

NWI-I00041 Information Retrieval Introduction

Dr.ir. Faegheh Hasibi
f.hasibi@cs.ru.nl

Nijmegen, August 31st, 2020



What is Information Retrieval?

“ Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information.

Gerard Salton, 1968



What is Information Retrieval?

“ Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information.

The information objects are not necessarily textual!

Application of IR is not only searching the web!



Dimensions of Information Retrieval

■ Information media:

- Text (the primary focus of the field)
- Images (e.g., legal, medical, and scanned documents)
- Audio (music and speech) and Video

■ Applications:

- Vertical search (e.g., product search, travel search)
- Enterprise search
- Desktop search
- Chatbots and conversational systems

■ Tasks:

- Ad hoc search
- Filtering and classification
- Question answering and passage retrieval



IR Core Issues

- Relevance
- Evaluation
- Information needs



Relevance

- A loose simple definition:
 - *A relevant document contains the information that a person was looking for when she submitted a query to the search engine.*
- Different “relevance criteria”
 - Topicality: is the document **about** the information request?
 - Readability: can I understand the text?
 - Authoritiveness: can I trust the text?
 - Child-suitability: is the text appropriate for children?
 - Etc.
- Distinction between user **topical relevance** and **user relevance**



Relevance

- A loose simple definition:
 - *A relevant document contains the information that a person was looking for when she submitted a query to the search engine.*
- Different “relevance criteria”
 - Topicality: is the document **about** the information request?
 - Readability: can I understand the text?
 - Authoritiveness: can I trust the text?
 - Child-suitability: is the text appropriate for children?
 - Etc.
- Distinction between user **topical relevance** and **user relevance**
- Different **retrieval models** are proposed to address the issue of relevance



Evaluation

- Develop evaluation measures and experimental procedures to compare ranking algorithms
- Early evaluation measures: Precision and Recall
- Typically test collections are used, consisting of queries and a list of relevant documents
 - Text REtrieval Conference (TREC)
 - <http://trec.nist.gov/>



Information Needs

- IR involve **users** and their **information needs**
- Queries are **poor representations** of **users' information needs**
- IR studies how people interact with search engines and express their information needs
 - Query expansion, clarification questions, and query suggestions are the techniques to provide more relevant results



Search Engines

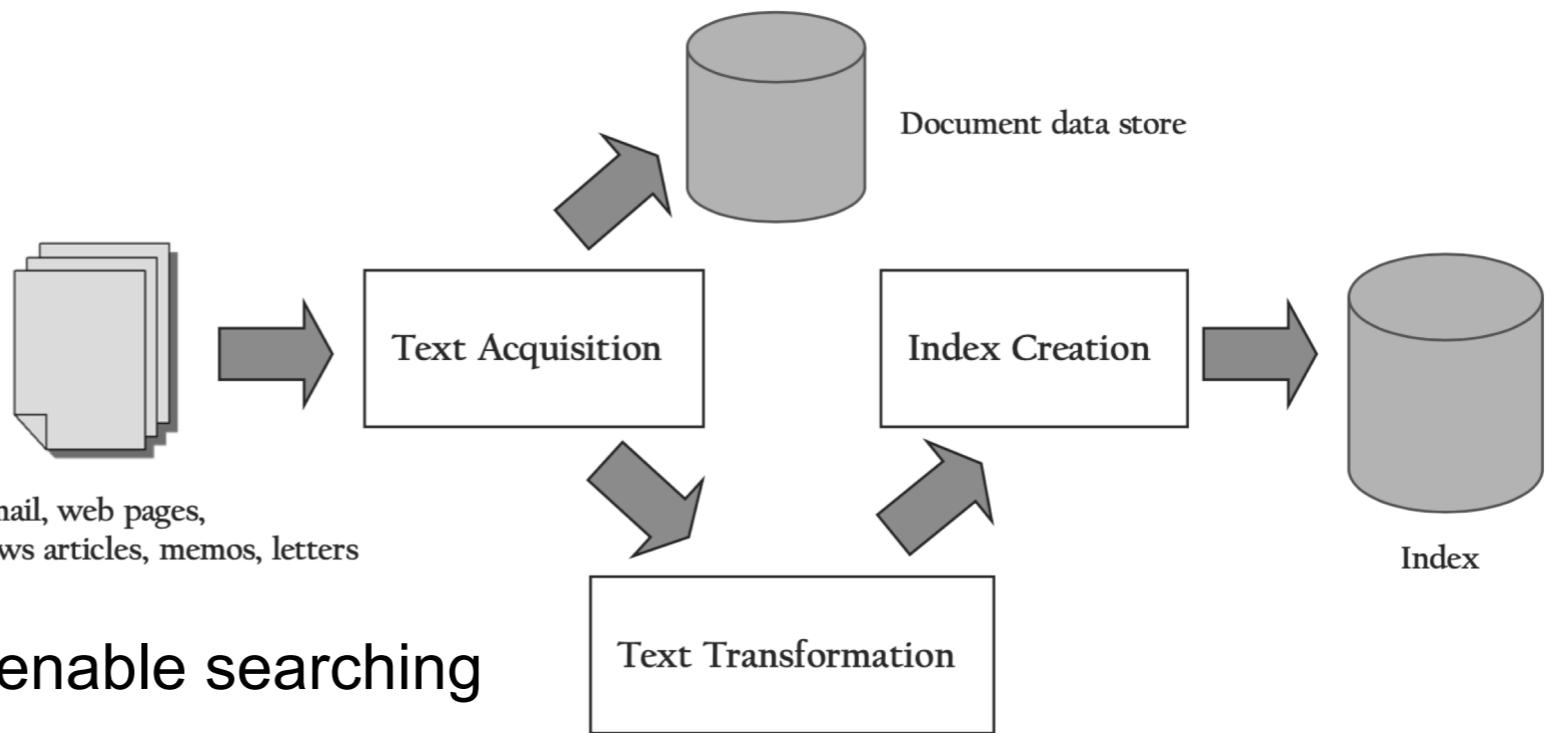
- Large scale IR application
- The main goal of search engines:
 - **Effectiveness** (quality): Retrieve the most relevant results
 - **Efficiency** (throughput): Retrieve results as quickly as possible



Search Engine Architecture (in a nutshell)

Indexing

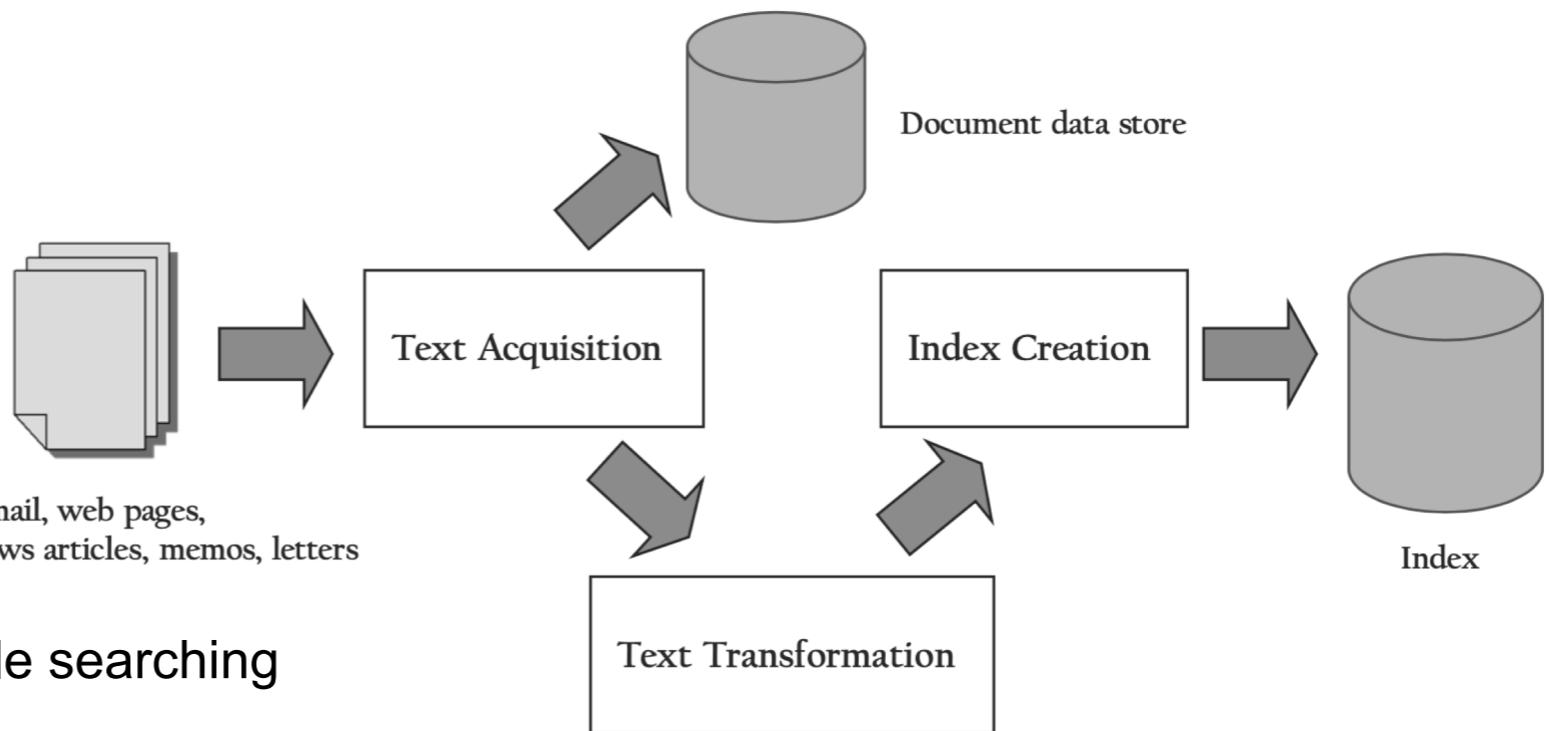
Builds the structures that enable searching



Search Engine Architecture (in a nutshell)

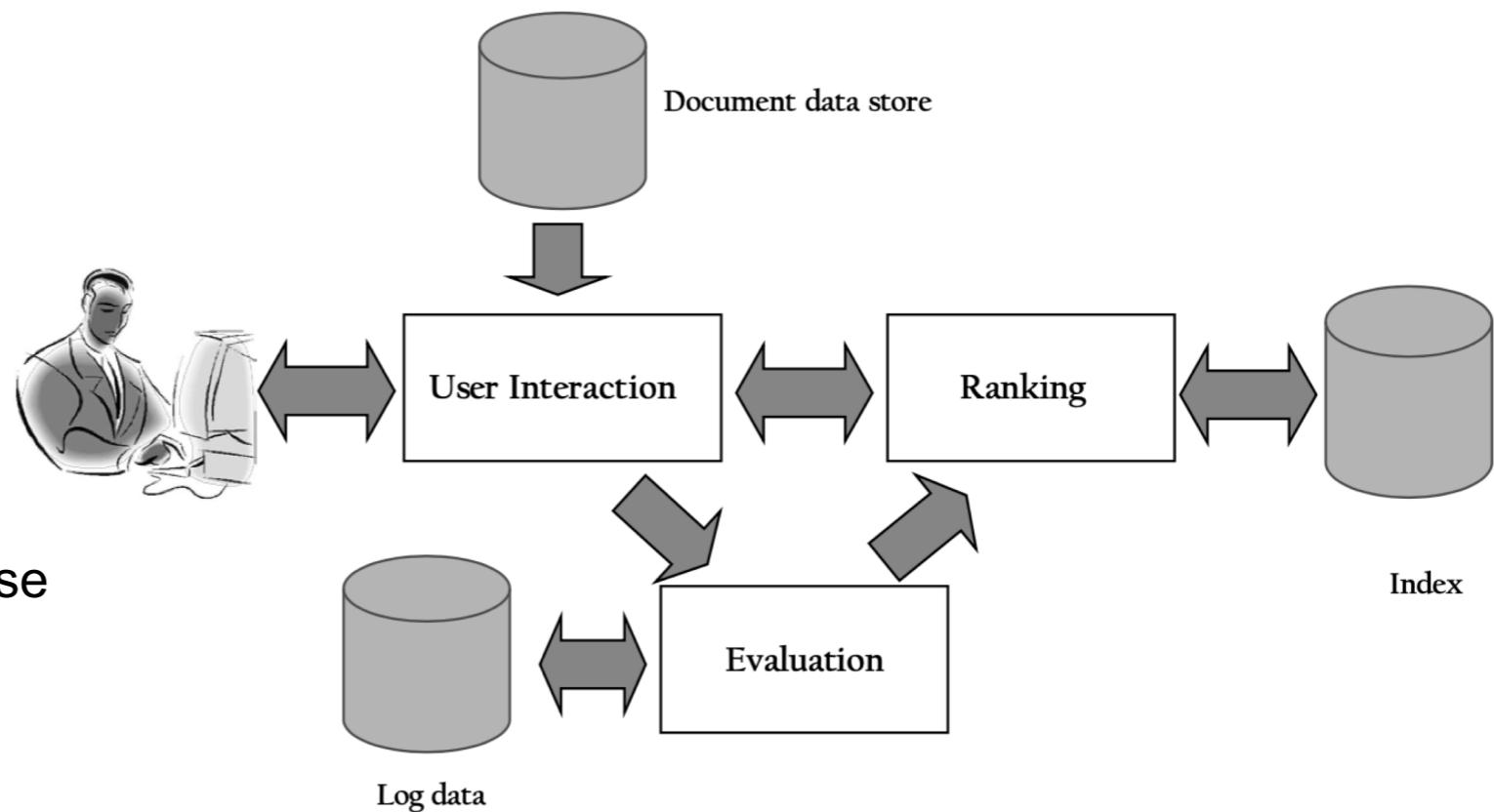
Indexing

Builds the structures that enable searching

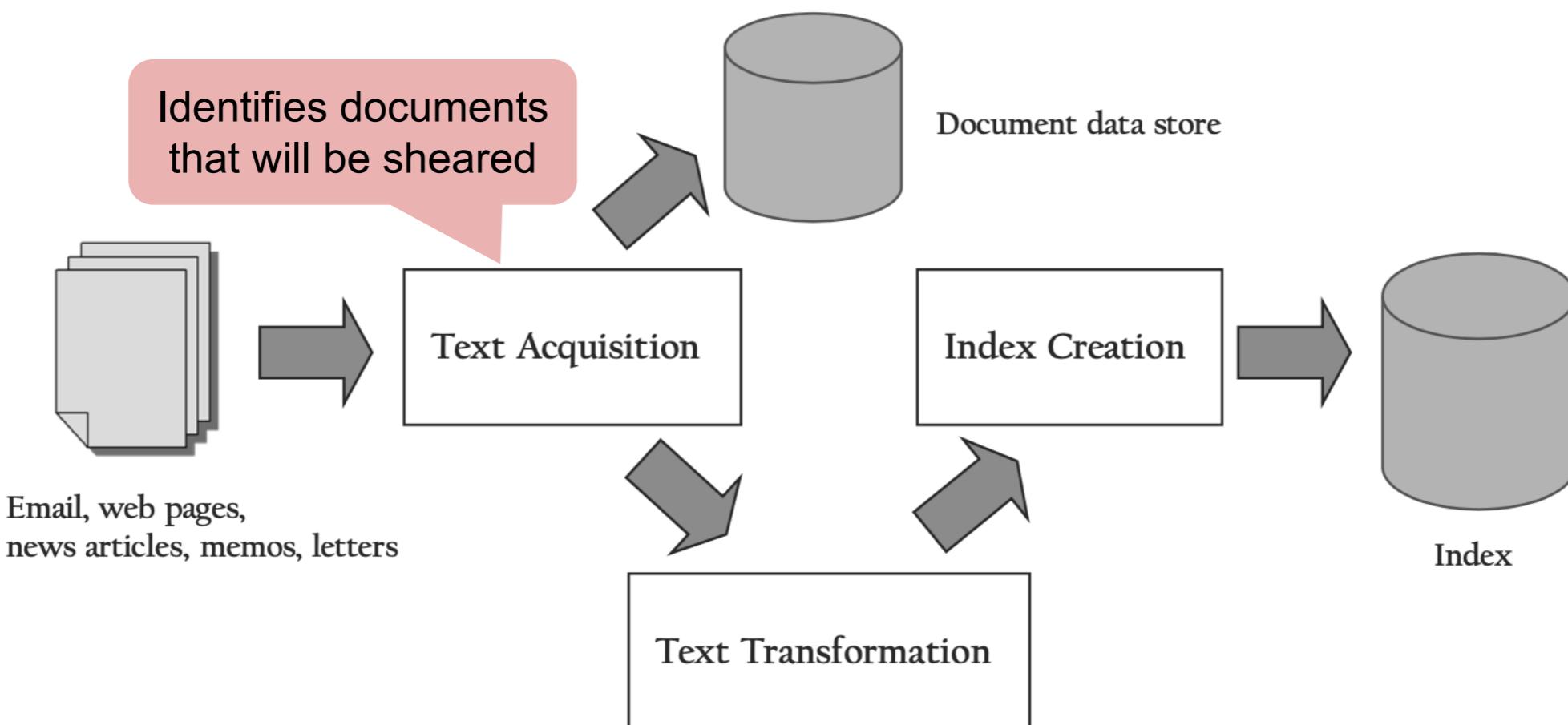


Query Processing

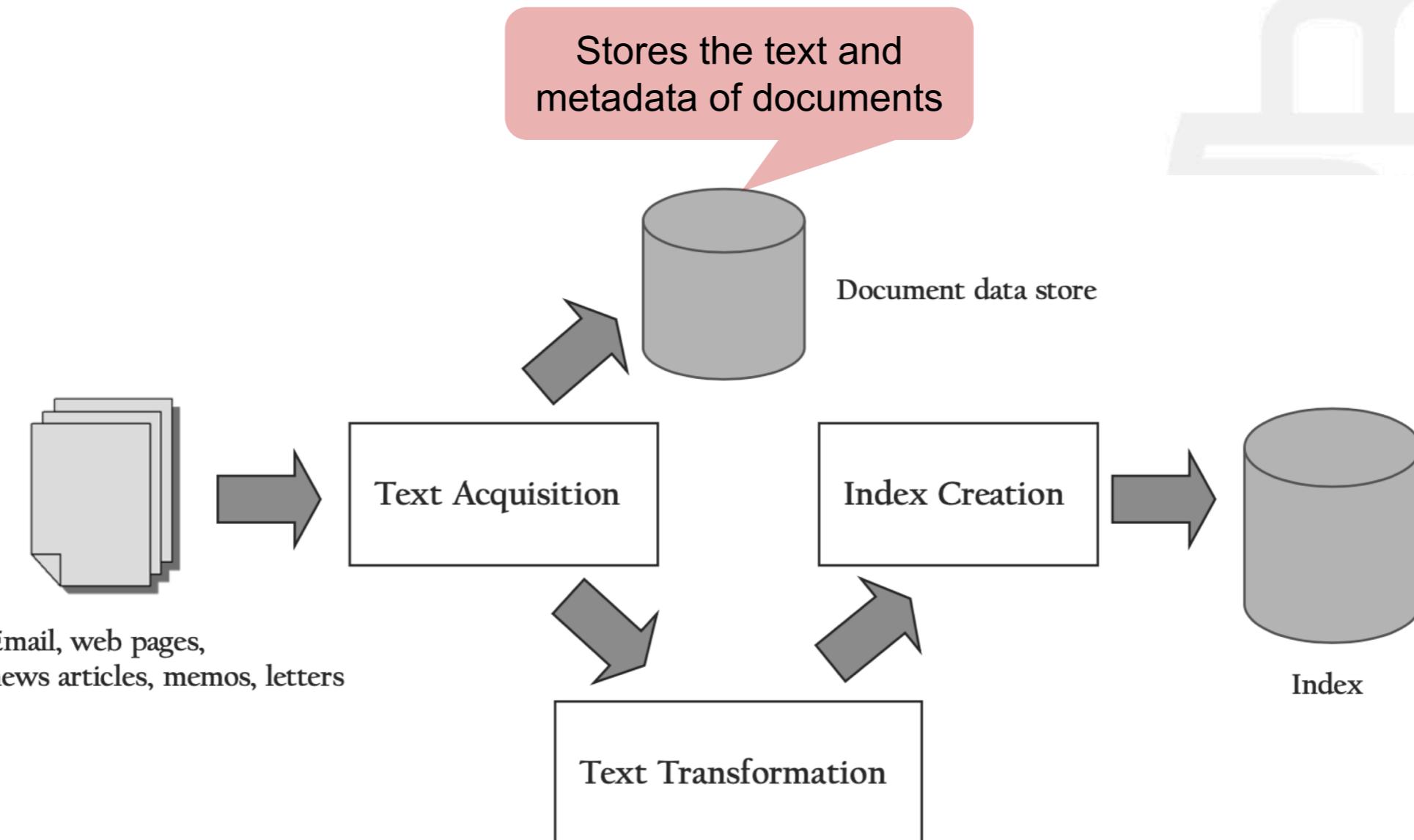
Generates the results using those structures and user's query



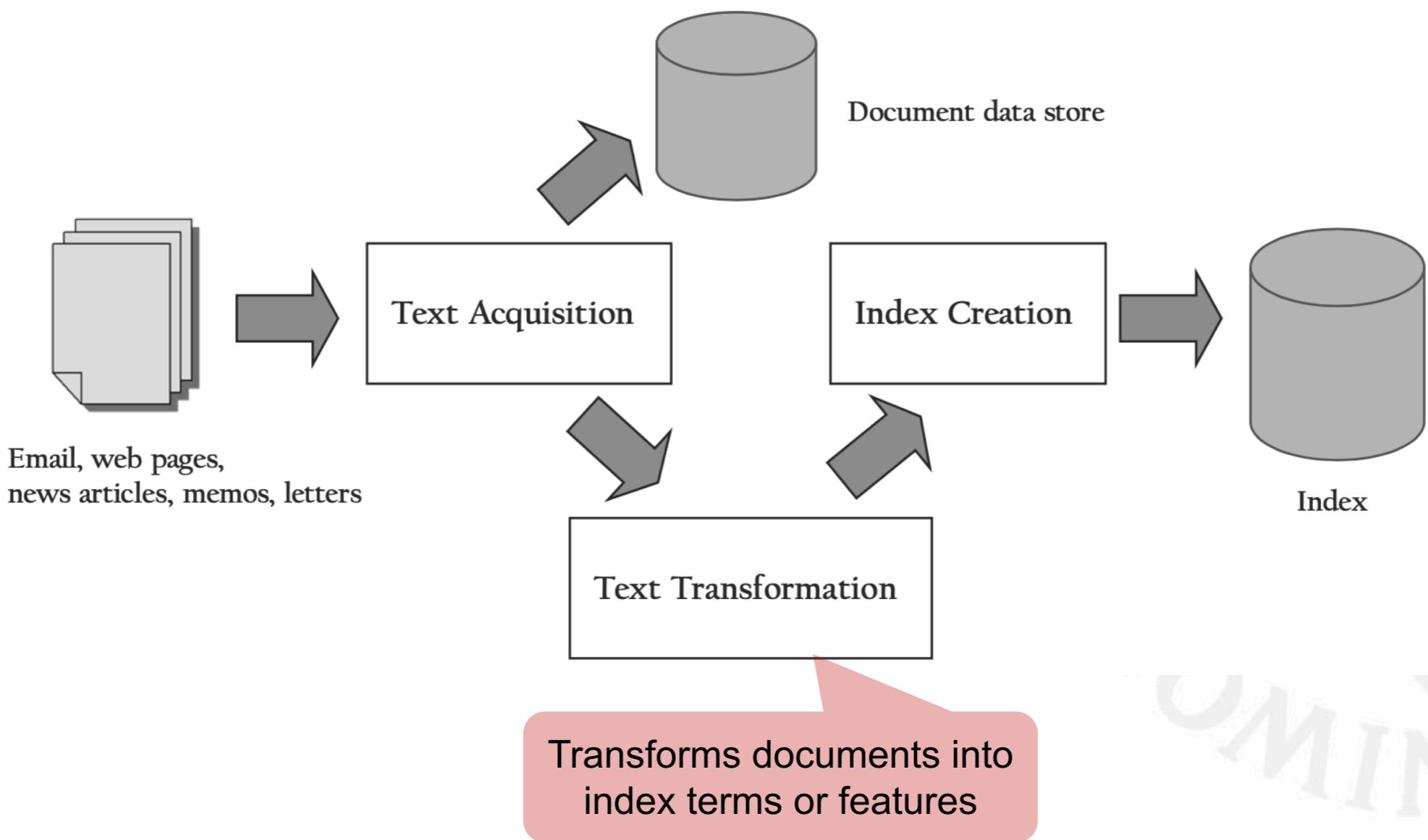
Indexing



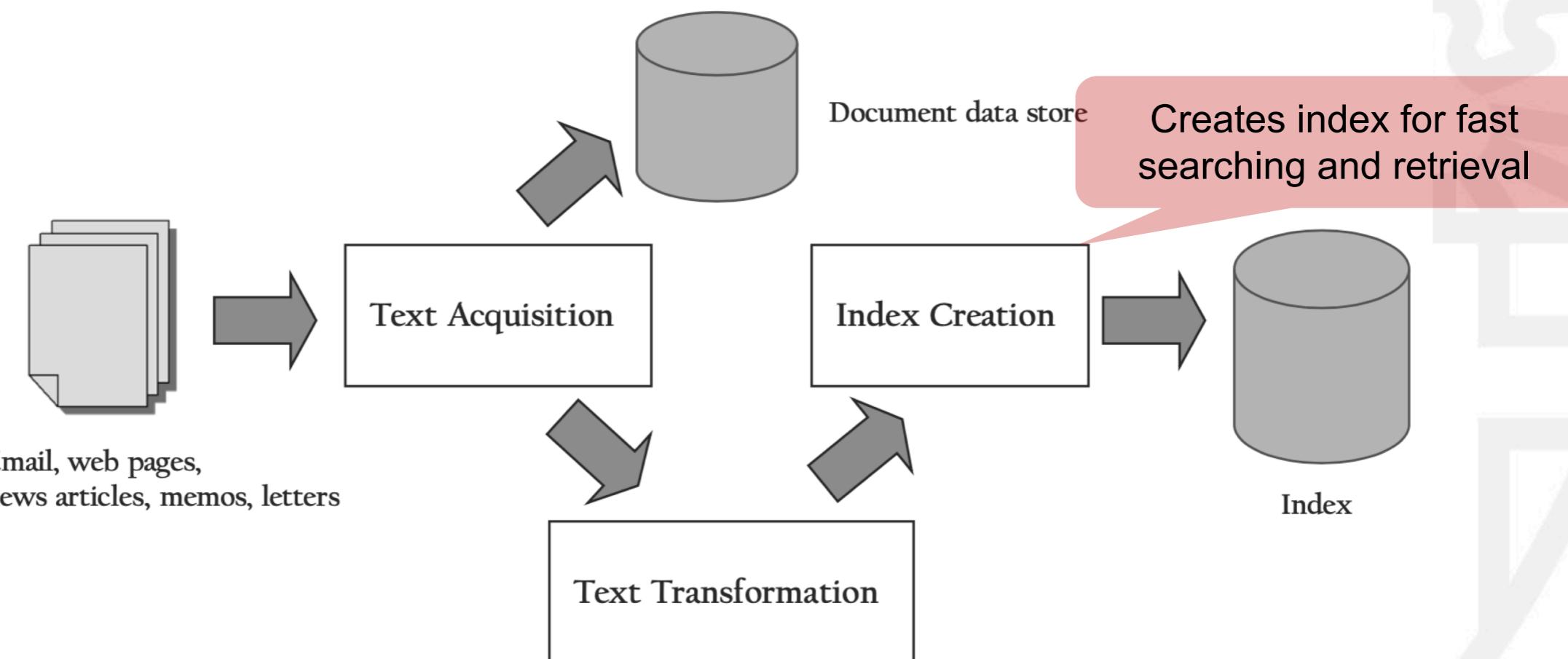
Indexing



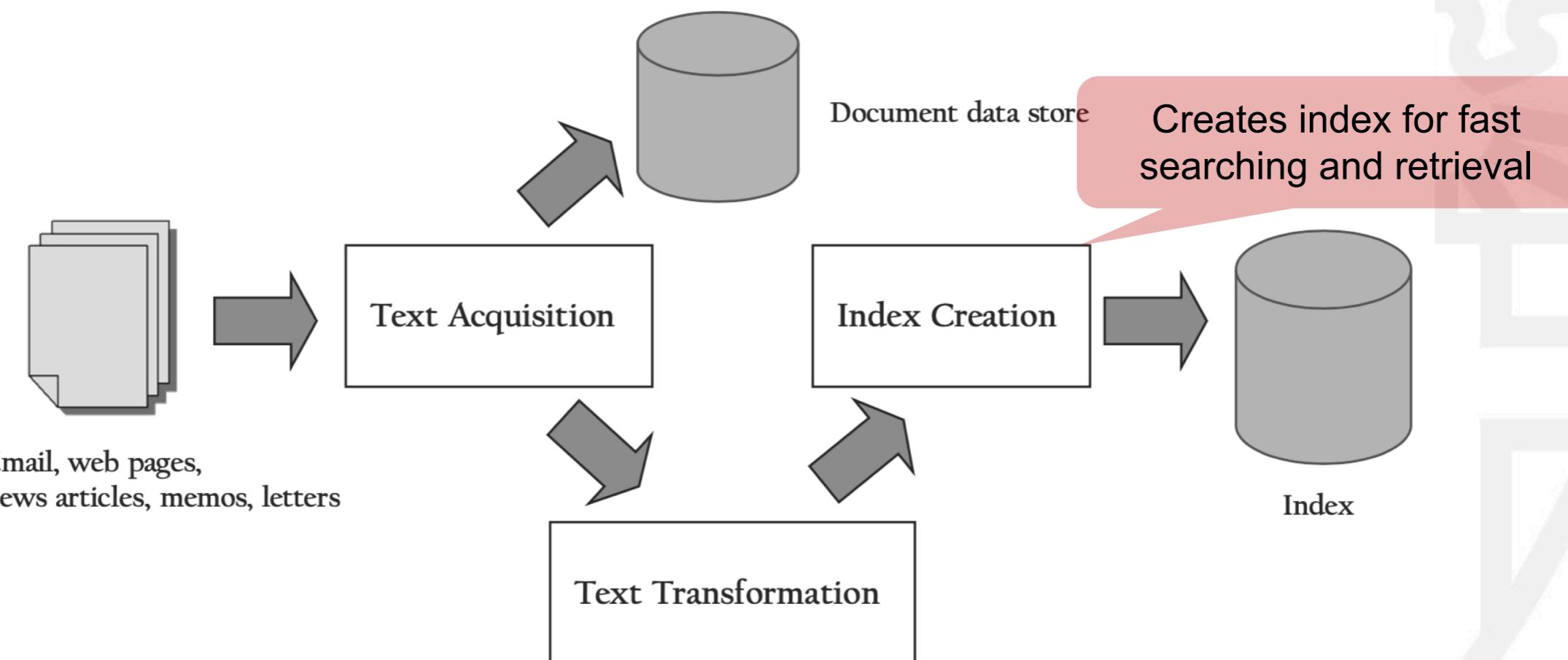
Indexing



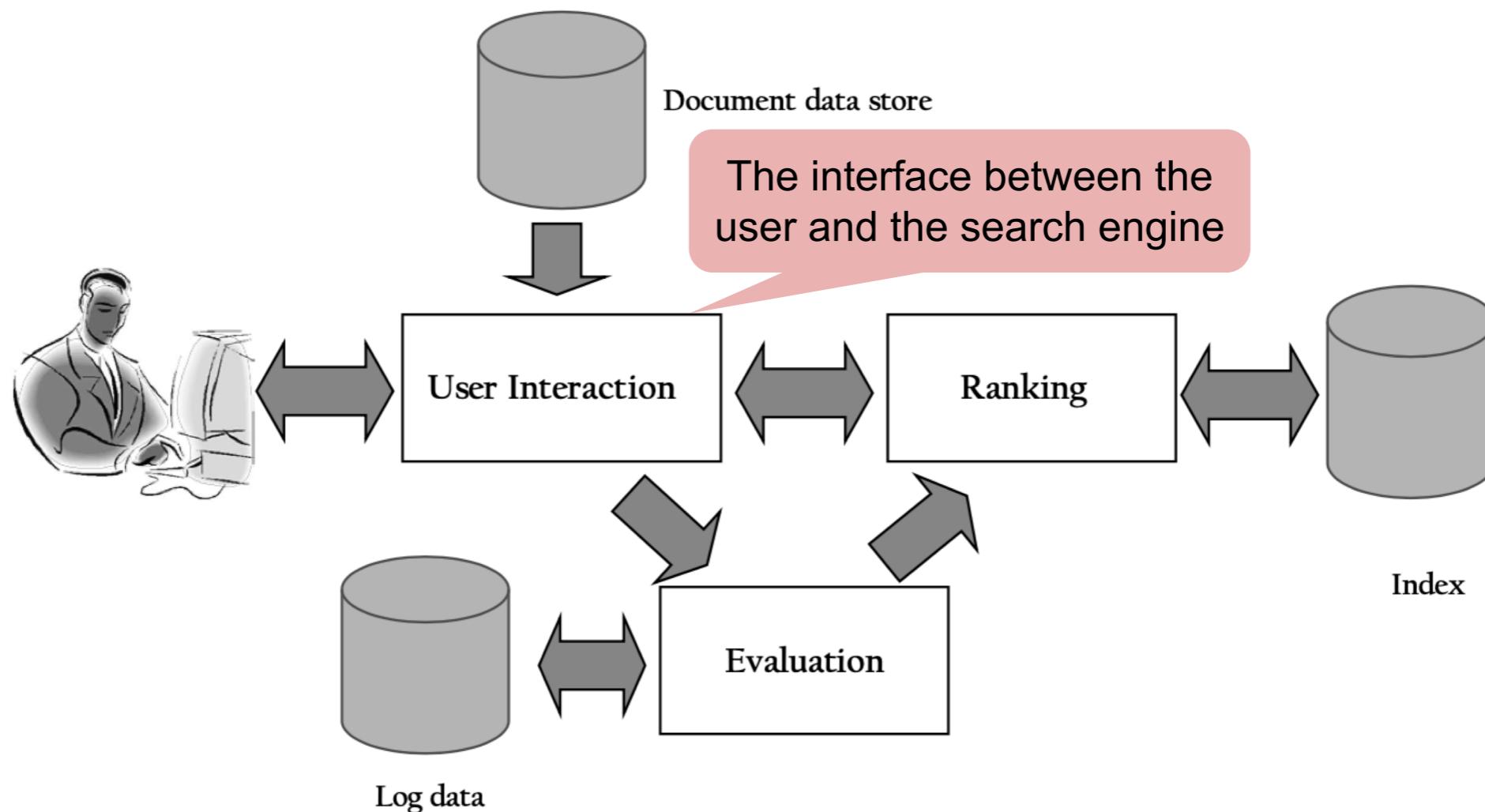
Indexing



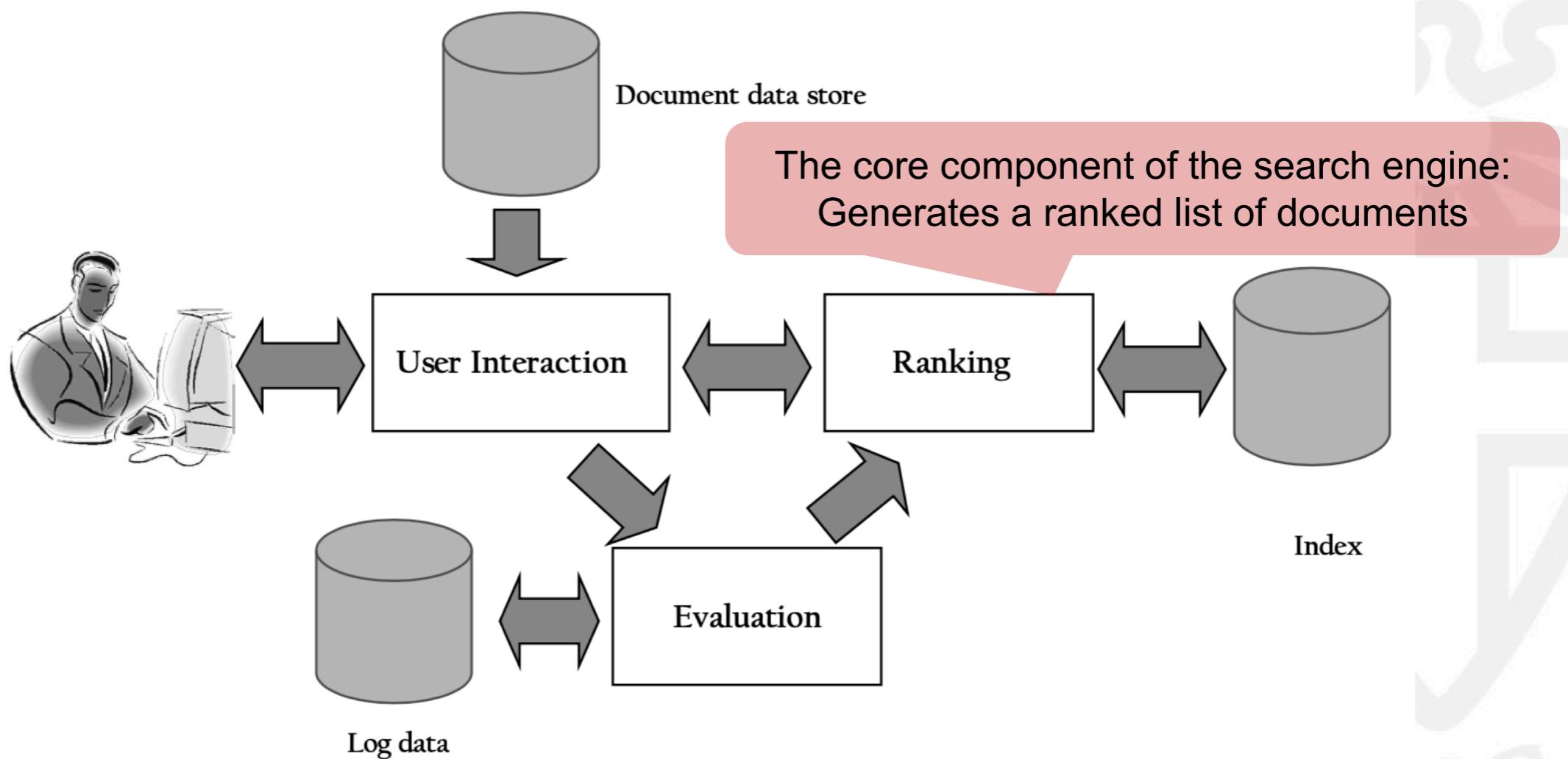
Indexing



Query processing



Query processing



Query processing

