

Finding Related Queries

David G. Deschenes

October 29, 2002

WWW Search Engines

Introduction & Background

Query Suggestion

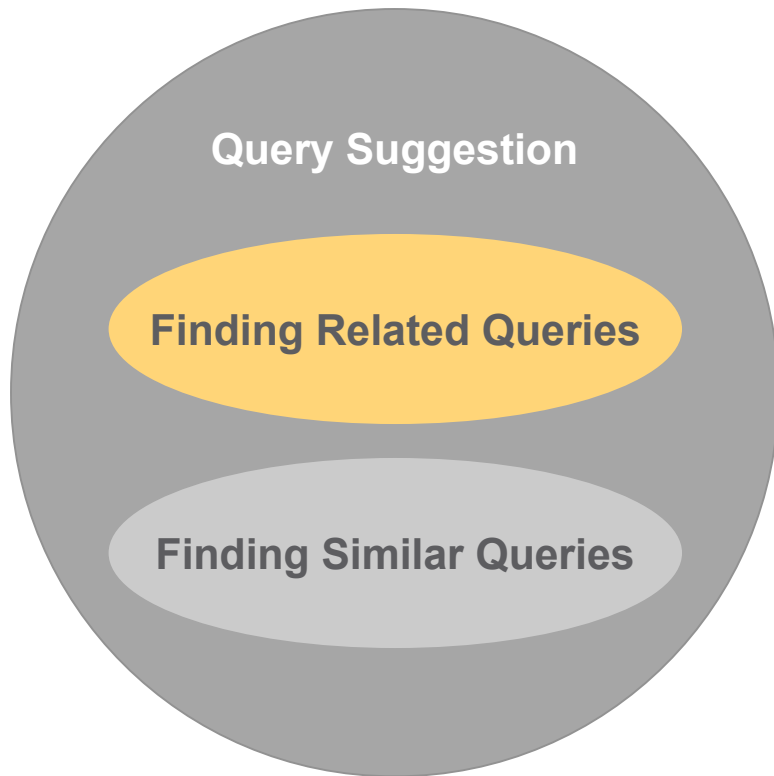
» Usage

- Help searchers refine their information needs
- Reduce time needed to fulfill information needs

» Commercial Implementations

- AltaVista Prisma
- Teoma Refine
- Lycos Suggested Search
- Ask Jeeves Related Search Terms
- DogPile & InfoSpace “Are you looking for?”

Query Suggestion Concepts

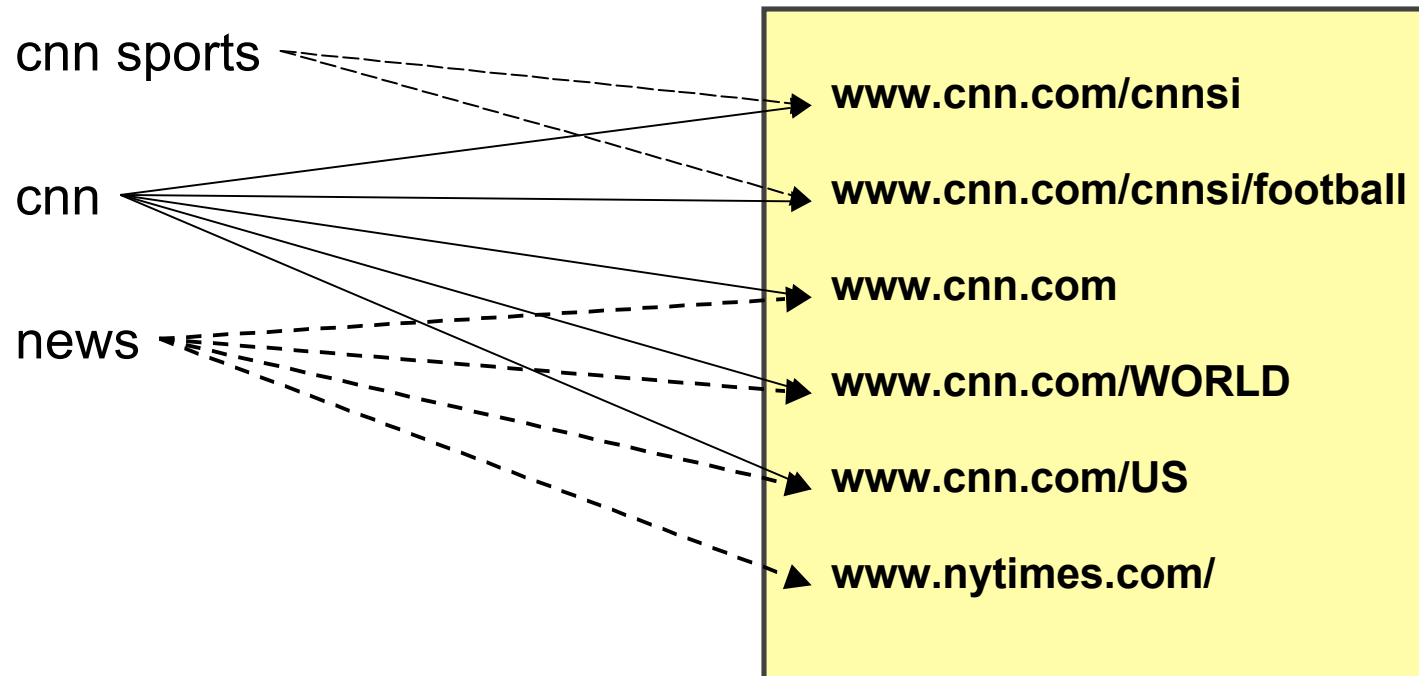


- » **Similar Queries**
 - Synonym Replacement
 - Word Co-occurrence
 - Spelling Suggestion
 - Textual Similarity
- » **Related Queries**
 - Shared Results

Uses of Related Queries

- » Aid web site designers and managers in determining which queries should, but do not successfully result in their sites
- » Provide searchers with alternate queries
- » Supply top queries for a given URL

How it should work?



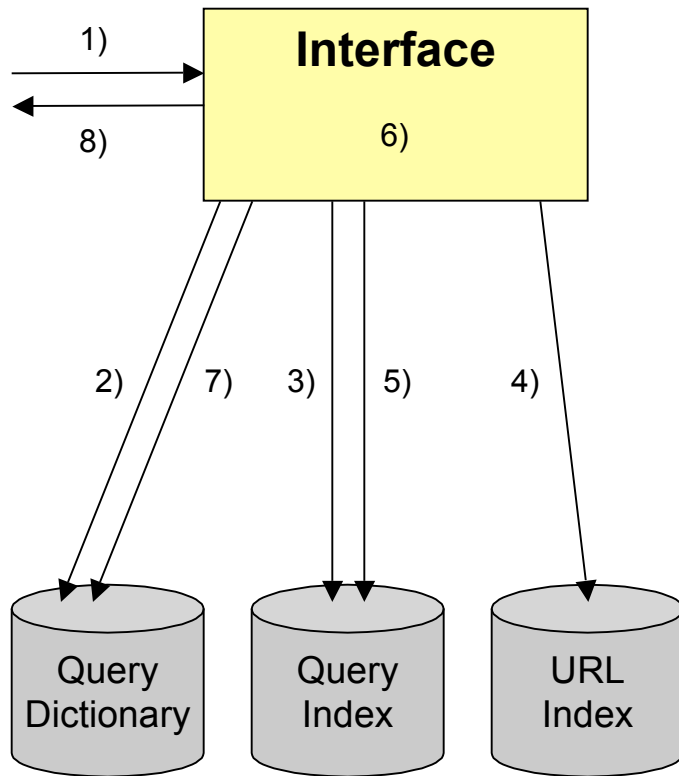
- » “cnn” and “news” are more strongly related
- » “cnn” and “cnn sports” are less strongly related
- » “cnn sports” and “news” are not related

Motivation

- » Little existing exploration into the discovery of related queries
- » Interesting application of search engine techniques
- » Potential to produce a useful tool
- » May at some point be integrated into a search engine

System Design

Design Overview



- 1) User submits query
- 2) Retrieve query ID
- 3) Retrieve URL IDs for query ID
- 4) Retrieve Query IDs for each retrieved URL ID
- 5) Retrieve URL Ids for each retrieved Query ID
- 6) Rank retrieved queries
- 7) Translate retrieved query IDs
- 8) Return ranked queries to user

Ranking Process

- » Each time a query, which results in a URL that is in the result set of the submitted query, is seen its rank value is incremented
- » The top ten queries are then sorted according to rank value and returned to the interface
- » Frequency across the entire query data set might also be used to weight queries during the ranking process

Inverted Indices

» Query Inverted Index

- Maintains Query \square URL Mappings

» URL Inverted Index

- Maintains URL \square Query Mappings

» Details

- Sorted binary files mapping object IDs
- Indices are loaded into memory and lookups are done using a binary search
- Results are retrieved from disk at lookup time

Indexed Data

» Excite

- Provided trace of approximately 1.2M unique queries

» Google

- Retrieving top 10 results for each of the queries in the Excite trace
- Retrieving in bunches of 1000 queries
- Using Google Web APIs

Comments, Questions?