

Notes 6

Recap of inverted index

- input user query -> query parsing -> throw that into inverted index -> find associated documents and return to user
- same approach for mapping stemming to actual queries
- query expansion: term A -> synonym of A

Judging Criteria

Brainstorm - What is important when evaluating a search engine?

- % of Relevant articles
- Customer satisfaction
- Speed
- How “Smart” the engine is -> if it can learn

What is actually used in industry

1. Dwell time on ranked result
2. Amount of scrolling -> more common for mobile searching

‘Correct’ metric

- The ideal goal is to satisfy users’ information needs
- We try to approximate this.

Metric approximation

1. Information need = Reflected my query
 - Categorize information need into : Navigational, Informational , Transactional
 - **Navigational:** User doesn’t know information need much
 - **Informational:** User knows query. Prefer more long clicks
 - **Transactional:** You get led to results
2. Satisfaction
 - We approximate satisfaction as less effort = satisfaction
 - Quality of search result . Higher quality = satisfaction

Classic IR Evaluation

1. Define collection
2. Fix set of queries

3. Set of relevance judgements. Check to see if you satisfy this metric. Not how you rank, etc.
- Relevance is with respect to information need. NOT the key words of the query
- Two ways: unranked retrieval sets vs ranked retrieval

Unranked

- Boolean retrieval:
- Precision: fraction of retrieved documents are relevant $p(\text{rel} | \text{retr})$. Return less, be more conservative
- Recall: fraction of relevant docs retrieved $p(\text{retr} | \text{rel})$. Return more.
- Choosing one over the other. Unless you have NO results or PERFECT ranking

Summarize precision and recall to single value.

- In order to compare different systems
- Computer F-measure: weighted harmonic mean of precision and recall. . Alpha balances trade off. F1 score is more sensitive to lower value than arithmetic average. The F1 score / harmonic mean tells you worst case .

Ranked

- Calculate precision and recall with respect to rank. At every precision, calculate precision and recall
- Decide which curve is better. Area under curve \Rightarrow effort user has to spend