Tightly Coupling Search and Structure

Susan T. Dumais
Microsoft Research

SIGIR’97 - Workshop on Information Reduction
July 31, 1997
Search and Structure

• Two Common Methods for Info Access
  – Search (Querying)
  – Structure (Navigation)

• Search -> Ranked List or Set of Results
  – e.g., Excite, Infoseek, Lycos, Alta Vista

• Structure -> Navigable, Browable Structure
  – e.g., Yahoo! … Dewey, LC, MeSH

• Both Methods Have Known Access Problems
Combining Search and Structure

• Separate-But-Equal
  – Both Search and Structure available, but not integrated

• Tightly Coupled Integration
  – Very powerful notion; limited exploration of rich design space; some possibilities:
    • Structure Search Results
    • Show Search Results in Context of Structure
Separate-But-Equal

• Provide Two Alternative Methods
• User Can Explore Ranked Search Results OR Topic Hierarchy
  – e.g., Excite, Infoseek, Lycos
82% Register of WWW Search Engines
More Like This: Click here for a list of documents like this one.
URL: http://coombs.anu.edu.au/CoombswebPages/SearchEngines.html

82% Publicize your home page using NTG Inter...
More Like This: Click here for a list of documents like this one.
URL: http://www.ntg-campus.com/ntg/public.htm
Summary: Announce-A-Site is a new service which submits a site's information to 20 different Indexes, Directories, What's New Lists for one low fee. New on the Net is a new service which submits a site's information to 20 different Indexes, Directories,
You searched for search engines

Sites 1 - 10 of 7,915,264

You are on the first page of search results. There are 7,915,264 sites, and you are viewing 10 of them. The first site is number 1.

Related Topics
Search engines
Internet directories
Searching & exploring the Internet

Tips
- Search only these results
- Search the whole Web

Sites 1 - 10 of 7,915,264

InterNIC Meta-Search Engines
The following list includes a number of sites that query multiple search engines simultaneously, and sites that provide direct access to multiple search engines. It was last updated...
1) **ELImnet Search Engines (에밀넷 검색 엔진)**

For 0.2% ¼ ½έ¶¶ Kor-Seek, ¼ ½έ¶¶ Mr. ‘UÃ£º À½£À½ £À½ Yahoo! Inf Lycos
Magellan Excite Open Text Index The Electric Labrary CNET's Shareware.com
AccuFin...
http://www.elim.co.kr/search/ [100%, 2 of 2 terms]

2) **Spiders, Crawlers and other Indexes to the Internet**

Spiders, Crawlers and other Indexes to the Internet Contents of this Page
Selected Indexes, Search Engines and Directories Indexes to Mailing Lists Le...
http://www.cs.unca.edu/~davidson/indexes.html [88%, 2 of 2 terms]

3) **Chez Marco's Search-The-WWW page**

Search the WWW! When you are looking for a WWW page on a general topic
Tightly Coupled Integration

• Structure Search Results … going beyond the ranked list
  – e.g., Scatter/Gather; Excite; Document “maps”
• Search Results Are Shown in the Context of an Organizing Structure (focus today)
  – e.g., Yahoo!, SuperBook, AMIT, Cat-A-Cone
Found 5 Category and 161 Site Matches for information retrieval.

Yahoo! Category Matches  (1 - 5 of 5)

Business and Economy: Companies: Information, Databases: Online Information Retrieval Services
Reference: Libraries: Information Science: Information Retrieval

Regional: Countries: Australia: Business: Companies: Information, Databases: Online Information Retrieval Services
Regional: Countries: United Kingdom: Reference: Libraries: Information Science: Information Retrieval

Yahoo! Site Matches  (1 - 15 of 161)

Business and Economy: Companies: Information, Document Delivery

- Access/Information - provides research, information-retrieval, and document delivery. Specializes in business, market background, environmental, and high-tech research, court monitoring, and records management.

Reference: Libraries: Information Science

- CATaloguing and Retrieval of Information Over Networks Applications(CATRIONA) - feasibility study to investigate the requirements for the development of applications programs and procedures to facilitate, using classification, and
13. Hypertext Spanning the Internet: <WWW>

The World-Wide Web, or WWW, is the newest information service to arrive on the Internet. The Web is based on a technology called hypertext. Most of the development has taken place at CERN, the European Particle Physics Laboratory; but it would be a mistake to see the Web as a tool designed by and for physicists. While physicists may have paid for its initial development, it’s one of the most flexible tools—probably the most flexible tool—for prowling around the Internet. Like Gopher and WAIS, the Web is very much under development, perhaps even more so. So don’t be surprised if it doesn’t occasionally work the way you’d like. It’s certainly worth playing with.

To try the Web, telnet to info.cern.ch. This will automatically drop you into a public-access client program (or browser, to use the Web’s terminology). This is a “line-oriented” browser that will work with a traditional terminal. Several other browsers are available; if you decide to install your own (and that’s highly recommended if you want to use the Web frequently), you can choose between the line-oriented browser, several browsers for the X Window System (the one called “Viola” or “ViolaWWW” is probably the most feature-rich), the NeXT UNIX workstations, the Macintosh, and PCs.

What Is Hypertext?

Hypertext is a method of presenting information where selected words in the text can be “expanded” at any time to provide other information about the word. That is, these words are links to other documents which may be text, files, pictures, anything. For the sake of illustration, let’s assume that your library had a hypertext card catalog. If you
Benefits of Integration

• “Structure-Guided Search” - Selection of relevant matches guided by context
• “Search-Guided Structure” - Structure view depends on query (e.g., fisheye)
• Other combinations?
• Mitigates some key IR problems
  – Word sense disambiguation
  – Recall enhancement
Open Issues

• Generating structures
• Scalability
• Integration with other sources of information - e.g., usage communities, collaboration, temporal, attributes
• Evaluation - need to go beyond static relevance judgements a la SuperBook expts