Information Storage and Retrieval (IS 313)

Lecture 1

Introduction to the Course

Dr. Ahmed Elsayed

2023 - 2024

Course Info.

- **Lectures:** Thursday (12-2, 2-4)
- **Instructor:** Dr. Ahmed Elsayed
  – Office hours: Wednesday 8:00 - 10:00
  – Contact: ahmedYakoup@fci.helwan.edu.eg
- **Assistant Instructor/TA: ??**
- **Textbook:**
  – Christopher D. M., Prabhakar R. and Hinrich S. **An Introduction to Information Retrieval.** Cambridge University Press Cambridge, England

Assessment Scheme

- Mid-Term Exam: 20 % [30 %]
- Project: 20 %
- Final Exam: 60 %   [50 %]
- **Total** :   **100 %**

General Rules

Introduction to
Information Retrieval

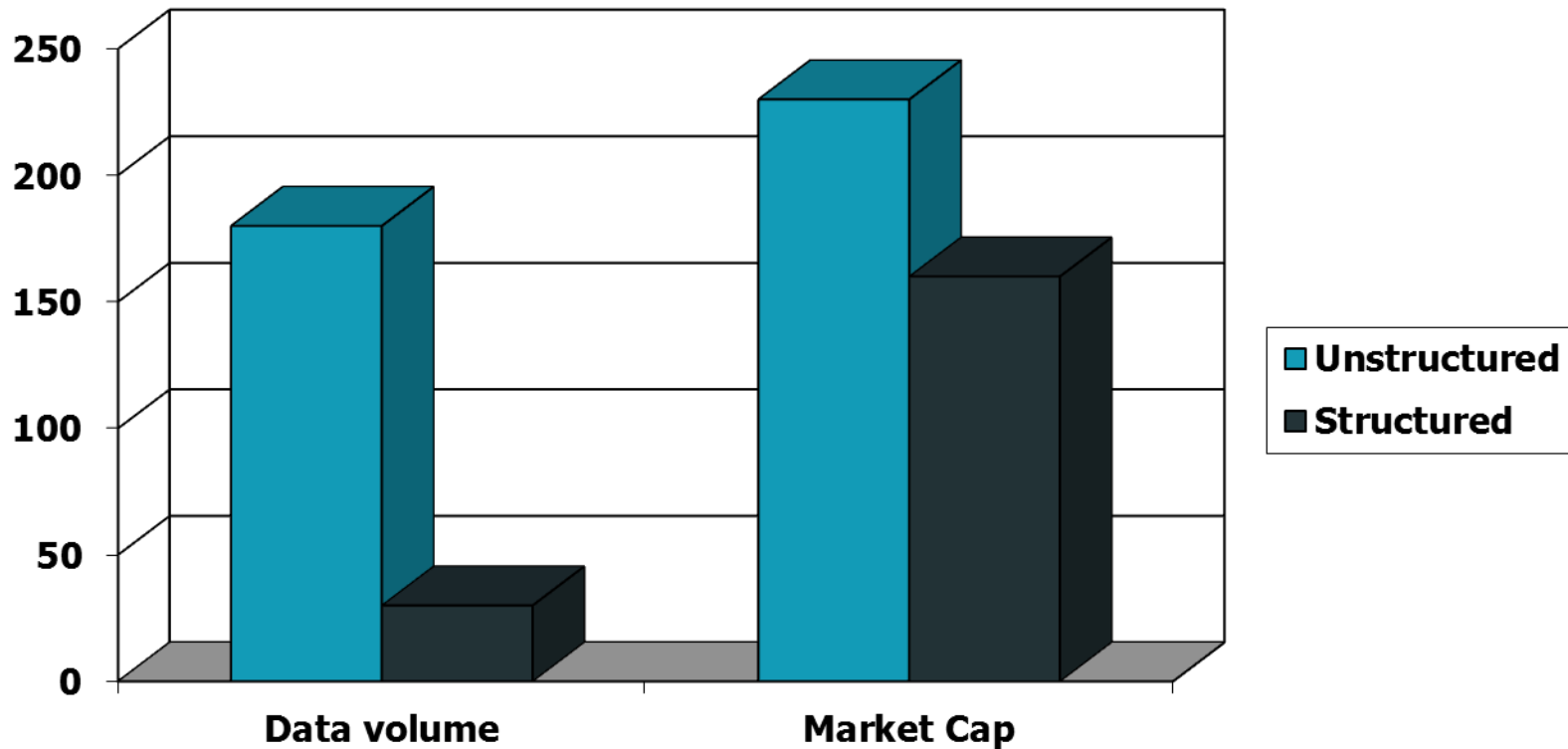## Introducing Information Retrieval
and Web Search

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).

    - These days we frequently think first of web search, but there are many other cases:
        - E-mail search
        - Searching your laptop
        - Corporate knowledge bases
        - Legal information retrieval

# Unstructured (text) vs. structured (database) data in the mid-nineties
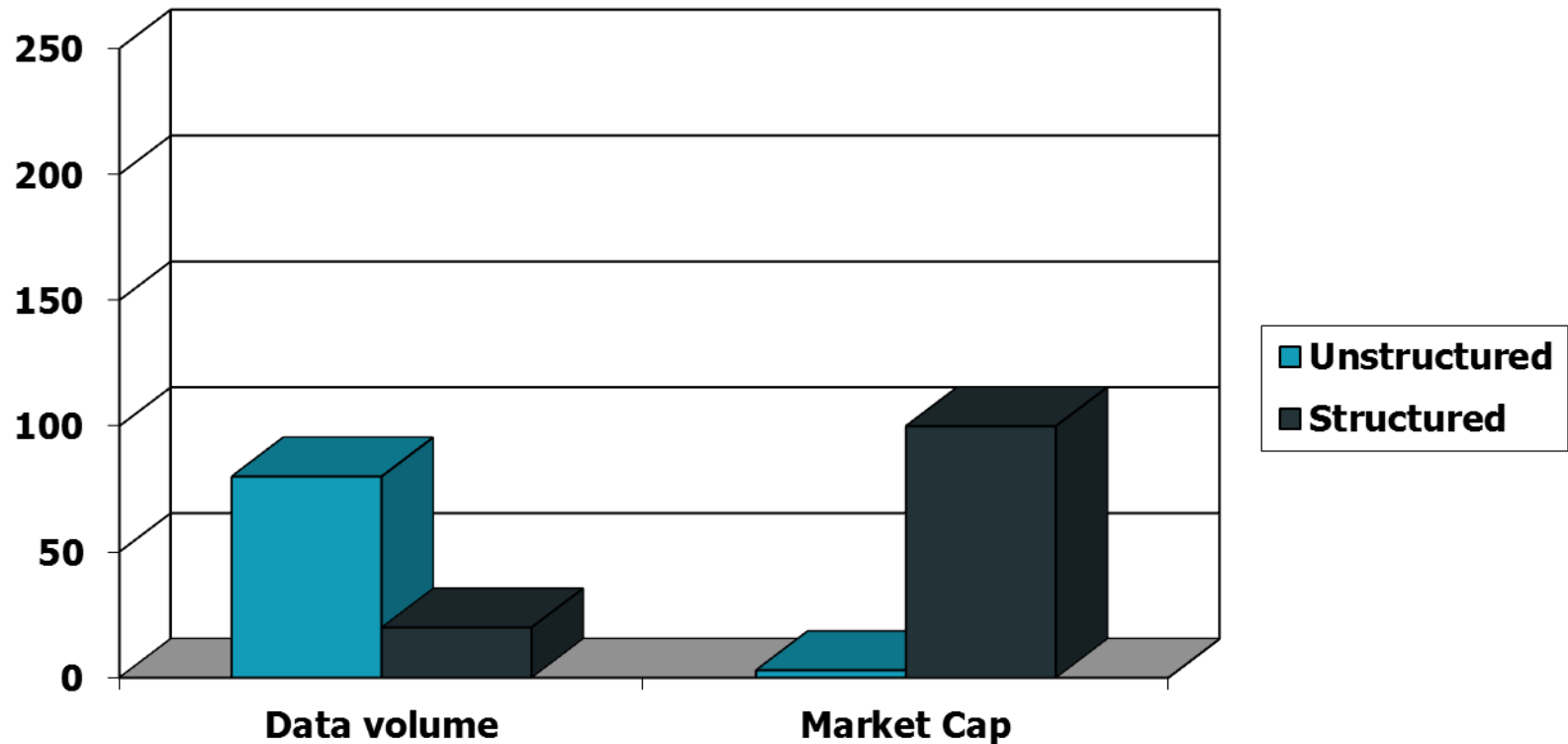
Unstructured (text) vs. structured (database) data today

Basic assumptions of Information Retrieval

- Collection: A set of documents
    - Assume it is a static collection for the moment

- Goal: Retrieve documents with information that is relevant to the user's information need and helps the user complete a task

The classic search model
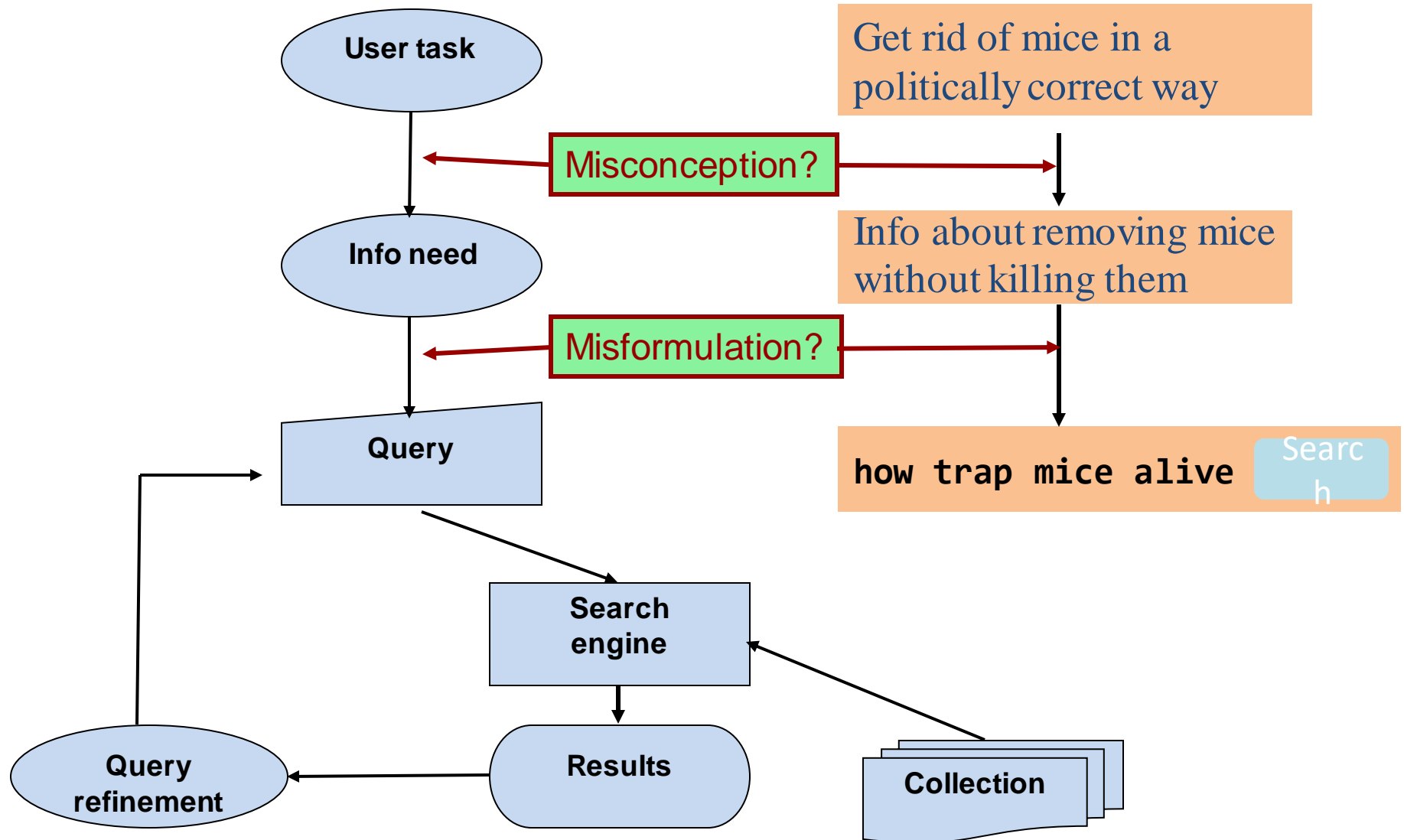
How good are the retrieved docs? Through efficient & effectiveness efficient--> time and space complexity effectiveness--> quality of result

- *Precision* : Fraction of retrieved docs that are relevant to the user's information need

- *Recall* : Fraction of relevant docs in collection that are retrieved

  - More precise definitions and measurements to follow later

Course syllabus

	Topic
1	**Introduction to the Course**
2	**Boolean Retrieval Model**
3	**The Term Vocabulary and Postings Lists**
4	**Scoring and Term Weighting**
5	**Vector Space Model**
6	**Evaluation in Information Retrieval**
7	**Index Construction**
8	**Dictionaries and Tolerant Retrieval**