CS317
Information Retrieval
Week 08

Muhammad Rafi
March 31, 2021

Relevance Feedback & Query Expansion

Chapter No. 9

Agenda

- Problem of IR Systems
  - Recall /Precision
- Query Refinement
  - Global vs Local methods
- Global Methods
- Relevance Feedback
  - Direct / Indirect / Pseudo relevance feedback
- Relevance Feedback in Vector Space
  - Rocchio Algorithm

Agenda

- When Relevance Feedback work
- Relevance Feedback on Web
- Query Expansion
  - Global vs Local
  - Automatic thesaurus generation

Problem with IR Systems

- Same concept may be referred by different words (synonymy), it has an impact on the recall of most information retrieval systems.
  - Users often attempt to address this problem themselves by manually refining a query.
  - Query refinement / expansion
- The methods for tackling this problem split into two major classes:
  - Global Methods
  - Local Methods

Query Refinement: Global Methods

- Global methods are techniques for expanding or reformulating query terms independent of the query and results returned from the query.
- The changes in the query wording will cause the new query to match other semantically similar terms.
  - Query expansion/reformulation with a thesaurus or WordNet
  - Query expansion via automatic thesaurus generation
  - Techniques like spelling correction

Query Refinement: Local Methods

- Local methods adjust a query relative to the documents that initially appear to match the query.
  - Relevance feedback- involve users to get the feedback from the results return for a given query.
  - Pseudo relevance feedback, also known as Blind relevance feedback.
  - (Global) indirect relevance feedback.
- Relevance feedback is one of the most used and most successful approaches.

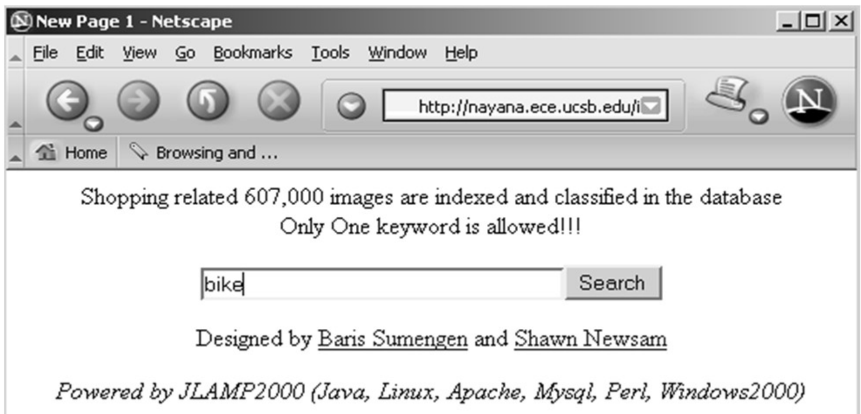
Relevance feedback

- The idea of relevance feedback (RF) is to involve the user in the retrieval process so as to improve the final result set.
  - The user issues a (short, simple) query
  - The system returns an initial set of retrieval results.
  - The user marks some returned documents as relevant or non-relevant.
  - The system computes a better representation of the information need based on the user feedback.
  - The system displays a revised set of retrieval results.

Sec. 9.1.1

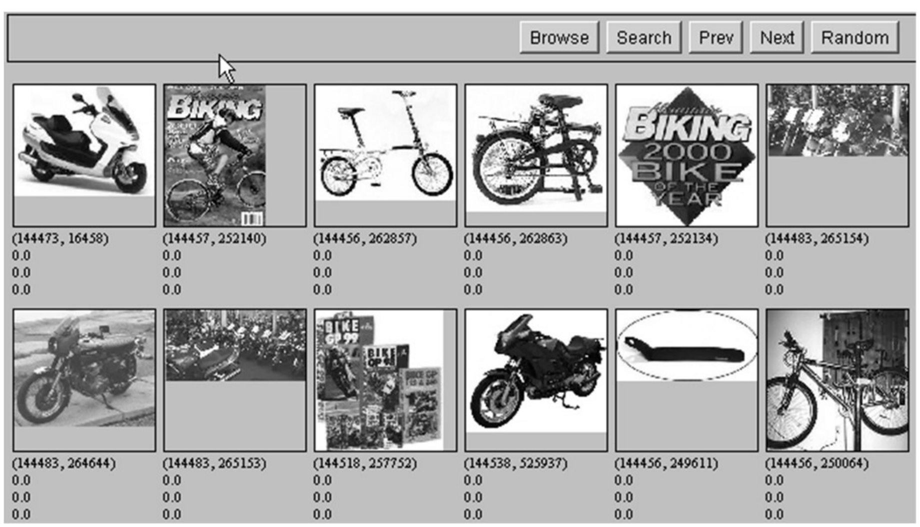
Relevance Feedback: Example

- Image search engine
  http://nayana.ece.ucsb.edu/imsearch/imsearch.html

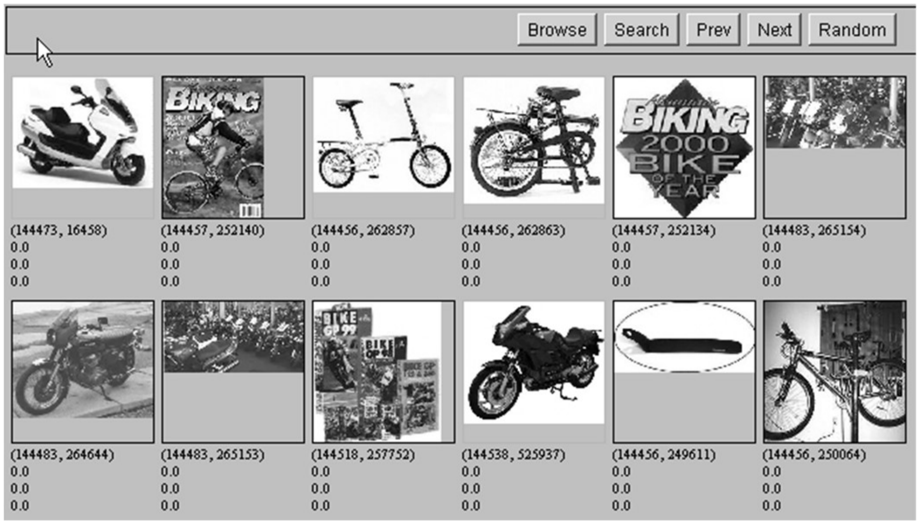


Sec. 9.1.1

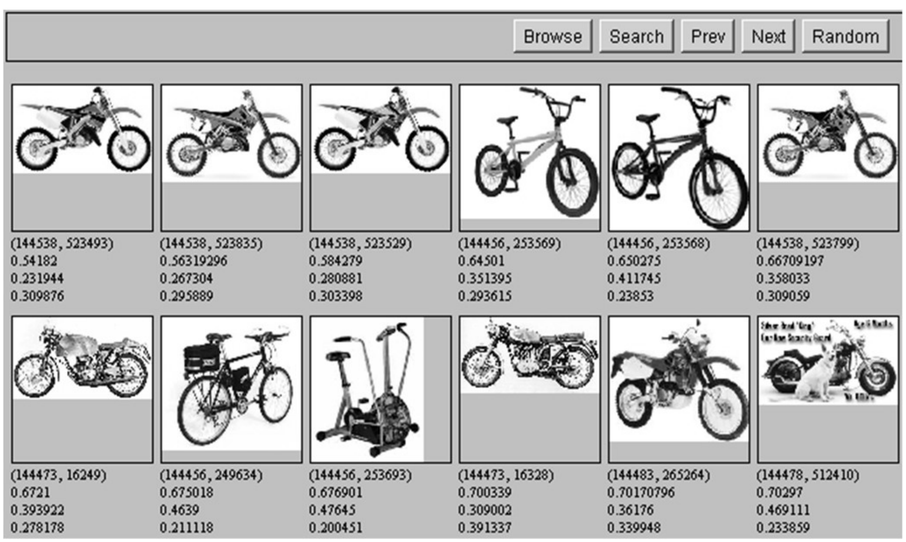
Results for Initial Query



Relevance Feedback



Results after Relevance Feedback



Sec. 9.1.1

Example 2: Initial query/results

- Initial query: *New space satellite applications*

1. 0.539, 08/13/91, NASA Hasn't Scrapped Imaging Spectrometer
2. 0.533, 07/09/91, NASA Scratches Environment Gear From Satellite Plan +
3. 0.528, 04/04/90, Science Panel Backs NASA Satellite Plan, But Urges Launches of Smaller Probes +
4. 0.526, 09/09/91, A NASA Satellite Project Accomplishes Incredible Feat: Staying Within Budget
5. 0.525, 07/24/90, Scientist Who Exposed Global Warming Proposes Satellites for Climate Research
6. 0.524, 08/22/90, Report Provides Support for the Critics Of Using Big Satellites to Study Climate
7. 0.516, 04/13/87, Arianespace Receives Satellite Launch Pact From Telesat Canada
8. 0.509, 12/02/87, Telecommunications Tale of Two Companies +

- User then marks relevant documents with "+".

Sec. 9.1.1

Expanded query after relevance feedback

2.074 new	15.106 space
30.816 satellite	5.660 application
5.991 nasa	5.196 eos
4.196 launch	3.972 aster
3.516 instrument	3.446 arianespace
3.004 bundespost	2.806 ss
2.790 rocket	2.053 scientist
2.003 broadcast	1.172 earth
0.836 oil	0.646 measure

Results for expanded query

1. 0.513, 07/09/91, NASA Scratches Environment Gear From Satellite Plan
2 2. 0.500, 08/13/91, NASA Hasn't Scrapped Imaging Spectrometer
1 3. 0.493, 08/07/89, When the Pentagon Launches a Secret Satellite, Space Sleuths Do Some Spy Work of Their Own
4. 0.493, 07/31/89, NASA Uses 'Warm' Superconductors For Fast Circuit
5. 0.492, 12/02/87, Telecommunications Tale of Two Companies
8 6. 0.491, 07/09/91, Soviets May Adapt Parts of SS-20 Missile For Commercial Use
7. 0.490, 07/12/88, Gaping Gap: Pentagon Lags in Race To Match the Soviets In Rocket Launchers
8. 0.490, 06/14/90, Rescue of Satellite By Space Agency To Cost $90 Million

Pseudo Relevance feedback

- Pseudo relevance feedback, also known as blind relevance feedback, provides a method for automatic local analysis.
- It automates the manual part of relevance feedback, so that the user gets improved retrieval performance without an extended interaction.
- This automatic technique mostly works. Evidence suggests that it tends to work better than global analysis.

Indirect Relevance feedback

- We can also use indirect sources of evidence rather than explicit feedback on relevance as the basis for relevance feedback. This is often called implicit (relevance) feedback.
- Implicit feedback is less reliable than explicit feedback, but is more useful than pseudo relevance feedback.
- Users are often reluctant to provide explicit feedback, it is easy to collect implicit feedback in large quantities for a high volume system, such as a web search engine.

Relevance Feedback in vector spaces

- We can modify the query based on relevance feedback and apply standard vector space model.
- Use only the docs that were marked.
- Relevance feedback can improve recall and precision
- Relevance feedback is *believed* to be most useful for increasing *recall* in situations where recall is important
  - Users can be expected to review results and to take time to iterate

Key concept: Centroid

- The centroid is the center of mass of a set of points
- Recall that we represent documents as points in a high-dimensional space
- Definition: Centroid $$\vec{\mu}(C) = \frac{1}{|C|}\sum_{d \in C} \vec{d}$$

where C is a set of documents.

Rocchio Algorithm

- The Rocchio algorithm incorporates relevance feedback information into the vector space model.

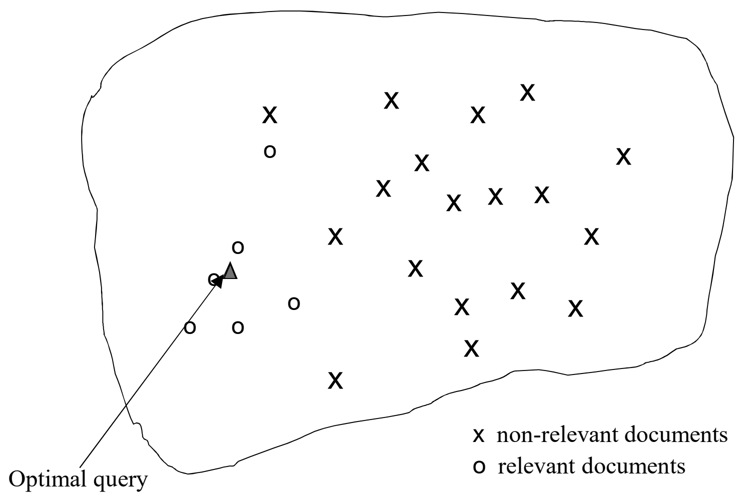
$$\vec{q}_{opt} = \arg\max_{\vec{q}} [\cos(\vec{q}, \vec{\mu}(C_r)) - \cos(\vec{q}, \vec{\mu}(C_{nr}))]$$

- The optimal query vector for separating relevant and non-relevant documents (with cosine sim.):

$$\vec{Q}_{opt} = \frac{1}{|C_r|}\sum_{\vec{d}_j \in C_r} \vec{d}_j - \frac{1}{N - |C_r|}\sum_{\vec{d}_j \notin C_r} \vec{d}_j$$

- Q_{opt} = optimal query; C_r = set of rel. doc vectors; N = collection size; C_{nr} = set of non-rel. doc
- Unrealistic: we don’t know relevant documents.

The Theoretically Best Query



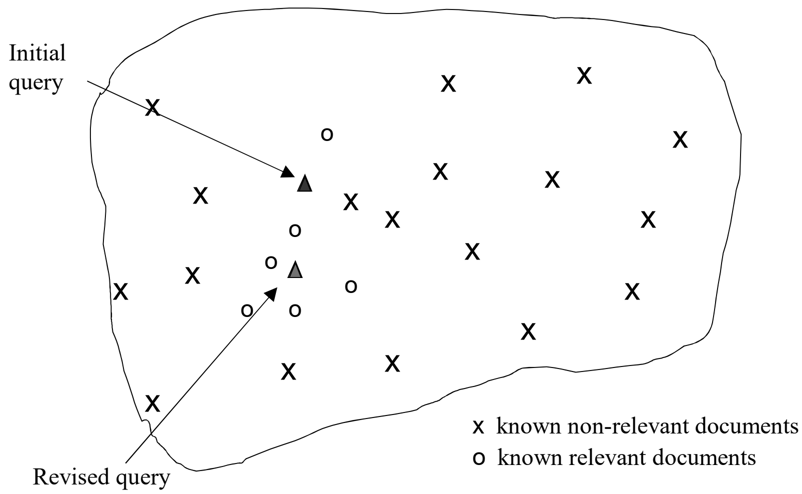
Rocchio 1971 Algorithm (SMART)

- Used in practice:

$$\vec{q}_m = \alpha \vec{q}_0 + \beta \frac{1}{|D_r|} \sum_{\vec{d}_j \in D_r} \vec{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{\vec{d}_j \in D_{nr}} \vec{d}_j$$

- q_m = modified query vector; q_0 = original query vector; α, β, γ: weights (hand-chosen or set empirically); D_r = set of known relevant doc vectors; D_{nr} = set of known irrelevant doc vectors
- New query moves toward relevant documents and away from irrelevant documents
- Tradeoff α vs. β/γ : If we have a lot of judged documents, we want a higher β/γ.
- Term weight can go negative
  - Negative term weights are ignored

Relevance feedback on initial query



23

Sec. 9.1.3

Relevance Feedback: Assumptions

- A1: User has sufficient knowledge for initial query.
- A2: Relevance prototypes are “well-behaved”.
  - Term distribution in relevant documents will be similar
  - Term distribution in non-relevant documents will be different from those in relevant documents
    - Either: All relevant documents are tightly clustered around a single prototype.
    - Or: There are different prototypes, but they have significant vocabulary overlap.
    - Similarities between relevant and irrelevant documents are small

Sec. 9.1.3

Violation of A1

- User does not have sufficient initial knowledge.
- Examples:
  - Misspellings (Brittany Speers).
  - Cross-language information retrieval (hígado).
  - Mismatch of searcher's vocabulary vs. collection vocabulary
    - Cosmonaut/astronaut

Sec. 9.1.3

Violation of A2

- There are several relevance prototypes.
- Examples:
  - Burma/Myanmar
  - Contradictory government policies
  - Pop stars that worked at Burger King
- Often: instances of a general concept
- Good editorial content can address problem
  - Report on contradictory government policies

Sec. 9.1.4

Excite Relevance Feedback

Spink et al. 2000

- Only about 4% of query sessions from a user used relevance feedback option
  - Expressed as “More like this” link next to each result
- But about 70% of users only looked at first page of results and didn’t pursue things further
  - So 4% is about 1/8 of people extending search
- Relevance feedback improved results about 2/3 of the time

Relevance Feedback: Problems

- Long queries are inefficient for typical IR engine.
  - Long response times for user.
  - High cost for retrieval system.
  - Partial solution:
    - Only reweight certain prominent terms
      - Perhaps top 20 by term frequency
- Users are often reluctant to provide explicit feedback
- It’s often harder to understand why a particular document was retrieved after apply relevance feedback

Relevance Feedback

- Approaches
  - Term weighting without query expansion
  - Query Expansion without term weighting
  - Query Expansion with term weighting

Relevance Feedback – Approaches

- Noreault (1979) tried an entirely different approach to relevance feedback. After users had selected a relevant document(s), he used this (these) document(s) as a "new query," effectively ranking (using the cosine correlation as implemented in the SIRE system) all the documents in the collection against this document(s). The top 30 of these retrieved documents were then added to those initially selected by the original query. He found on average that this added 5.1 new relevant documents per query. This feedback method would work with any type of retrieval system such as Boolean systems, provided some method existed for selecting related documents, as the query itself is not modified during the retrieval process.

Relevance Feedback – Approaches

- Attar and Fraenkel (1981) used an approach based on local feedback only. They produced an ordered list of the terms in all the documents retrieved in a first iteration search, with the list ordered by the frequency of a term in that set and by its "distance" from the initial query. These terms were shown to a user for selection as new query terms or sifted by an automatic procedure for addition to the query. They found that both expert and nonexpert users could correctly select terms from the suggested list, but that their automatic sifting mechanism could not do this well. Note that terms are selected from all retrieved documents, not just relevant ones, so that this technique could be used as a feedback technique for queries retrieving no relevant documents on the first iteration.

Relevance Feedback – Approaches

- Dillon et al. (1983) used a hybrid approach to provide relevance feedback to a Boolean environment. They devised a new query based only on terms from previously retrieved documents, with terms weighted by a formula similar to the term precision formula used by Salton (but with significant differences). The weighted terms were not used, however, to produce a ranked list of retrieved documents, but were used to automatically construct a revised Boolean query. Results suggest that this method can be effective if very careful attention is paid to the construction of the revised Boolean query.

Relevance feedback summary

- Relevance feedback has been shown to be very effective at improving relevance of results.
- Its successful use requires queries for which the set of relevant documents is medium to large.
- Full relevance feedback is often onerous for the user, and its implementation is not very efficient in most IR systems.

Relevance feedback summary

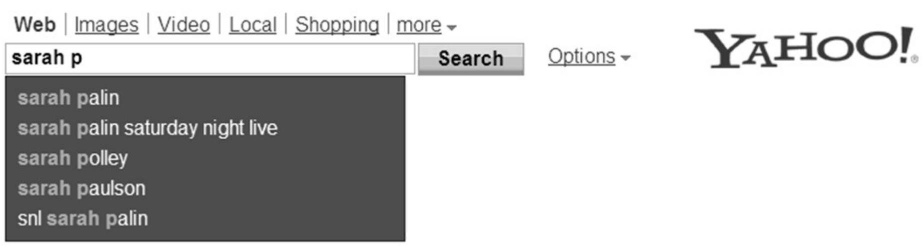
- Other use of relevance feedback
  - Following a changing information need (e.g., names of car models of interest change over time)
  - Maintaining an information filter (e.g., for a news feed).
  - Active learning (deciding which examples it is most useful to know the class of to reduce annotation costs).

Sec. 9.2.2

Query Expansion

- In relevance feedback, users give additional input (relevant/non-relevant) on documents, which is used to reweight terms in the documents
- In query expansion, users give additional input (good/bad search term) on words or phrases
- Query expansion help in extending user queries with related terms in order to solve the lexical gap problem in Information Retrieval

Query assist



Would you expect such a feature to increase the query volume at a search engine?

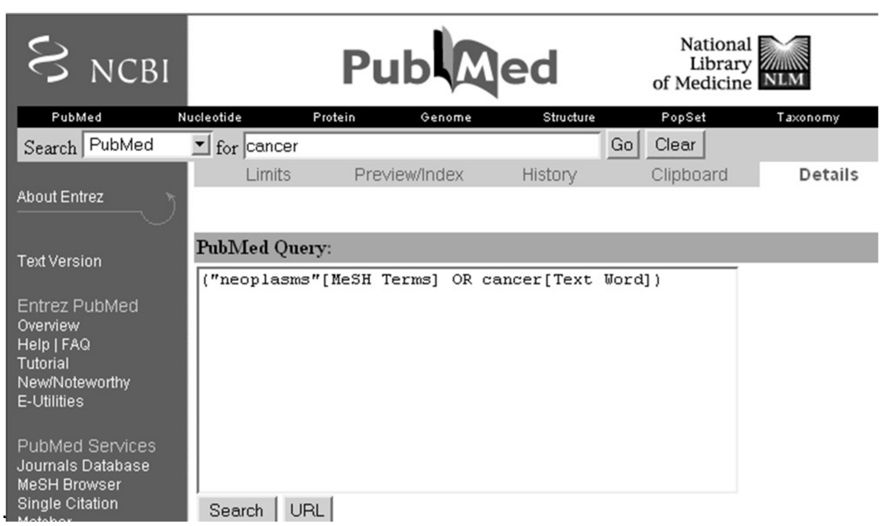
Sec. 9.2.2

How do we augment the user query?

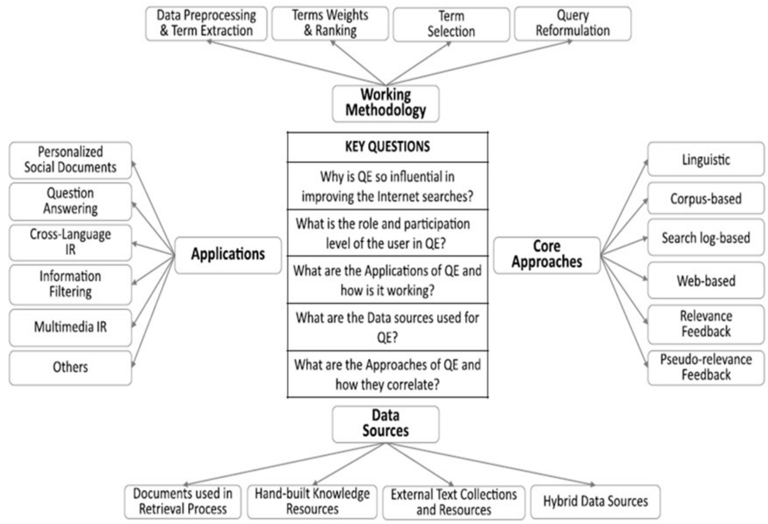
- Manual thesaurus
  - E.g. MedLine: physician, syn: doc, doctor, MD, medico
  - Can be query rather than just synonyms
- Global Analysis: (static; of all documents in collection)
  - Automatically derived thesaurus
  - Refinements based on query log mining
- Local Analysis: (dynamic)
  - Analysis of documents in result set

Sec. 9.2.2

Example of manual thesaurus



Relevance Feedback



Conclusion

- Relevance Feedback is an important part of learning the query intent from a user.