

Information Retrieval

CISC437/637, Lecture #23

Ben Carterette

Copyright © Ben Carterette

1

A Database for the WWW

- Imagine the web stored in a relational DBMS
- Three challenges:
 - How would we search it?
 - How would we know whether we found what we wanted?
 - How would we avoid getting bad results?

Copyright © Ben Carterette

2

Text Search

- Most-accessed field would be long textual information
 - News articles, patents, web pages, books, ...
- What is the best way to search within this data?
 - grep for keywords in it?
 - Store it in a DBMS and use SQL queries with ... LIKE '%keyword1%' AND LIKE '%keyword2%' ...?
- If there's enough data, these approaches are very slow and very inefficient
 - Hash and B+-tree indexes don't help

Copyright © Ben Carter@cs

3

Relevance

- In a DBMS, records either match a SQL query or they don't—there is no middle ground
- In free-text search, some documents can be better matches than others
 - Matching documents may not be relevant
 - Documents that don't match may be relevant
- **Relevance** describes the usefulness of a document to a particular user

Copyright © Ben Carter@cs

4

query

relevant results (for one user)

top-ranked results

relevant results (for a different user)

relevance

About 14,000,000 results (0.16 seconds)

Relevance - Wikipedia, the free encyclopedia
Relevance is a term used to describe how pertinent, connected, or applicable something is to a given matter. A thing is relevant if it serves as a means to ...
Epidemiology - Politics - Logic - Economics
en.wikipedia.org/wiki/Relevance - Cached - Similar

Relevance (information retrieval) - Wikipedia, the free encyclopedia
In the context of information science and information retrieval, relevance describes how well a retrieved set of documents (or a single document) meets the ...
en.wikipedia.org/wiki/Relevance_(information_retrieval) - Cached - Similar
Show more results from en.wikipedia.org

Relevance, Inc | We Know Software | Think Relevance
Mar 2, 2010 ... Relevance, Inc. - A software company in Durham, NC ... We are very pleased to announce three new additions to the Relevance team ...
thinkrelevance.com/ - Cached - Similar

Relevant | Define Relevant at Dictionary.com
bearing upon or connected with the matter in hand; pertinent; a relevant remark. Use relevant in a sentence - See images of relevant ...
dictionary.reference.com/browse/relevant - Cached - Similar

Relevance - Definition and More from the Free Merriam-Webster ...
Definition of relevance from the Merriam-Webster Online Dictionary with audio pronunciations, thesaurus, Word of the Day, and word games.
www.merriam-webster.com/dictionary/relevance - Cached - Similar

RELEVANT Magazine - Covering God, Life and Progressive Culture
May 8, 2010 ... So join our RELEVANT Advisory Board. You'll be included in occasional surveys and will get samples and other free stuff from time to time ...
www.relevantmagazine.com/ - Cached - Similar

Relevance: communication and cognition - Google Books Result
Dan Sperber, Debra Wilson - 1995 - Psychology - 326 pages
This revised edition includes a new Preface outlining developments in Relevance Theory since 1985, discussing the more serious criticisms of the theory and ...
books.google.com/books?id=m8Q1148704... - Cached - Similar

Ted Vinzani (relevance) on Twitter
May 8, 2010 ... (to @vinzani) Are you satisfied w/ Your Sales & Profitability? Relevance (edit) - 32 yrs Sales/Marketing Pro- Relevance (tag & Sales Psychology) - 28 ...
twitter.com/relevance - Cached - Similar

churchrelevance.com
The Visual Metaphor by David McCordless, Disclaimer: Church Relevance received a complimentary review book, and I'd post some Amazon affiliate links ...
churchrelevance.com/ - Cached - Similar

Relevance - Making Stuff That Matters - Google Books Result

Spam

- It is very easy for anyone to get new records into our WWW database
 - Just put a page on the web and get it crawled
- Therefore it's very easy to spam results
 - Just put the query terms you want to spam somewhere in your page's content
- Over time this effectively makes a DBMS useless