

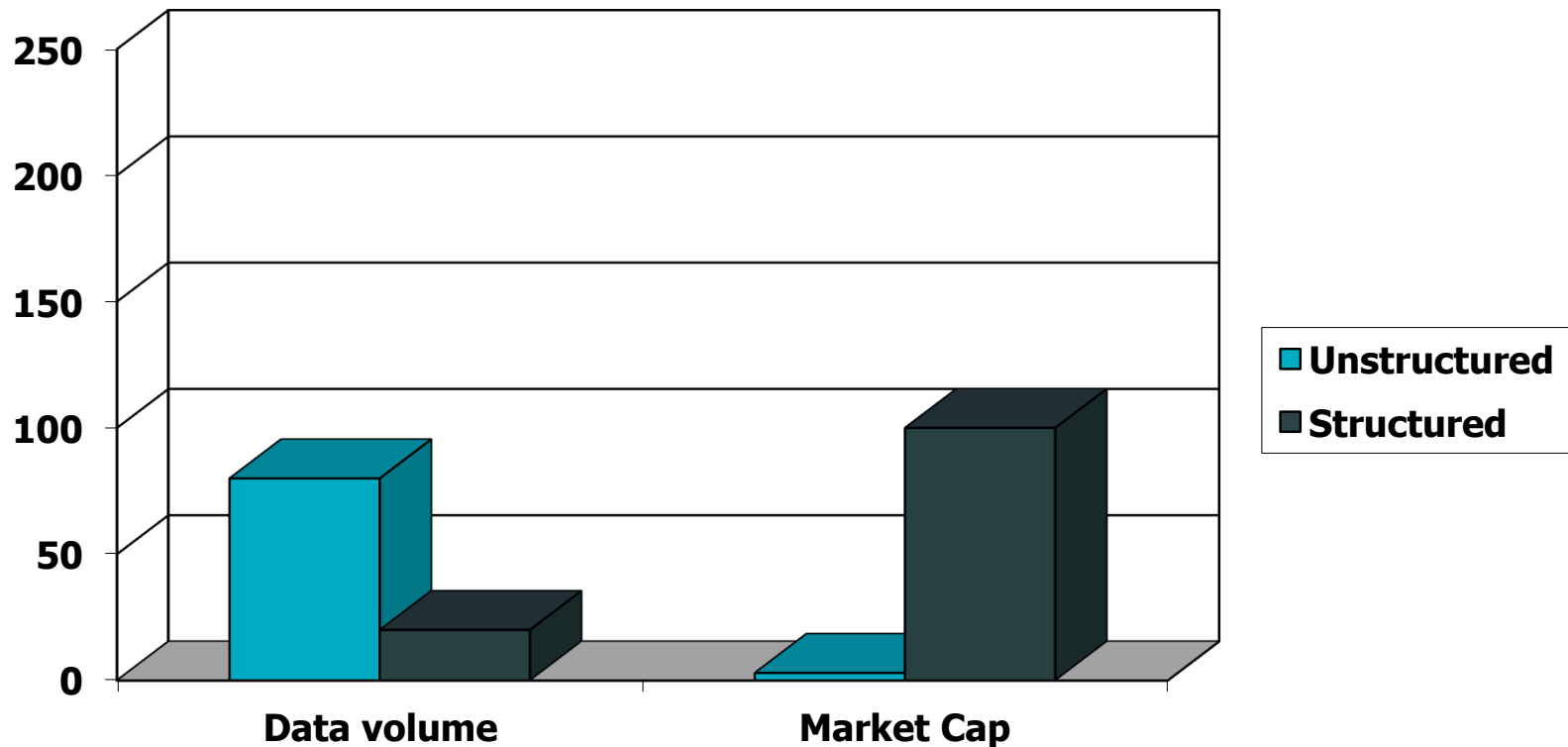
Introduction to **Information Retrieval**

Introducing Information Retrieval
and Web Search

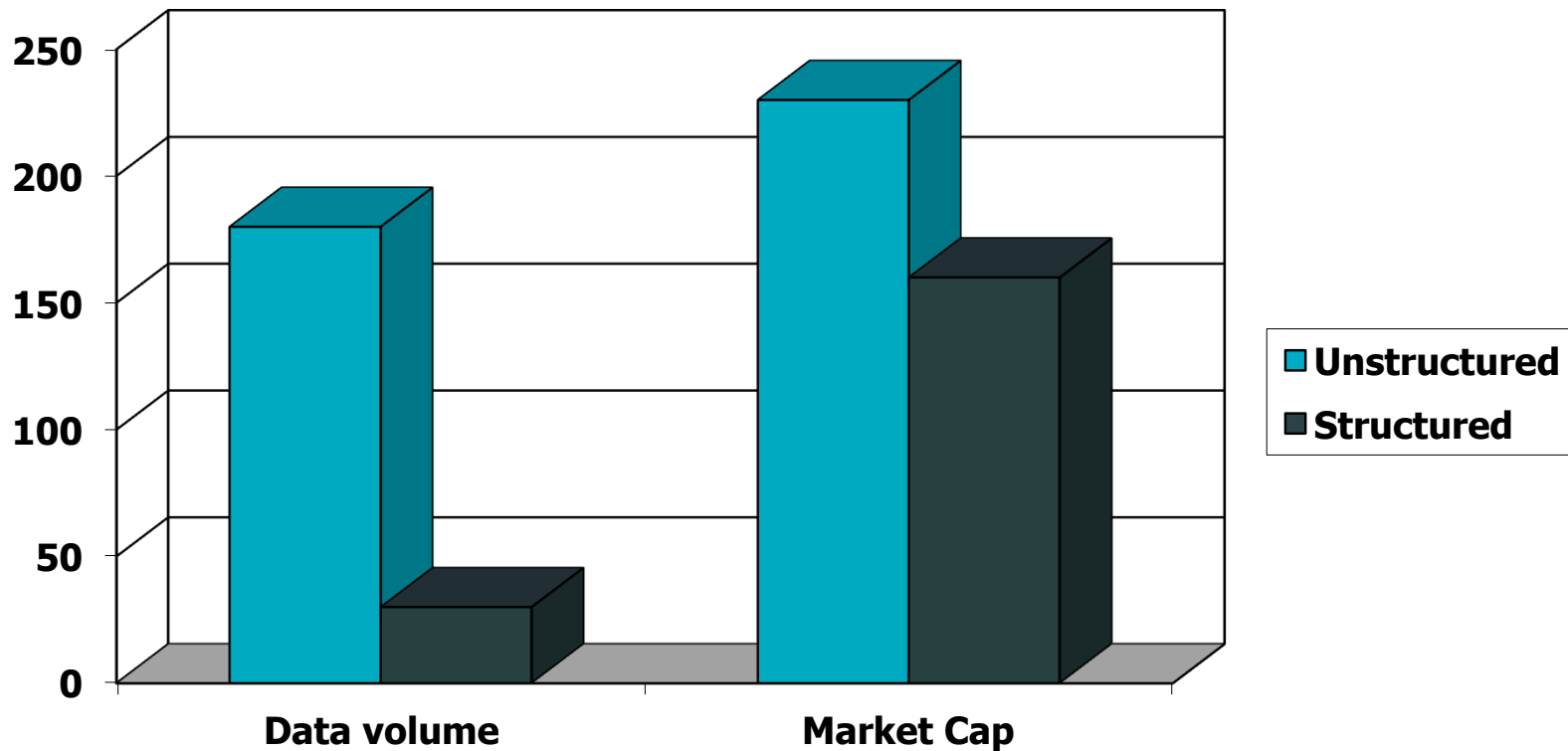
Information Retrieval

- 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



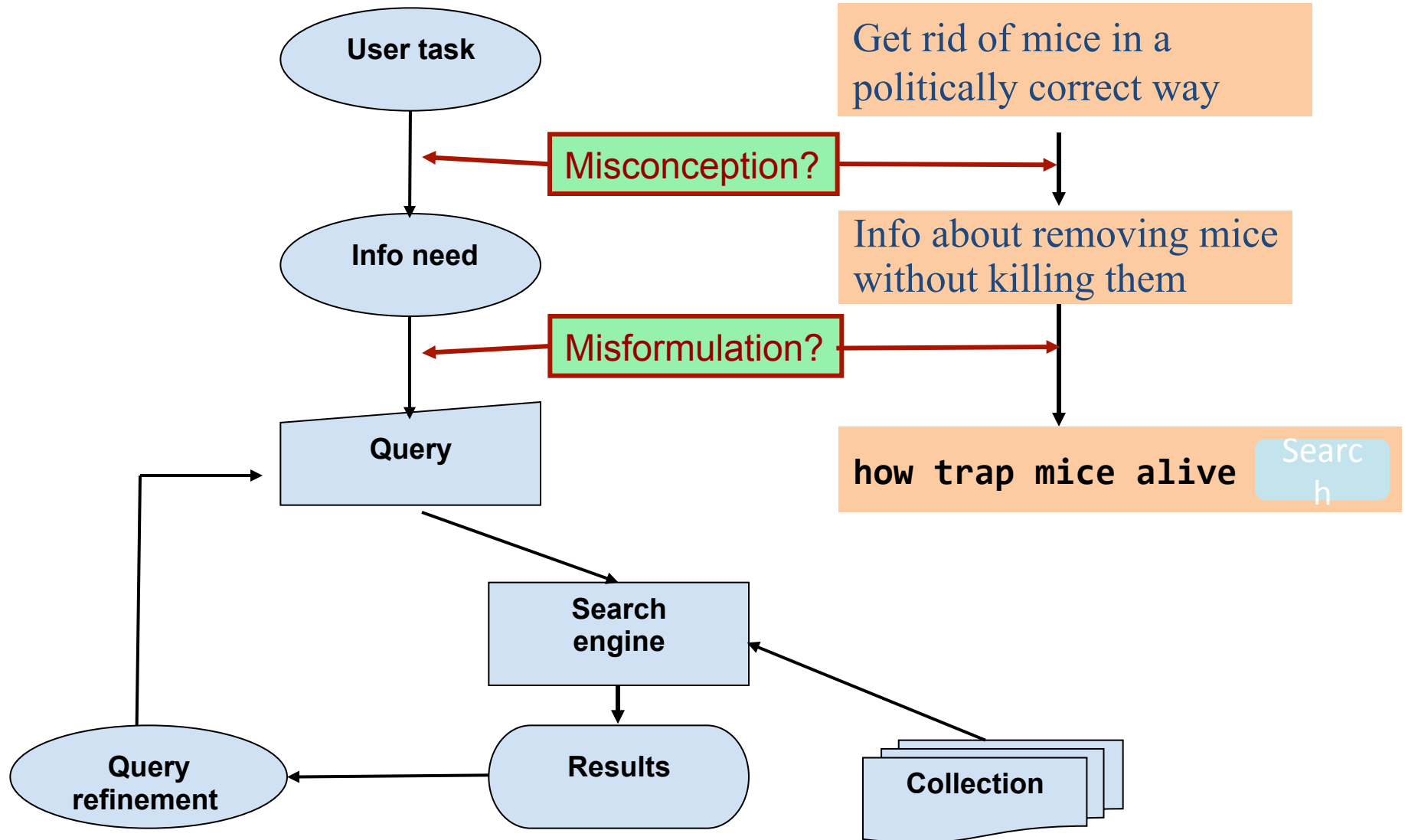
Unstructured (text) vs. structured (database) data today



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?

- *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

Introduction to **Information Retrieval**

Introducing Information Retrieval
and Web Search