Unit 8: Search Engines **LH 3**

8.1 Characteristics of search engine

8.2 Search Engine functionality

8.3 Ranking of Web pages

Search Engines

- Search engines are a program that searches for and identifies items in a database that correspond to keywords or characters specified by the user, used especially for finding particular sites on the World Wide Web.

- Search engines utilize automated software applications (referred to as robots, bots, or spiders) that travel along the Web, following links from page to page, site to site. The information gathered by the spiders is used to create a searchable index of the Web.

- Every Search Engine (Yahoo, Google. Bing) has different algorithm to rank page. So same search will have different result in different search Engine.

There are three basic components of a search engine as listed below:

- Web Crawler
- Database
- Search Interfaces

Web crawler

- It is also known as **spider** or **bots.** It is a software component that traverses the web to gather information.

Database

- All the information on the web is stored in database. It consists of huge web resources.

Search Interfaces

- This component is an interface between user and the database. It helps the user to search through the database.

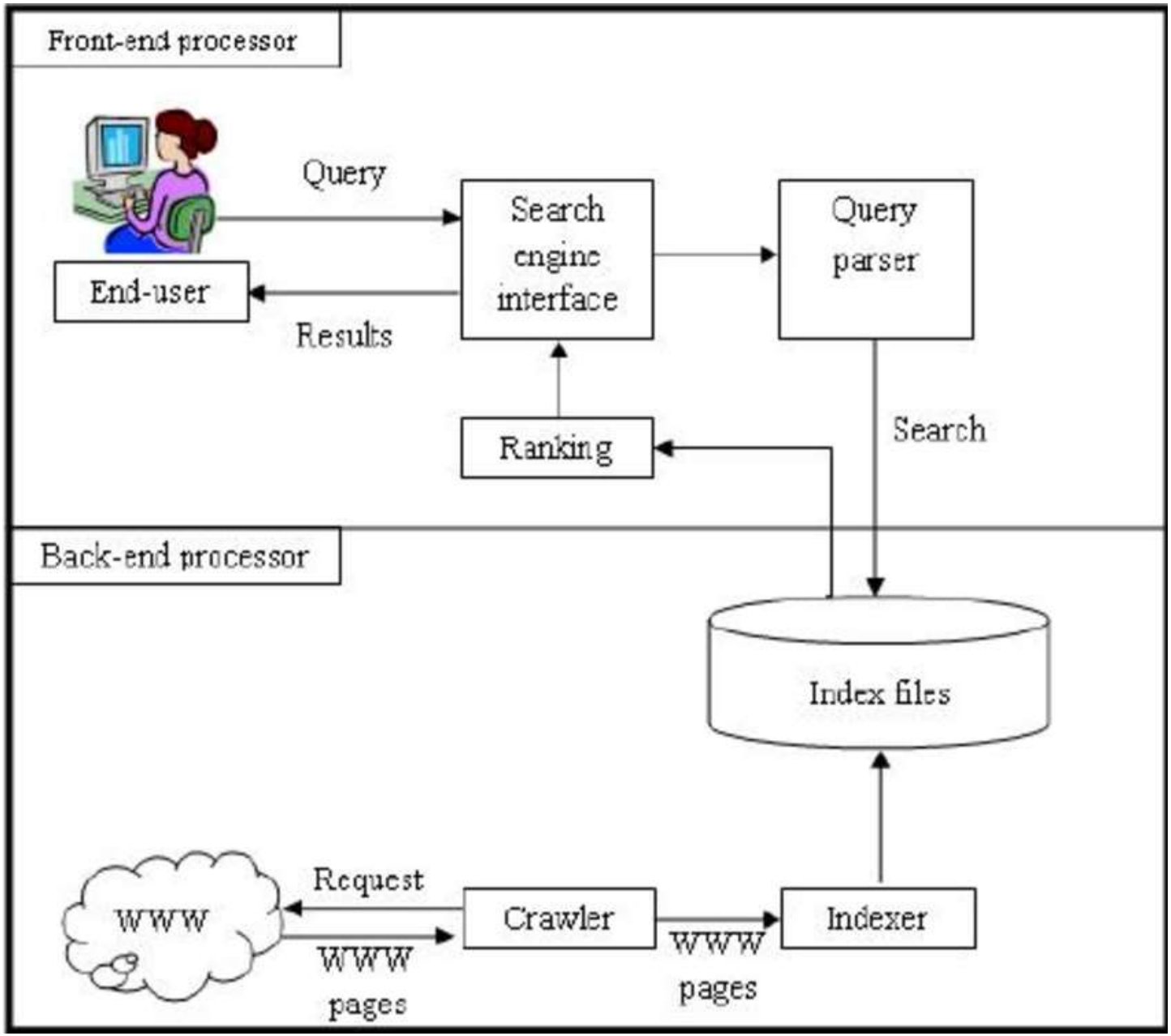
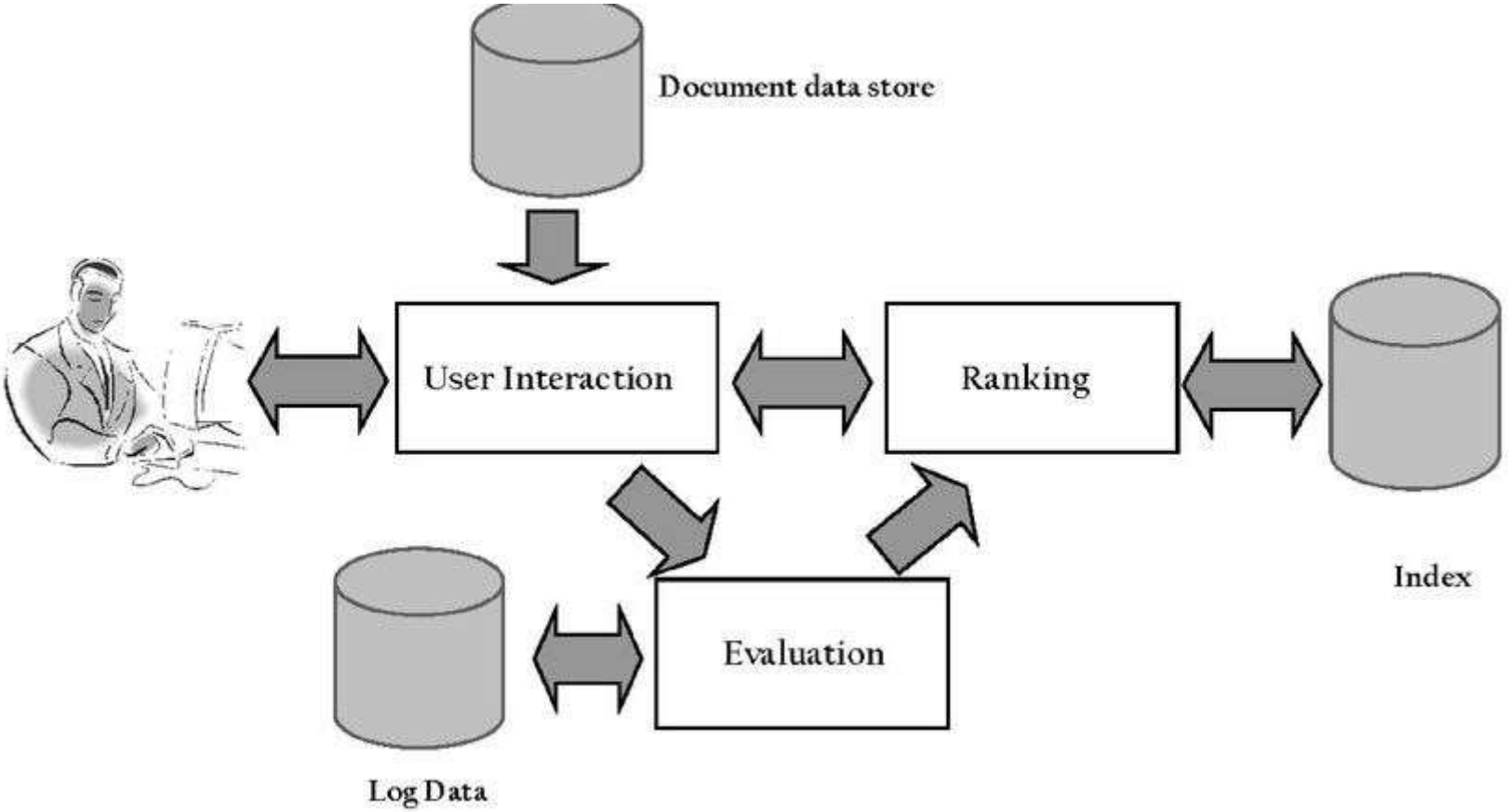


Fig: Architecture of Search Engine

Search Engine Architecture is Mainly Divided into two section.

- Front End (Query Parse)
- Back End (Index File)

Front End (Query Parse)



User Interaction:

- Support creation and refinement of query, display of results.

Ranking

- Uses query and indexes to generate ranked list of documents.

Evaluation:

- Monitors and measures effectiveness and efficiency.

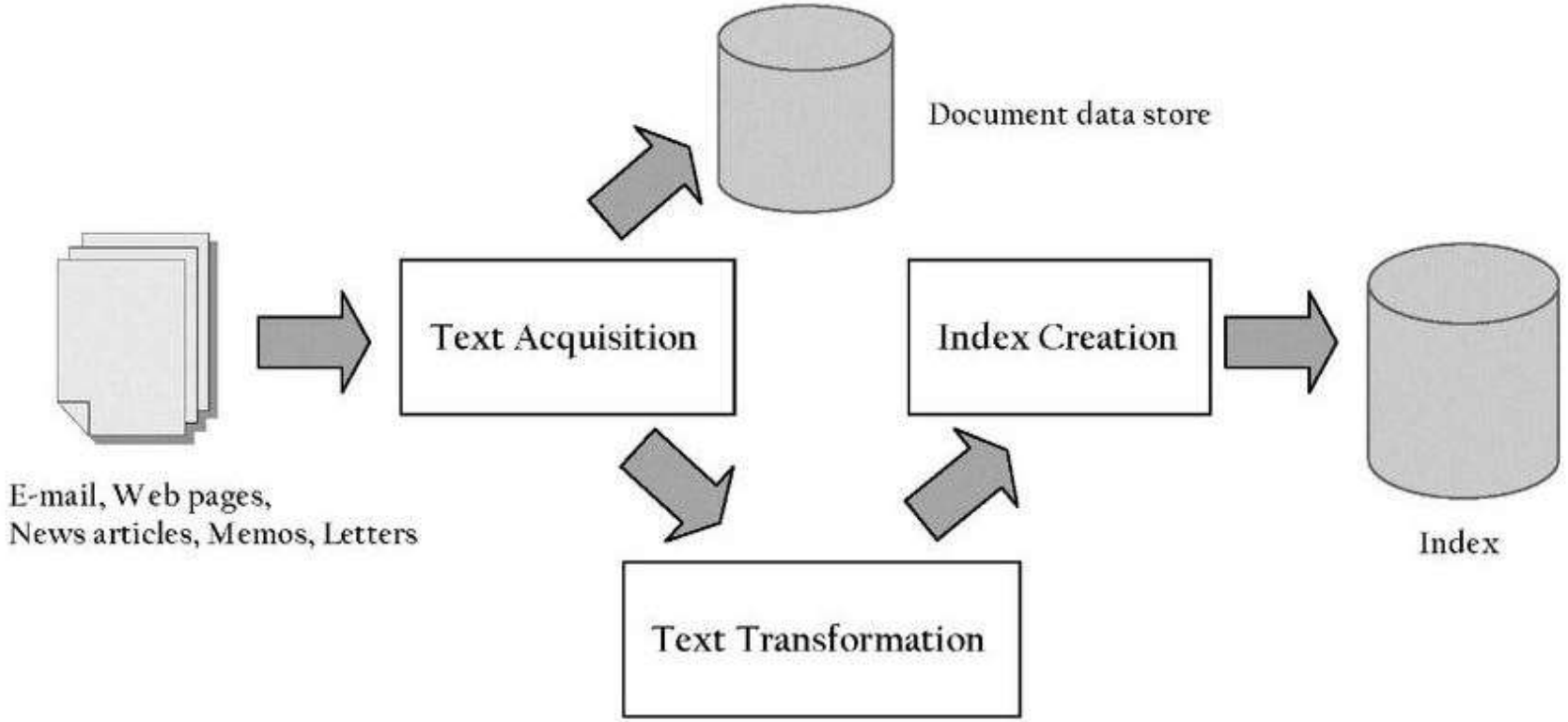


Fig:Indexing Process

Text Acquisition:

- Identifies and stores documents for indexing

Text Transformation

- Transforms documents into index terms or features

Index creation

Takes index terms and creates data structure (indexes) support fast searching

Characteristics of Search Engine

i) **Unedited**

- Anyone can enter content Quality issues; Spam

ii. **Varied information types**

- Phone book, brochures, catalogs, dissertations, news reports, weather, all in one place!

iii. **Different kinds of users**

- Lexis-Nexis: Paying, professional searchers.
- Online catalogs: Scholars searching scholarly literature.
- Web: Every type of person with every type of goal.

iv. **Scale**

- Hundreds of millions of searches/day; billions of docs.

Functions of Search Engine

Crawling:

- Web spider or web crawler a robust software of search engine that sorts out a website address and relevant content on the Internet for storing in a database of the search engine. The crawler can scan old website data and fresh new information. Crawlers possess the ability to crawl websites and fetch a considerable amount of data simultaneously. It crawls until it gets to the best possible result that involves hyperlinks and external or internal links.

Indexing

- The indexing function of a search engine first excludes any unnecessary and common articles such as "the," "a" and "an." After eliminating common text, it stores the content in an organized way for quick and easy access. Search engine designers develop algorithms for searching the web according to specific keywords and keyword phrases. Those algorithms match user-generated keywords and keyword phrases to content found within a particular website, using the index. .

Storage

- For easy and fast access to search results, the search engine database should store web content in the search engine database. Content fetched to the user the whole dependent on the storage space allocated. Search engines such as Yahoo and Google can store massive amounts of data, offering an immense source of information for the user.

Results

- Hyperlink are directed towards the websites shown on the search engine result page when a keyword or keyword phrase is searched. After analyzing the search term, the crawler goes through indexing and finds a suitable match of typed keywords. Algorithms provide the relevant data, and every search engine works on its algorithms and appropriately returns varied results.