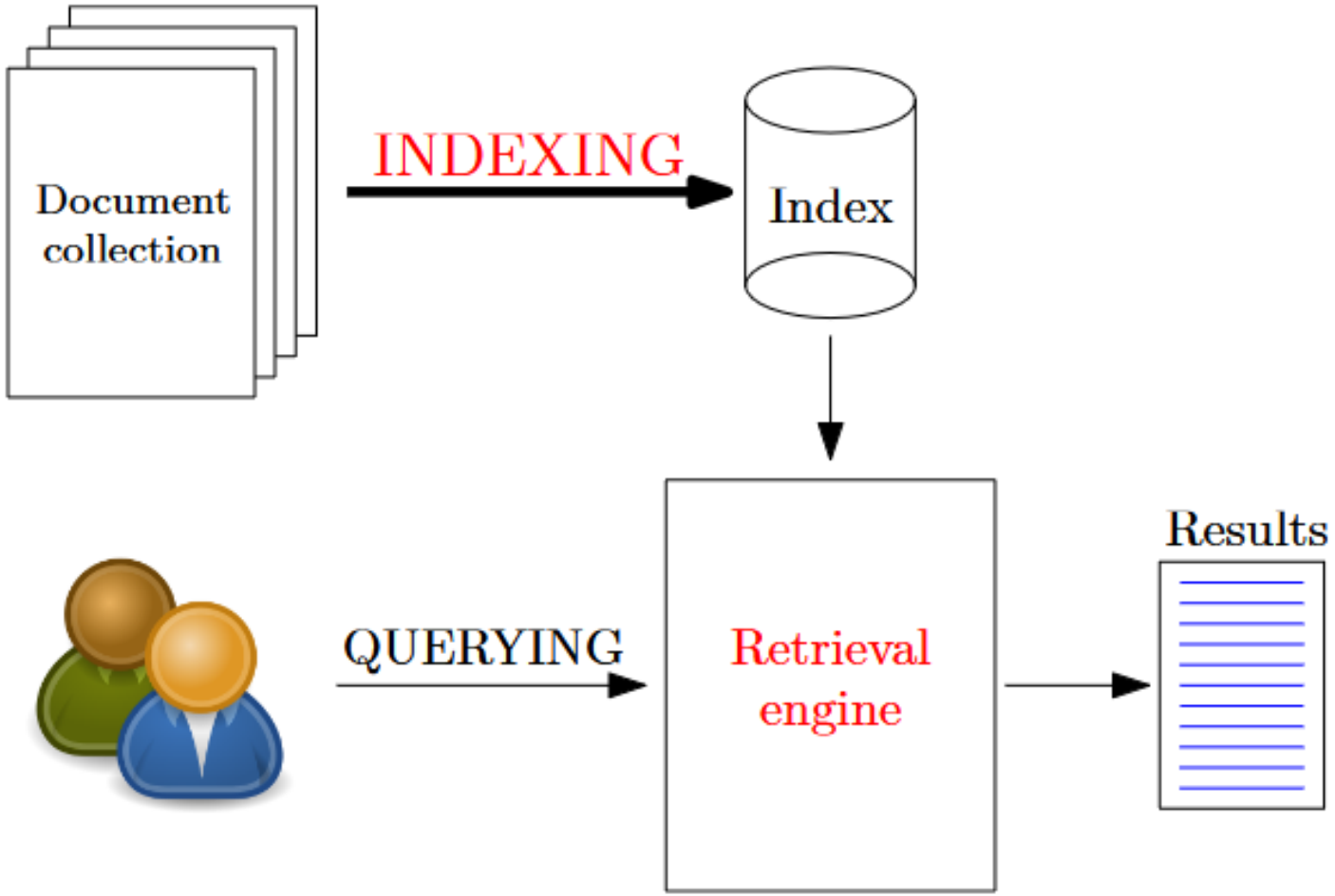


Unstructured Data

- No explicit structure but implicit features (characters, grammar, pixels, waves, etc)



IR Systems


Document collection
INDEXING
Index
QUERYING
Retrieval engine
Results

Steps

1. Document acquisition: how is the document collection obtained/constructed?
2. Indexing: representing documents so that retrieval is easy
3. Retrieval: matching the user query against documents in the collection
4. Evaluation: how to determine whether the system did well?

IR Tasks

- Adhoc retrieval (keyword based retrieval)
- Text categorization and clustering
- Information recommendation
- Information extraction
- Question answering
- Multimedia search