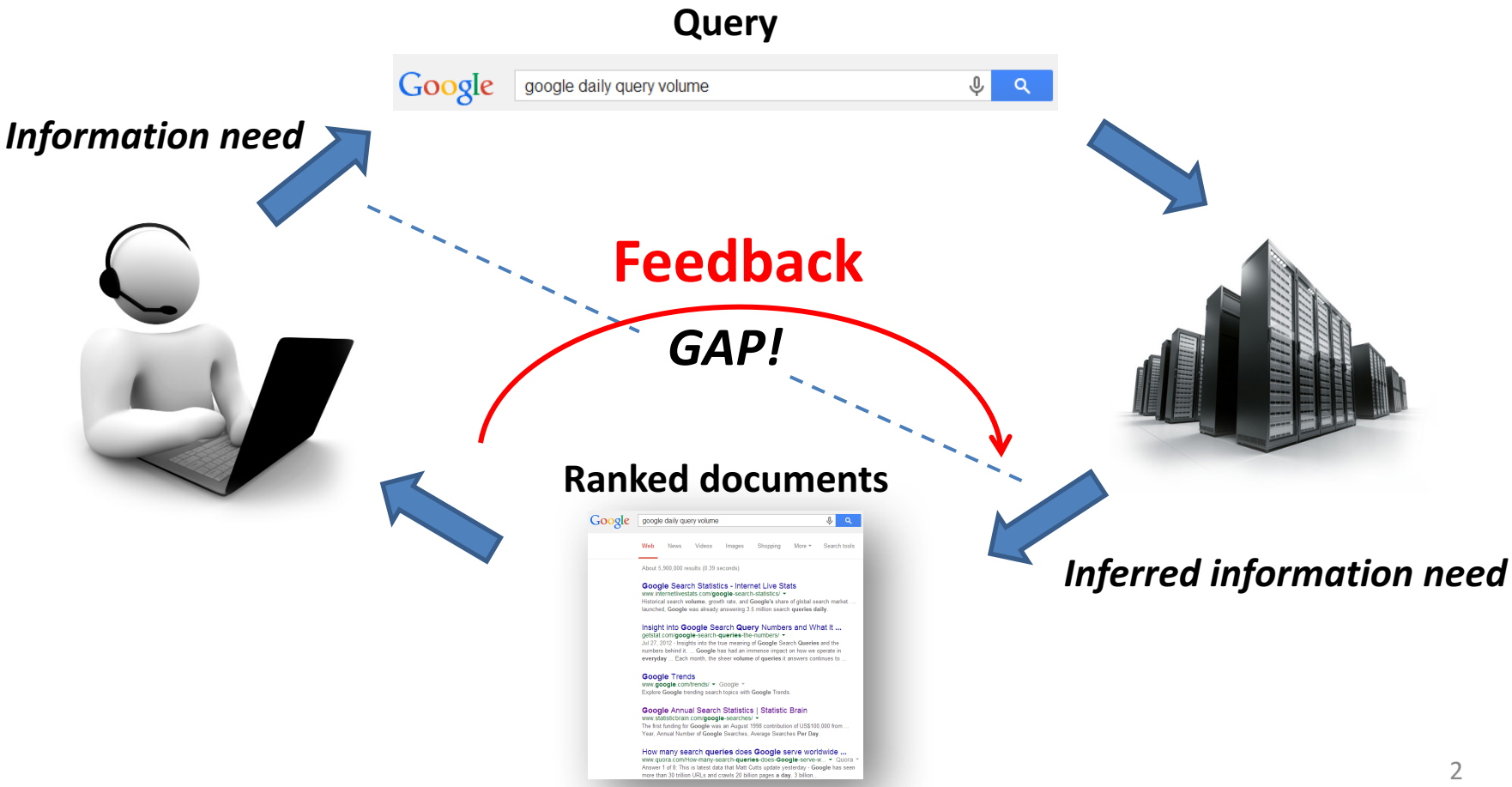
Relevance Feedback

Slides borrowed from Stanford and Hongning Wang with modifications

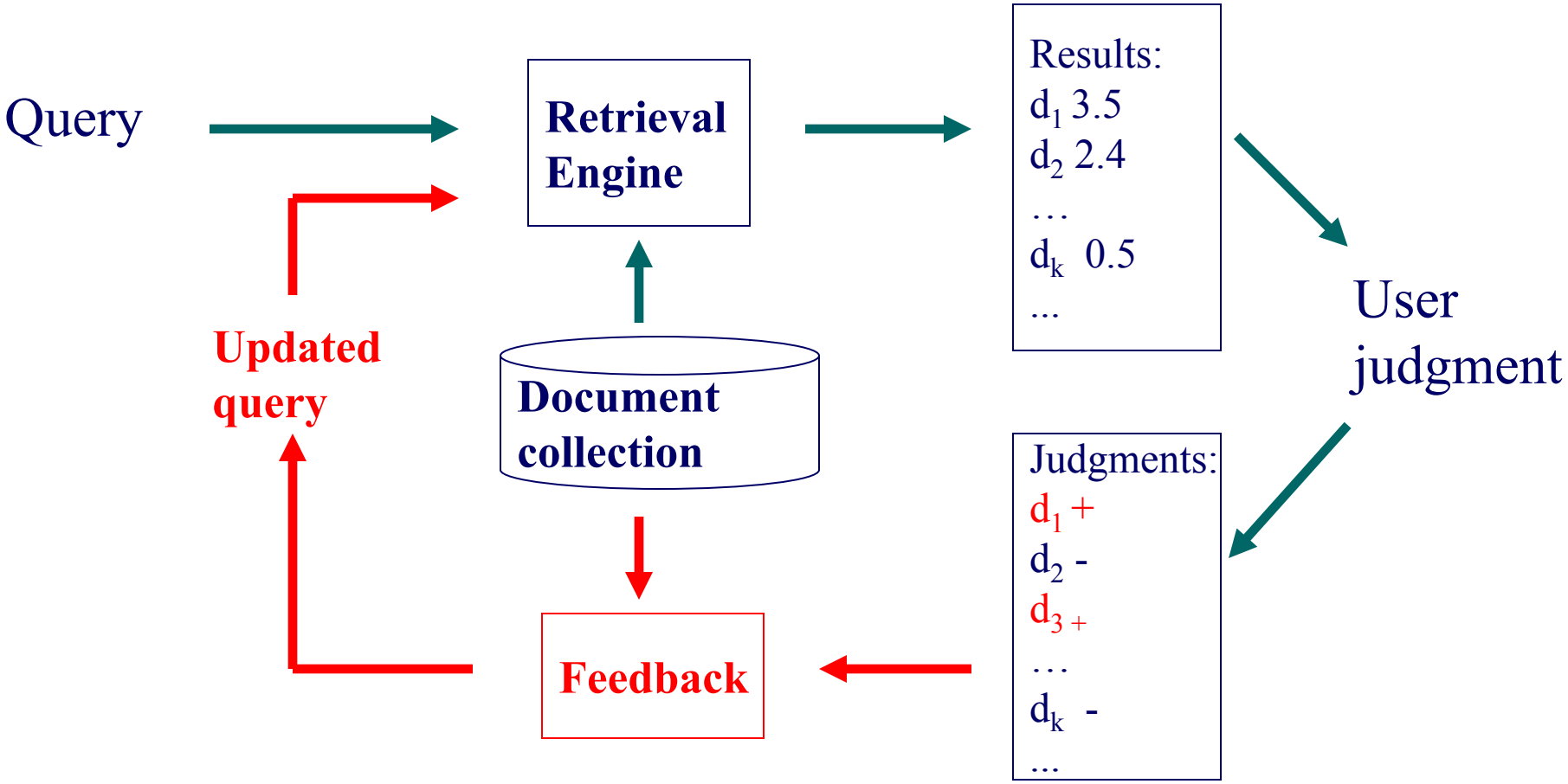
User feedback

should be

- An IR system ~~is~~ an interactive system



Relevance feedback



Basic idea in feedback

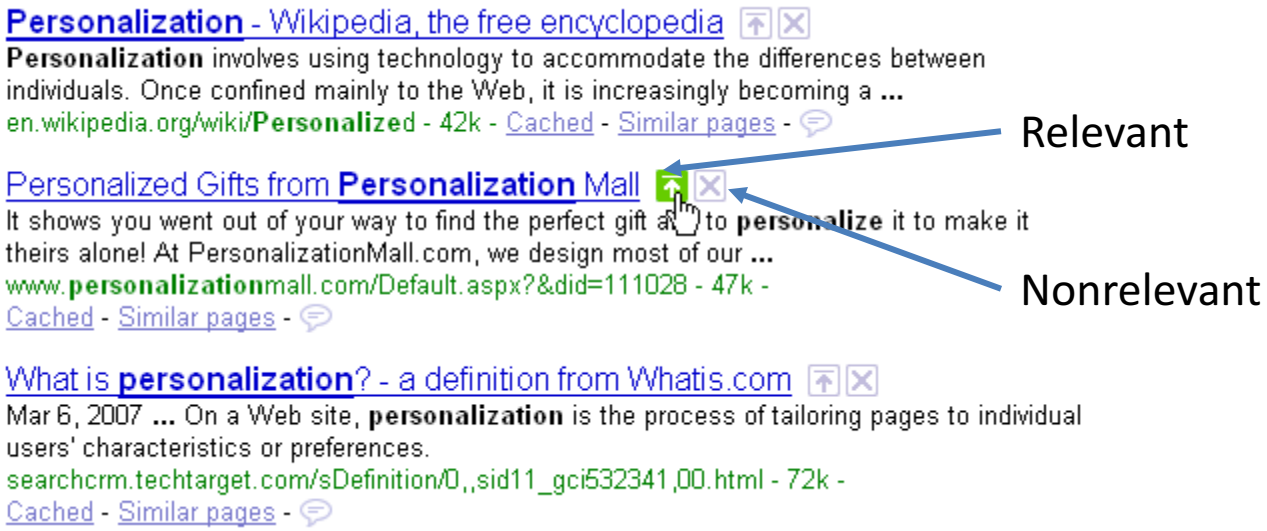
- Query expansion
  - Feedback documents can help discover related query terms
  - E.g., query="information retrieval"
    - Relevant docs may likely share very related words, such as "search", "search engine", "ranking", "query"
    - Expand the original query with such words will increase recall and sometimes also precision

Basic idea in feedback

- Learning-based retrieval
  - Feedback documents can be treated as supervision for ranking model update
  - In “learning-to-rank”

Relevance feedback in real systems

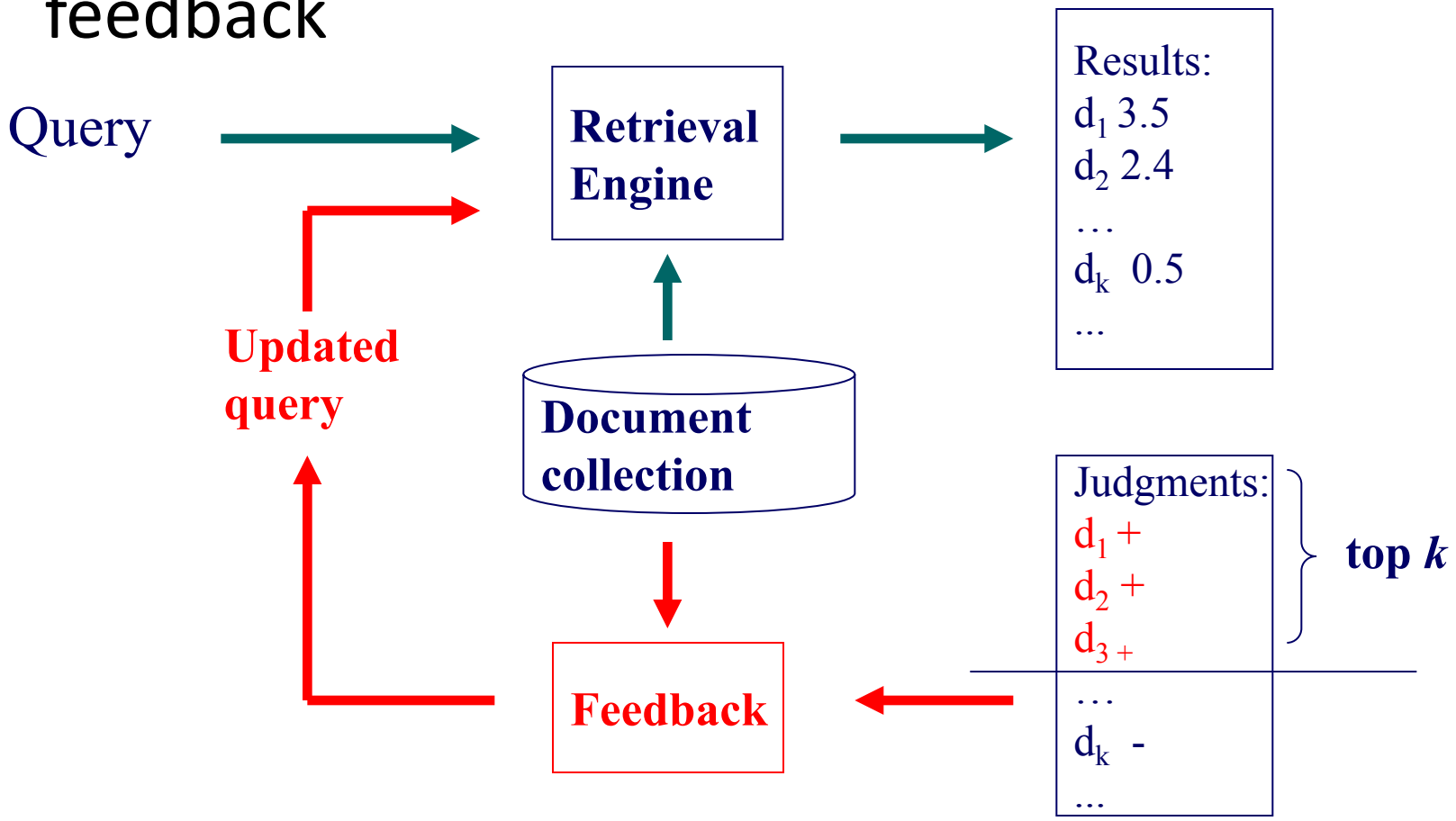
- Google used to provide such functions



  - Vulnerable to spammers though

Pseudo relevance feedback

- What if the users are reluctant to provide any feedback



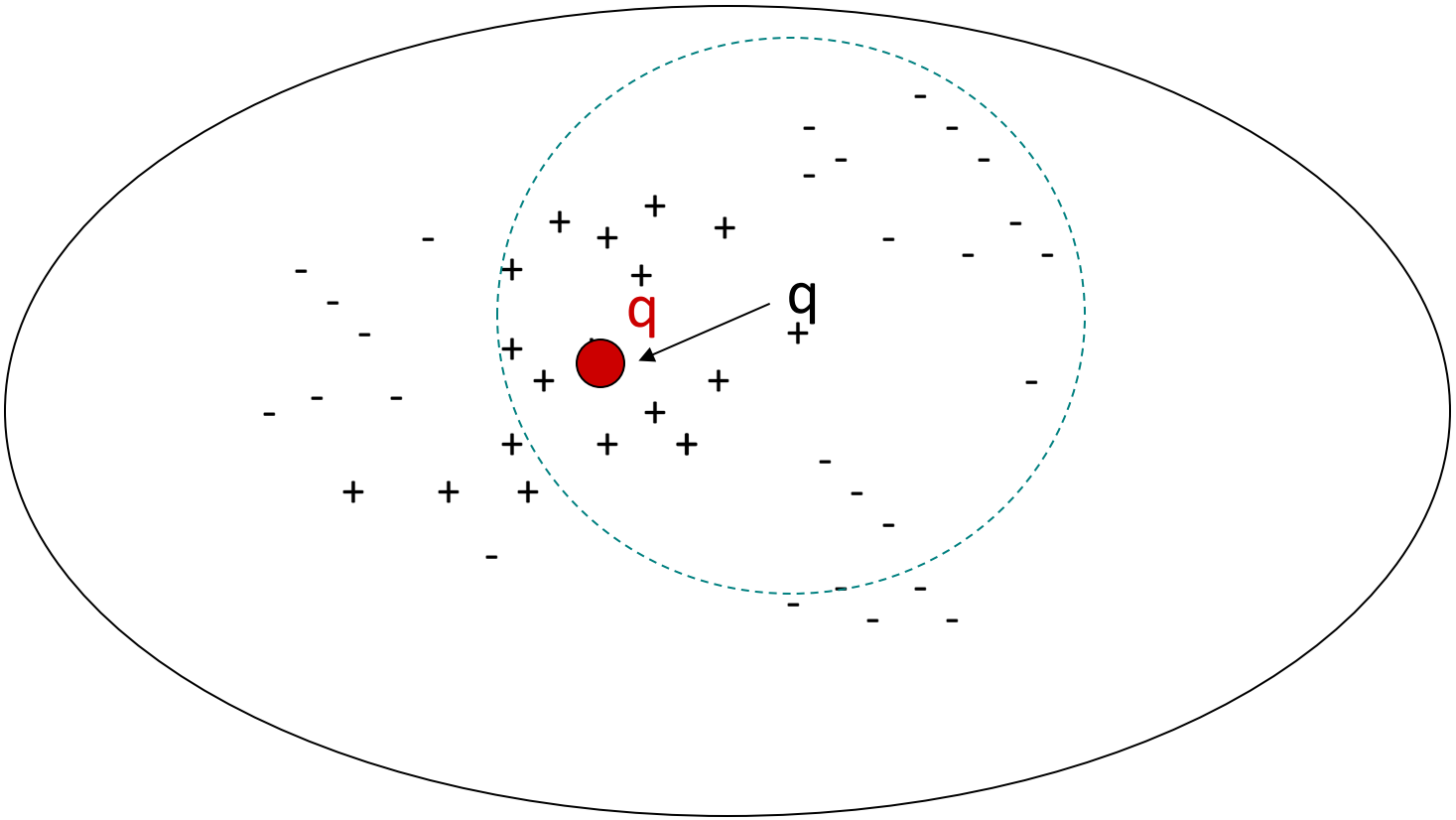
Feedback techniques

- Feedback as query expansion
  - Step 1: Term selection
  - Step 2: Query expansion
  - Step 3: Query term re-weighting
- Feedback as training signal
  - in learning to rank

Relevance feedback in vector space models

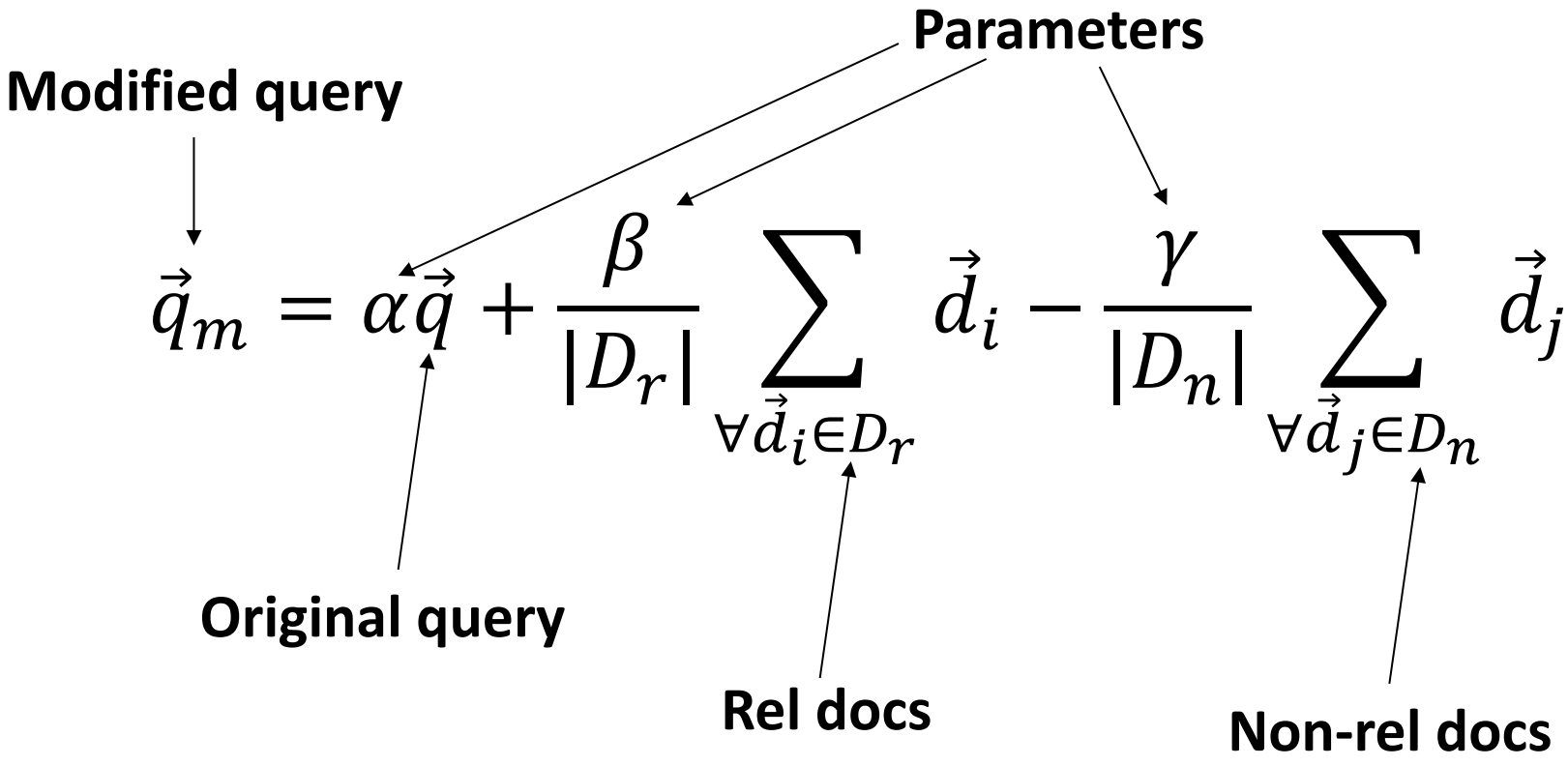
- General idea: query modification
  - Adding new (weighted) terms
  - Adjusting weights of old terms
- The most well-known and effective approach is Rocchio [Rocchio 1971]

Illustration of Rocchio feedback



Formula for Rocchio feedback

- Standard operation in vector space



Rocchio in practice

- Negative (non-relevant) examples are not very important (why?)
- Efficiency concern
  - Restrict the vector onto a lower dimension (i.e., only consider highly weighted words in the centroid vector)
- Avoid “training bias”
  - Keep relatively high weight on the original query
- Can be used for relevance feedback and pseudo feedback
- Usually robust and effective

What you should know

- Purpose of relevance feedback
- pseudo relevance feedback
- Rocchio relevance feedback for vector space models

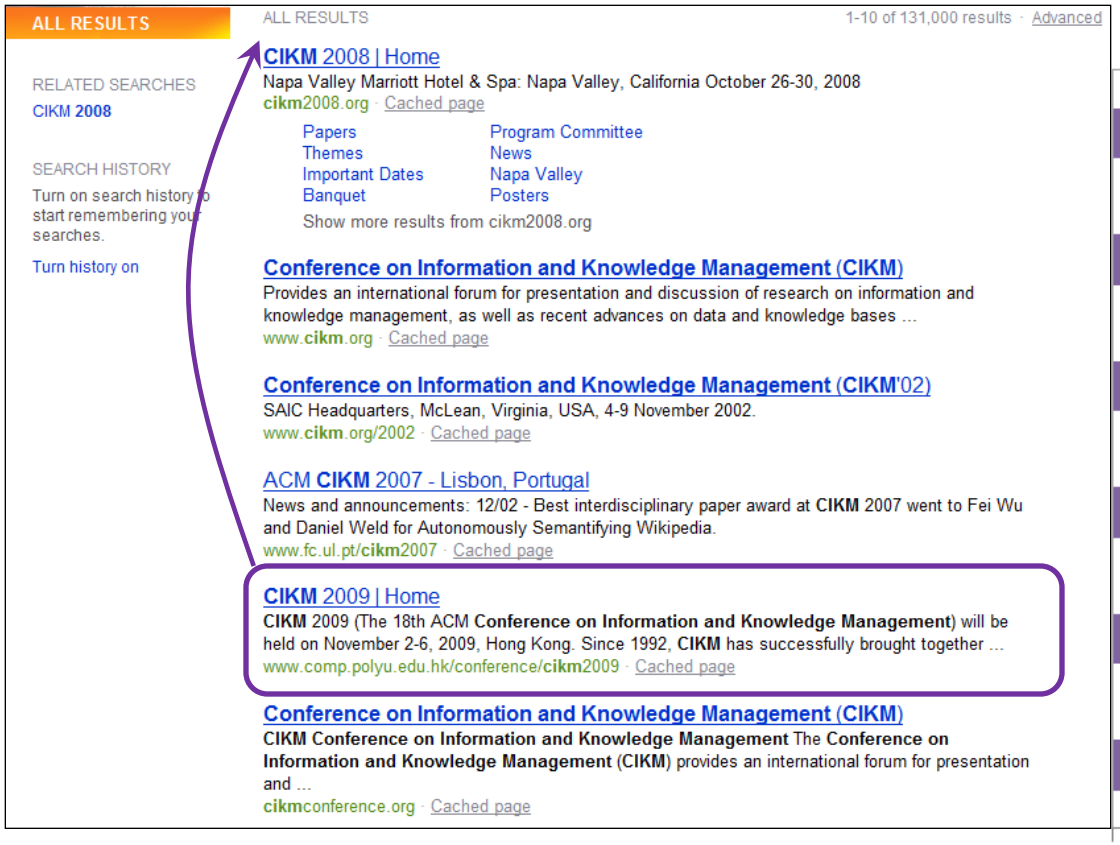
Today's reading

- Chapter 9. Relevance feedback and query expansion
  - 9.1 Relevance feedback and pseudo relevance feedback

Implicit User Feedback
-user clicks

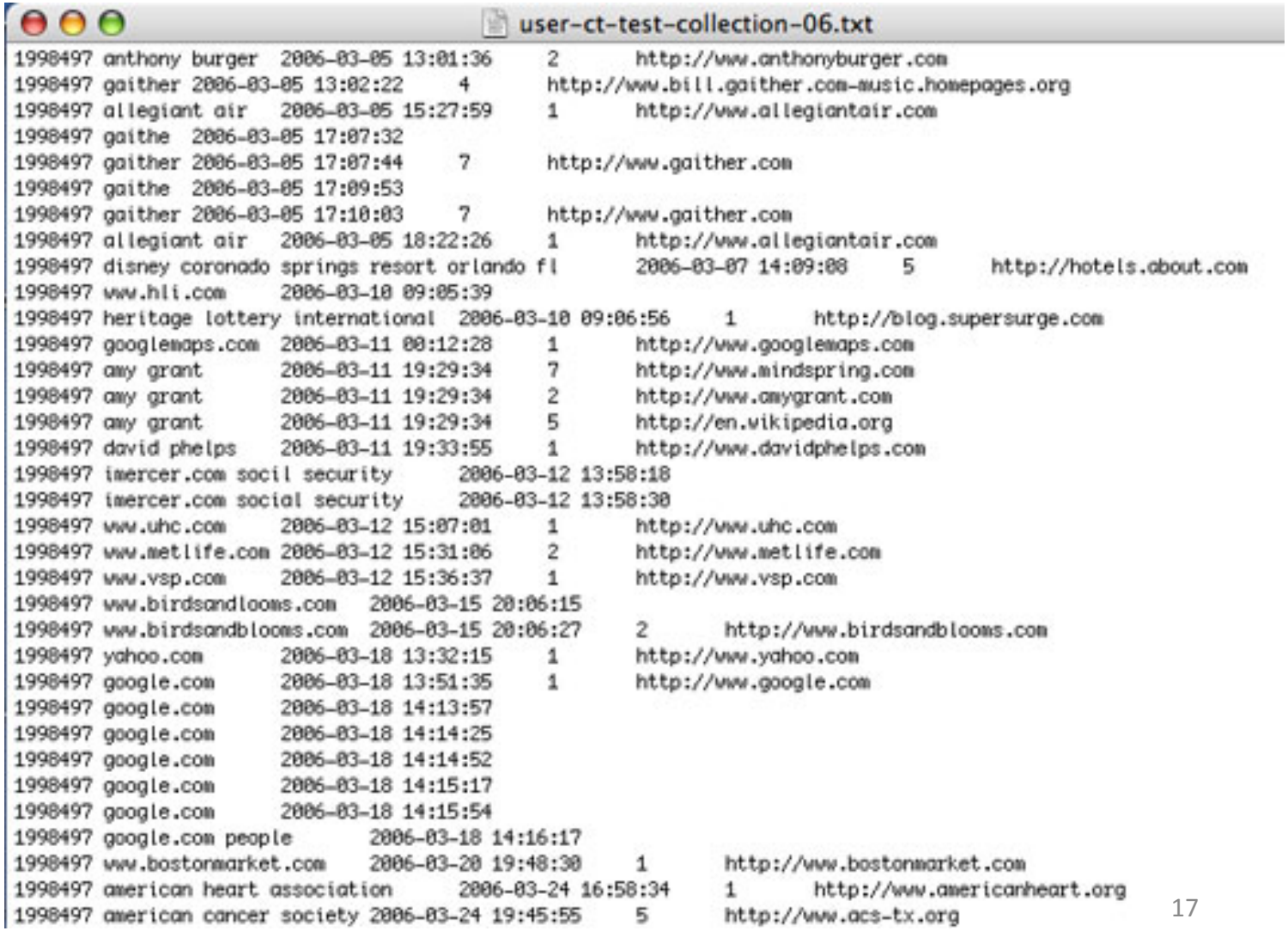
User Behavior

- Adapt ranking to user clicks?



Web search click log

An example

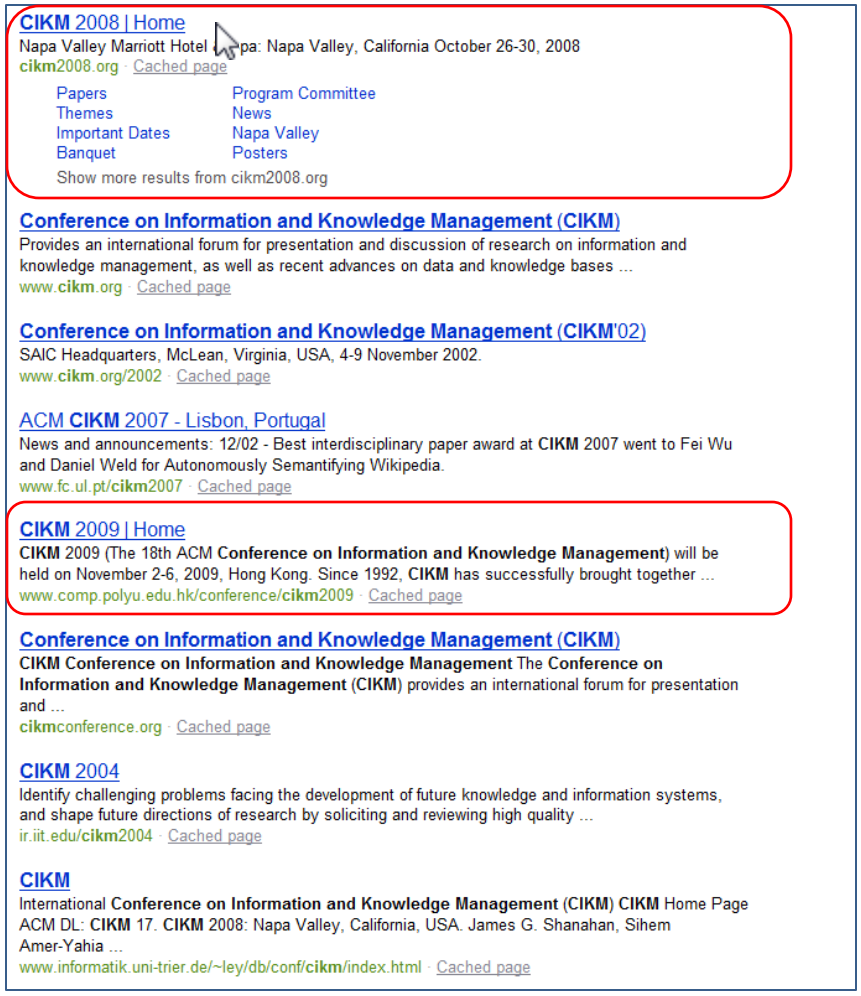


Web Search Click Log

- How large is the click log?
  - bing search logs: 10+ TB/day

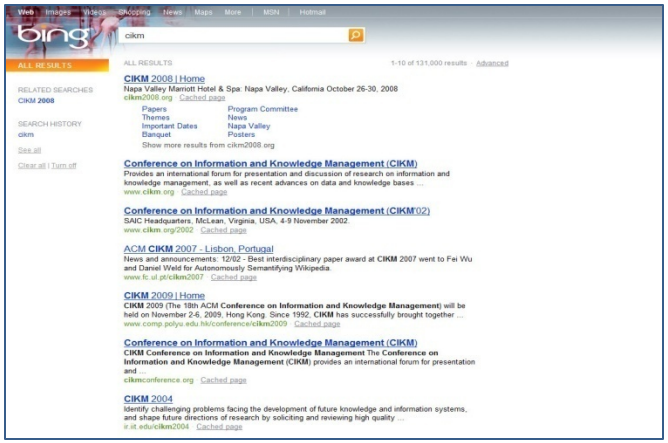
  - In existing publications:
    - [Silverstein+99]: 285M sessions
    - [Craswell+08]: 108k sessions
    - [Dupret+08] : 4.5M sessions (21 subsets * 216k sessions)
    - [Guo +09a] : 8.8M sessions from 110k unique queries
    - [Guo+09b]: 8.8M sessions from 110k unique queries
    - [Chapelle+09]: 58M sessions from 682k unique queries
    - [Liu+09a]: 0.26PB data from 103M unique queries

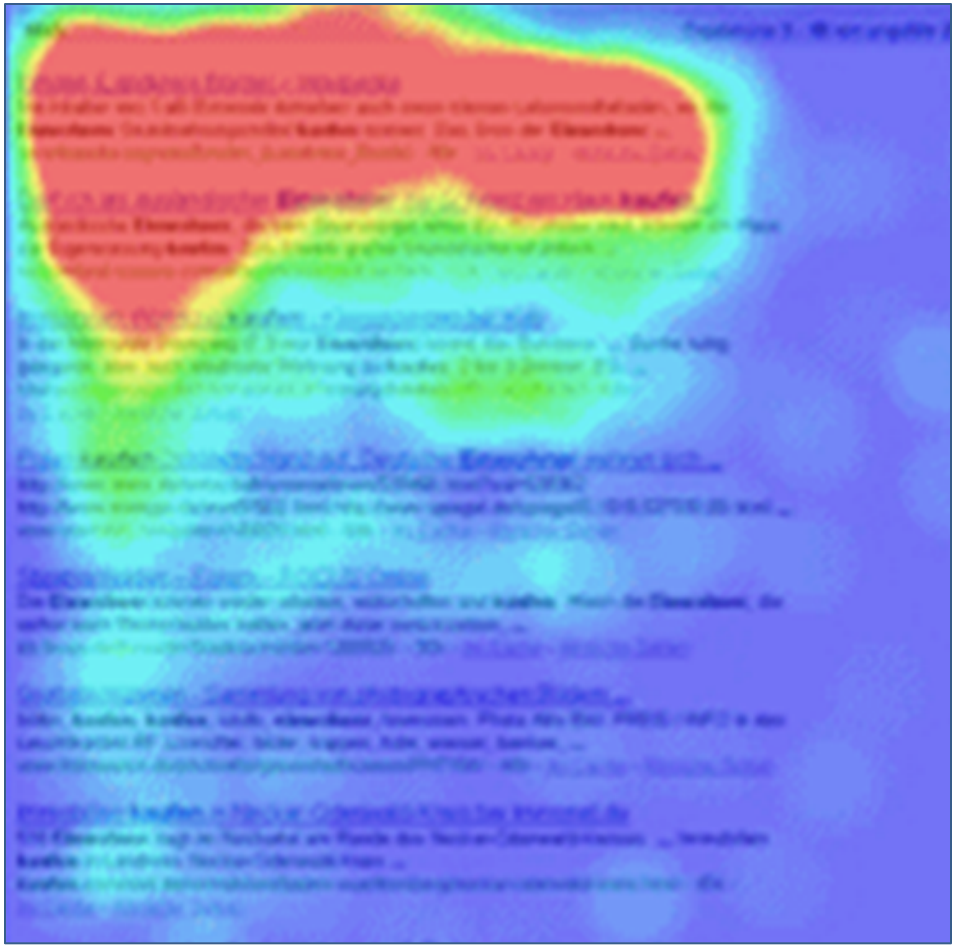
Interpret Clicks: an Example



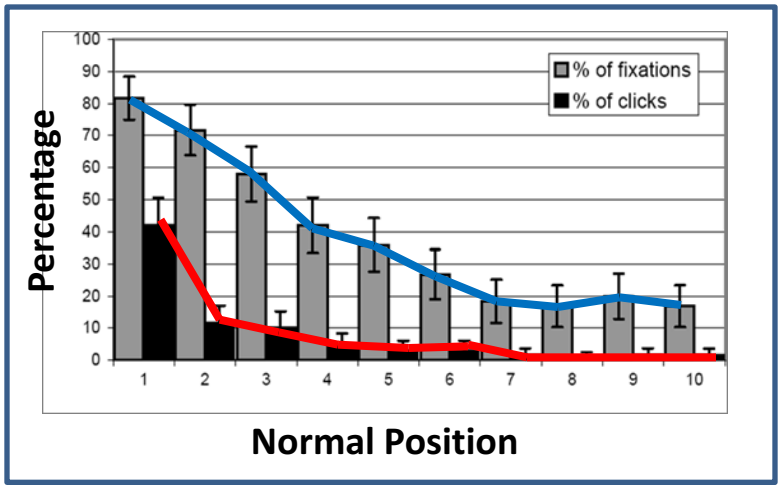
- Clicks are good…
  - Are these two clicks equally "good"?
- Non-clicks may have excuses:
  - Not relevant
  - Not examined

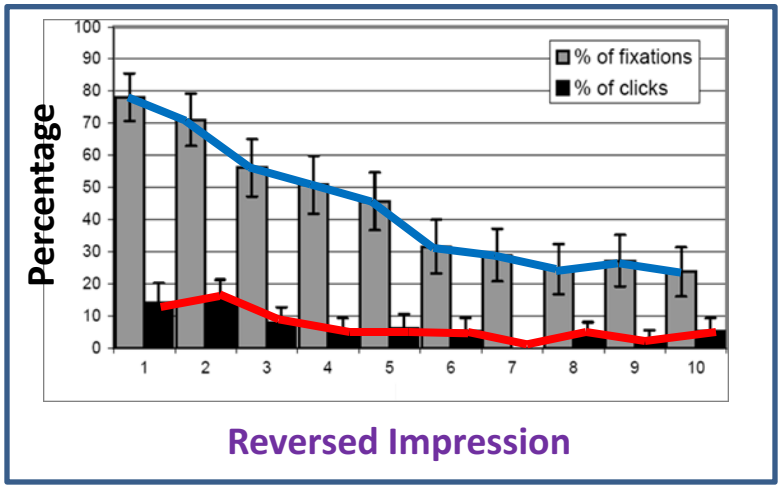
Eye-tracking User Study





Click Position-bias





- Higher positions receive more user attention (eye fixation) and clicks than lower positions.
- This is true even in the extreme setting where the order of positions is reversed.
- "Clicks are informative but biased".

[Joachims+07]

User behavior

- User behavior is an intriguing source of relevance data
  - Users make (somewhat) informed choices when they interact with search engines
  - Potentially a lot of data available in search logs
- But there are significant caveats
  - User behavior data can be very noisy
  - Interpreting user behavior can be tricky
  - Spam can be a significant problem
  - Not all queries will have user behavior

Features based on user behavior

From [Agichtein, Brill, Dumais 2006; Joachims 2002]

- Click-through features
  - Click frequency, click probability, click deviation
  - Click on next result? previous result? above? below>?
- Browsing features
  - Cumulative and average time on page, on domain, on URL prefix; deviation from average times
  - Browse path features
- Query-text features
  - Query overlap with title, snippet, URL, domain, next query
  - Query length

Incorporating user behavior into ranking algorithm

- Incorporate user behavior features into a ranking function like BM25F
  - But requires an understanding of user behavior features so that appropriate V_j functions are used
- Incorporate user behavior features into *learned* ranking function
- Either of these ways of incorporating user behavior signals improve ranking

What you should know

- Clicks as implicit relevance feedback
- Positional bias
- modeling approaches for click models