

Relevance Feedback, Query Expansion, and Inputs to Ranking Beyond Similarity

Lecture 10
CS 410/510
Information Retrieval on the Internet

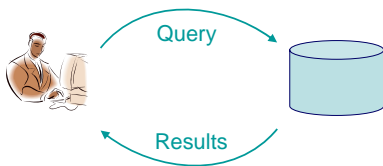
Outline

- Query reformulation
- Sources of relevance for feedback
- Using relevance feedback
 - Vector model
 - Probabilistic model
- Clustering approaches
 - Local analysis
 - Global analysis
- Examples of relevance feedback and query expansion
- Inputs to document ranking other than similarity

CS 510 Winter 2007

2

The basic problem



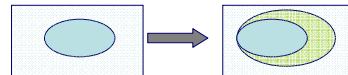
Queries are not always successful
New queries do not learn from previous failure

CS 510 Winter 2007

3

Possible solutions

- Retrieve new documents



– Add new terms to the query (*query expansion*)

- Re-order the documents



– Re-weight terms

CS 510 Winter 2007

4

How?

- Where do we get information for adding query terms or re-weighting documents?
- Use relevant documents to help find more relevant documents
- Assumption:
 - Relevant documents are more similar to each other than they are to non-relevant documents

CS 510 Winter 2007

5

How?

How do we get relevant documents?

1. Ask the user
 - System returns documents for initial query
 - User marks which ones are relevant
 - System uses as input to new query
 - Assumption: user will recognize documents that are useful, or close to what he wants
2. Assume top-ranked documents are relevant; automatically reformulate query

CS 510 Winter 2007

6

