Day 2 Notes

Search engine

1. Crawler & indexer: for effecient access
2. User input: Defines scope of search engine
3. Query parser
4. Rank: match query with document
5. Domain specific database : pushes from the web. google indexes it and makes it searchable.

- Google crawls on its own or content providers push it to google.

6. Post processing and display results

Abstraction

Indexed Corpus:

1. Crawler
2. Doc analyzer
3. Indexer (for effecient access)

Ranking proceduer + user:

4. User (query) => query rep => indeces => [ranker] -> results -> user feedback

–post image–

Query

- Designed representation of users info. Via NL or some form

Document

- Representation of information that potentially satisfies users' info need
- Text image video, etc.

Relevance

- Relatedness between documents and users' info needs
- Multiple perspectives: topical, semantic, temporal, spatial, etc.

Definition of IR:

–fill note–

Web craler

- Automated prog

–fill note–

Search Engine

- Search query logs: Record users' interaction history with search engine
- User modeling: Understanding users' longitudinal information need. Assess users' satisfaction towards search engine output

Browsing vs Querying

-