V. Users and Queries

- Information Needs and Queries
- Query Transformation and Refinement
- Cross-Language Search
Information Needs and Queries

Information Needs

- An information need is the underlying cause of the query that a user submits to a search engine
  - Sometimes called information problem: information need is generally related to a task
- Categorized using variety of dimensions
  - Number of relevant documents being sought
  - Type of information that is needed
  - Type of task that led to the requirement for information
Information Needs and Queries

Queries

- A query can represent very different information needs
  - May require different search techniques and ranking algorithms to produce the best rankings (see part on retrieval models)

- A query can be a poor representation of the information need
  - User may find it difficult to express the information need
  - User is encouraged to enter short queries both by the search engine interface, and by the fact that long queries often do not work
Information Needs and Queries

Interaction

- Interaction with the system occurs . . .
  - During query formulation and reformulation
  - While browsing the results

- Key aspect of effective retrieval
  - Users cannot change ranking algorithm but can change results through interaction
  - Helps refine description of information need
    - E.g., same initial query, different information needs
    - How does a user describe what she does not know?
Belkin et al. (1982) proposed a model called Anomalous State of Knowledge (ASK) hypothesis. Difficult for people to define exactly what their information need is, because that information is a gap in their knowledge. Search engine should look for information that fills those gaps. Interesting ideas, little practical impact (yet).
Information Needs and Queries

Keyword Queries

- Query languages in the past were designed for professional searchers (intermediaries)

User query:
Are there any cases which discuss negligent maintenance or failure to maintain aids to navigation such as lights, buoys, or channel markers?

Intermediary query:
NEGLECT! FAIL! NEGLIG! /5 MAINT! REPAIR! /P NAVIGAT! /5 AID EQUIP! LIGHT BUOY "CHANNEL MARKER"

Remarks:
- Intermediary query has OR-semantics
- ! is a wildcard (NEGLECT! matches neglected, neglects, neglect, etc.)
- /5 means “within five words”
- /P means “in the same paragraph”
- ” indicate a phrase
Information Needs and Queries

Keyword Queries

- Simple, natural language queries would enable everyone to search
- Current search engines do not perform well (in general) with natural language queries
- People trained (in effect) to use keywords
  - Compare average of about 3–4 words per web query to average of 30 or more words/CQA query (community-based question answering)
  - What would you answer someone approaching you just asking tropical fish?!
- Keyword selection is not always easy
  - Query refinement techniques can help