

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



Back



Forward



Stop



Refresh



Home



Search



Favorites



Print



Font



Mail

Links



Address http://www.excite.com/search.gw?trace=a&search=search+engines

excite

Search Results

Home

Search

[Win a '98 VW Bug!](#) · [Free Email](#) · [Yellow Pages](#) · [My Channel](#) · [Chat](#) · [Stock Quotes](#) · [Travel](#)Elizabeth@... Good night, green grass
Jason@... Good night, mail boxKnow when your family and
friends are online, and
contact them **INSTANTLY!**[Search Tips](#)
[Power Search](#)
[Search the News](#)

search engines

Search

[Help](#)[Select words
to add to
your search..](#)

engine

database

searching

servers

lists

einet

aliweb

gopher

searches

submit

Try these first:

[Web sites about Arts and Humanities > Filmmaking](#)[Web sites about ... > Directories and Searching](#)Top 10 matches. [About Your Results?](#)[Show Titles only](#) [List by Publisher](#)**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>*Summary:* Also, parallel search in USENET News groups databases: Alta Vista News; Dejanews; Excite News]. Search the SUSI databases suite [Simple Unified Search Interface] (Nexor,UK).**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,



Address http://www.infoseek.com/Titles?qt=search+engines&col=WW&sv=IS&k=noframes&nh=10



infoseek®



You searched for **search engines**

Sites 1 - 10 of 7,915,264

 [Looking for the Best Exclusive Sports Coverage? Click Here...CBS]

[Click here now!](#)

• [news center](#)

[NRTC Invests in New Internet Search Engine](#)

[Inktomi and UOL](#)

[Partner to Develop First High Performance Search Engine for South American and](#)

• [smart info](#)

- [People & Business](#)
- [Stocks/Companies](#)
- [Street Maps](#)
- [Shareware/Chat](#)
- [Desk Reference](#)
- [Infoseek's Investor](#)

Related Topics

- [Search engines](#)
- [Internet directories](#)
- [Searching & exploring the Internet](#)

Search input field

seek

[Tips](#)

Search **only** these results Search **the whole Web**

Sites 1 - 10 of 7,915,264 [Hide Summaries](#) [next 10](#)

[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 ...

85% <http://ds2.internic.net/tools/meta.html> (Size 11.7K)

[Search Engine Tutorial](#)

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



Address <http://www.yahoo.com/>



YAHOO!



[Today's Web Events](#)
[Tour de France](#)




[Yahoo! Chat](#)

[options](#)

[Yellow Pages](#) - [People Search](#) - [Maps](#) - [Classifieds](#) - [News](#) - [Stock Quotes](#) - [Sports Scores](#)

- [Arts and Humanities](#)
[Architecture](#), [Photography](#), [Literature](#)...
- [Business and Economy \[Xtra!\]](#)
[Companies](#), [Investing](#), [Employment](#)...
- [Computers and Internet \[Xtra!\]](#)
[Internet](#), [WWW](#), [Software](#), [Multimedia](#)...
- [Education](#)
[Universities](#), [K-12](#), [College Entrance](#)...
- [Entertainment \[Xtra!\]](#)
[Cool Links](#), [Movies](#), [Music](#), [Humor](#)...
- [Government](#)
[Military](#), [Politics \[Xtra!\]](#), [Law](#), [Taxes](#)...
- [Health \[Xtra!\]](#)
[Medicine](#), [Drugs](#), [Diseases](#), [Fitness](#)...
- [News and Media \[Xtra!\]](#)
[Current Events](#), [Magazines](#), [TV](#), [Newspapers](#)...
- [Recreation and Sports \[Xtra!\]](#)
[Sports](#), [Games](#), [Travel](#), [Autos](#), [Outdoors](#)...
- [Reference](#)
[Libraries](#), [Dictionaries](#), [Phone Numbers](#)...
- [Regional](#)
[Countries](#), [Regions](#), [U.S. States](#)...
- [Science](#)
[CS](#), [Biology](#), [Astronomy](#), [Engineering](#)...
- [Social Science](#)
[Anthropology](#), [Sociology](#), [Economics](#)...
- [Society and Culture](#)
[People](#), [Environment](#), [Religion](#)...



[Stock Quotes](#) - [NFL Preseason](#) - [Yahoo! Chat](#) - [Tour de France](#) - [Exploring Mars](#)

[Categories](#) - [Sites](#) - [AltaVista Web Pages](#) | [Headlines](#) | [Net Events](#)

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**](#)

[Regional: Countries: United Kingdom: Business: Companies: **Information**: Databases: Online **Information Retrieval** Services](#)

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 online classification and

- 51 The Whole Internet User's Guide & Catalog
- 2 +Preface
- 1 +1. What Is This Book About?
- +2. What Is the Internet?
- +3. How the Internet Works
- +4. What's Allowed on the Internet?
- +5. Remote Login
- +6. Moving Files: ftp
- +7. Electronic Mail
- +8. Network News
- +9. Finding Software
- +10. Finding Someone
- 1 +11. Tunneling through the Internet: Gopher
- 1 +12. Searching Indexed Databases: <WAIS>
- 39 > 13. Hypertext Spanning the Internet: <WWW>
- 4 Getting Started
- 7 The Web and Gopher
- 4 +Using WAIS Resources
- 10 +Other Internet Resources
- 5 +Navigating the Web
- 6 Where the Web Is Going
 - Other Kinds of Documents
 - Making Your Own Links
 - 1 Collaborating with Others
 - 1 Hypertext Editors
 - 2 How Can I help?
- +14. Other Applications
- +15. Dealing With Problems
- 3 +Resources on the Internet
- 2 +The Whole Internet Catalog
- +Appx. A Getting Connected to the Internet
- +Appx. B International Network Connectivity
- +Appx. C Acceptable Use
- 2 Glossary

The Whole Internet User's Guide & Catalog
13. Hypertext Spanning the Internet: <WWW>

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. *

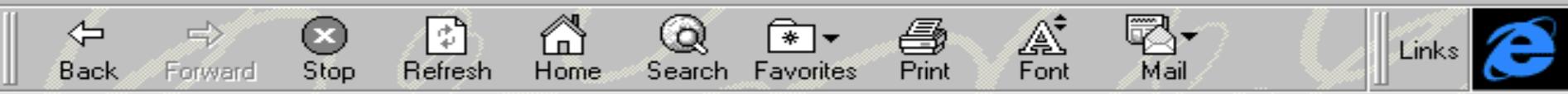
What Is Hypertext?

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

Query: Search

History List:

Found:



Address http://superbook.bellcore.com/~std/amit/kw.html

Tree Navigator Tool

Search...

Restricted Search...

Show Shared

Set Focus

Open Document

Go... Base

Prev Next

- Search Hit Score
- Frequency

- Mark Rosenstein's Sailing Page
 - + .
 - Sailing Magazines
 - + .
 - Sailing World -- On Deck
 - + .
 - **Sailing World Instruction Hotl**
 - + .
 - + .
 - One Design Resources
 - + .
 - International Contender Association
 - + .
 - **Swedish Contender Associat**
 - + .
 - + .
 - + .
 - Philadelphia Wooden Boat Factory
 - + .
 - **North End Shipyard Schoone**
 - + .

Search

Search for: sailing in Maine

Maximum Hits: 20



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