

Access to the Deep Web
a panel presentation and discussion

NSDL Annual Meeting

11/15/04

Speakers

- Walt Warnick, Science.gov
 - <http://science.gov>
- Sumir Meghani, Yahoo
 - <http://yahoo.com>
- Ammy Vogtlander, Elsevier
 - <http://www.scirus.com>
- John M. Saylor, NSDL (moderator)

Haiku

Deep treasures emerge

From cooperatively

Shared discovery

(inspired by Lee Zia)

Crossing the Divide Between Surface and Deep Web Access

Walter L. Warnick, Ph.D.
U.S. Department of Energy (DOE)
Office of Scientific and Technical
Information (OSTI)

NSDL Annual Meeting
November 2004

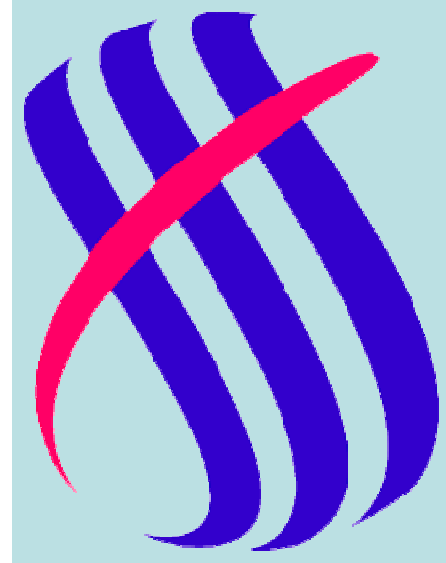
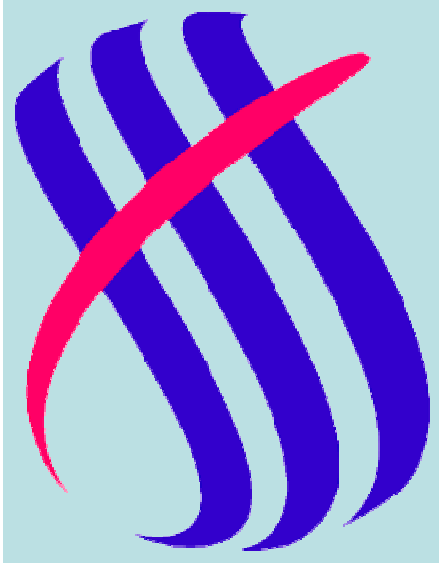
OSTI

Delivering science

since 1947

Our Mission

To advance science and sustain technological creativity by making R&D findings available and useful to DOE researchers and the American people.



Today we are
accomplishing our mission
by facilitating greater
access to the deep Web

- ❖ Through enhanced metasearch capabilities
- ❖ Through exposure of deep Web content to commercial Web crawlers
- ❖ Through exposure of deep Web content to other interested parties

OSTI coordinates the DOE-wide Scientific and Technical Information (STI) Program

Involves Headquarters Program Offices, Field
Offices, National Laboratories, other DOE
facilities and grantees

Collaboration enables OSTI to link to distributed
STI full-text documents across the DOE complex
via URLs in metadata

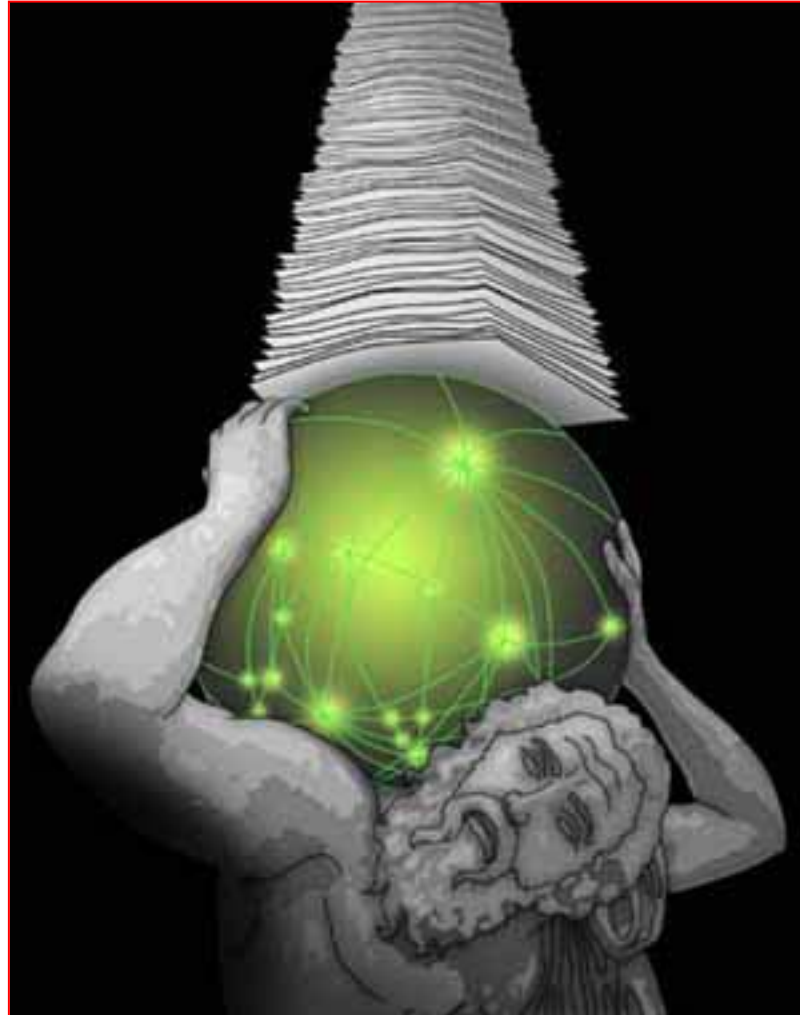
OSTI has managed DOE's R&D results for over 55 years and maintains a central repository

- ❖ Over 5 million electronic citations of DOE and other energy-related research information
- ❖ 1.5 million DOE-sponsored full-text reports
- ❖ Approximately 80,000 full-text reports publicly accessible electronically through our product called the Information Bridge

From the days
when
information
was difficult
to deliver



To today's Information Age *overload ...*



... “search” has been an evolutionary – and at times a revolutionary – process.

New and important steps have been taken to benefit researchers and the American public

Without much notice –
we've crossed a great
divide ...

between surface Web

and deep Web search

Most patrons search for information along the surface of the Web



YAHOO!

Google™

YAHOO!

Google™

YAHOO!

Google™

And thanks to our friends at Yahoo! and Google, patrons can quickly access lots of wonderful information.

In fact, Yahoo! and Google have done such a good job, that many patrons think: “If it’s not on Yahoo! or Google, then it simply doesn’t exist.”



YAHOO!

Google™

YAHOO!

Google™

YAHOO!

Google™

Deep Web

But, as the NSDL community well knows, there's much, much more to the World Wide Web picture than most citizens realize

The Deep Web Is Huge

**Deep
Web**

**Surface
Web**

Most STI is in databases in the Deep Web

- government databases
- journals

YAHOO!

Google™

YAHOO!

Google™

Google™

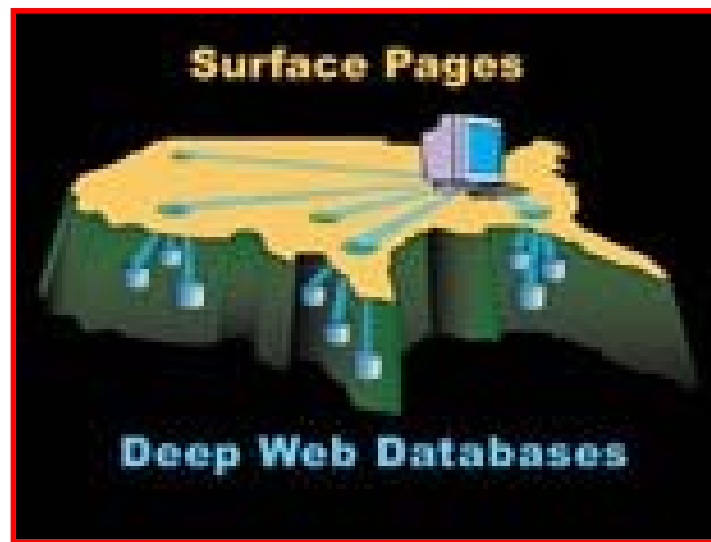
YAHOO!

Google™

YAHOO!

Deep Web

Estimates are that the deep Web, where most science information resides, is vastly larger than the surface Web

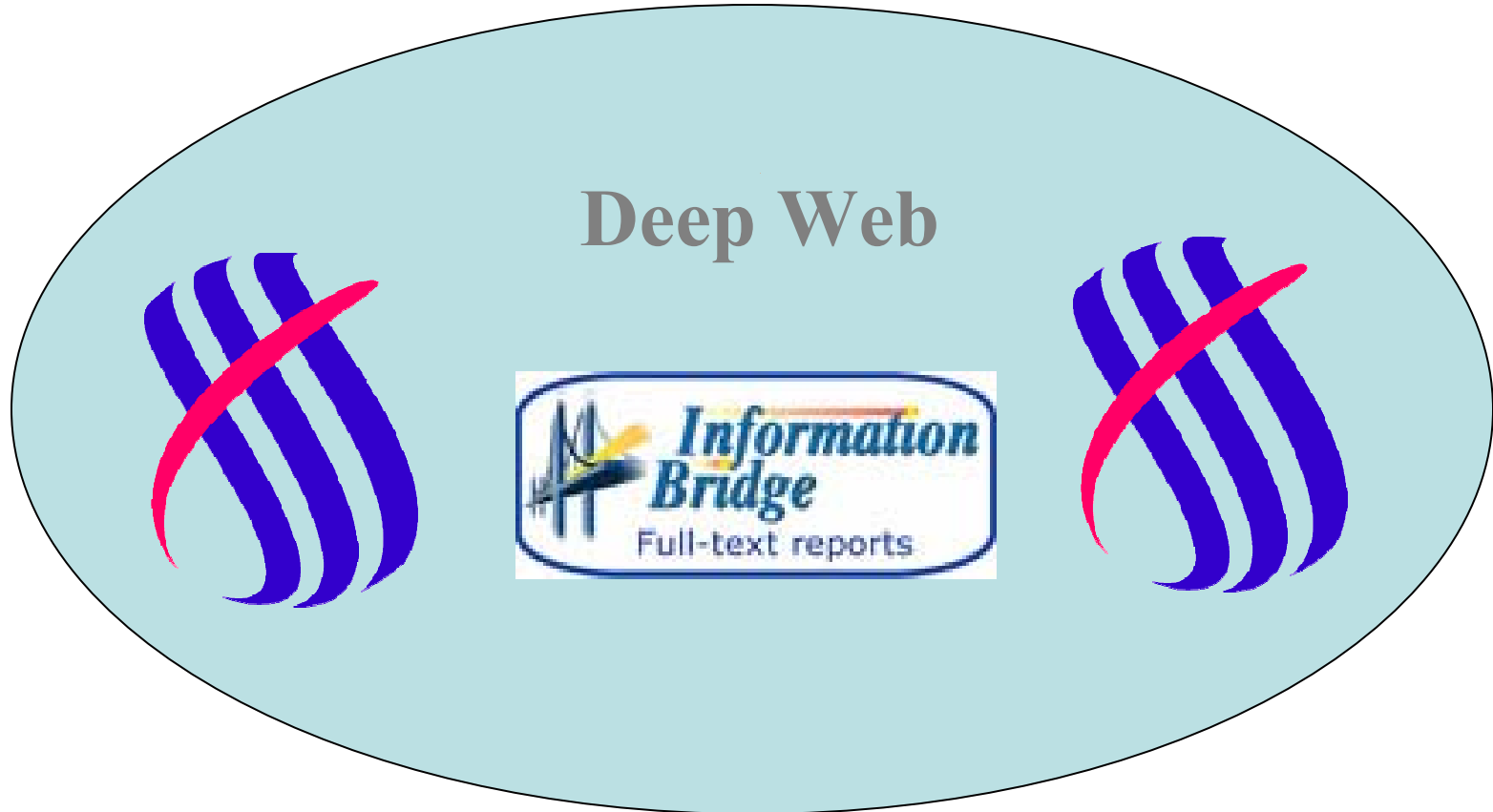


In 2003, UC Berkeley's School of Information Management and Systems estimated with 95 percent confidence:

surface web: **167 TB**

deep Web: between
66,800 TB and **91,850 TB**.

OSTI's Information Bridge is one example of a deep Web collection





YAHOO!

Google™

YAHOO!

Google™

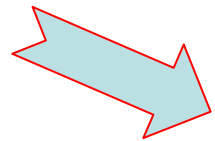
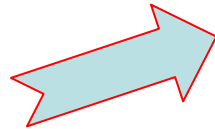
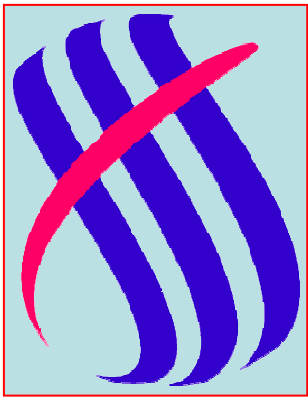
YAHOO!

Google™

Deep Web

How best to open deep Web
access to researchers and the
public?

We've followed a two-pronged approach



Developing, enhancing
Metasearch capability

Exposing our deep Web
content to surface Web

Metasearch

- ❖ **Searching multiple deep Web databases via a single query**
- ❖ **First government deployment was made by OSTI in 1999**
- ❖ **Applied to largest collection of government R&D databases through Science.gov in 2002**
- ❖ **Patrons often overwhelmed by too many hits**
- ❖ **Introduced meta-ranking of metasearch results in 2004**
- ❖ **R&D of next-generation algorithm for deployment in 2005**
- ❖ **Goal to approach ultimate in precision searching**

Hosted by OSTI, 12 science agencies (representing over 96 percent of US R&D budget) participate in delivering science information to citizens through Science.gov

Science.gov 1.0

- ❖ Metasearch
- ❖ Launched December 7, 2002

Science.gov 2.0

- ❖ Launched May 11, 2004
- ❖ Relevancy Ranking (Quick Rank)
- ❖ Advanced Search, Other Enhancements

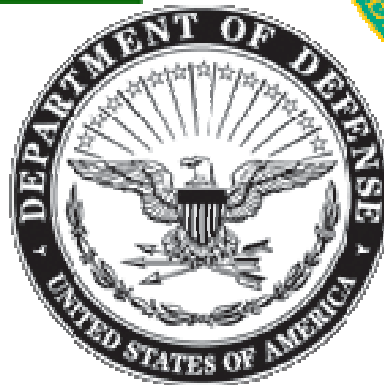
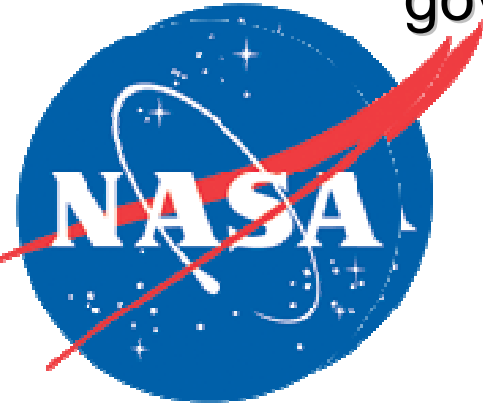
Science.gov 3.0
(Launch 2005)

- ❖ Alerts
- ❖ Fielded Search
- ❖ Boolean Search
- ❖ Enhance Relevancy Ranking (MetaRank)

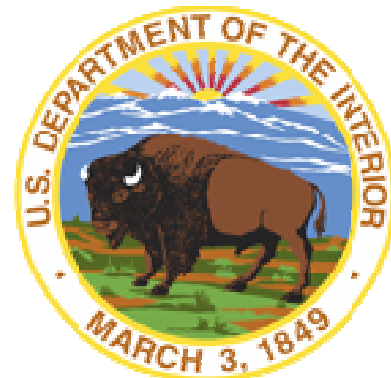
Science.gov 4.0
(Launch 2006)

- ❖ Enhance Relevancy Ranking (DeepRank)
- ❖ Full-text Relevancy Ranking, Science.gov 4.0 Grid

While Science.gov opened unprecedented public access to government science, it could not touch the Yahoo! /Google audience



With support from the National Archives and Records Administration



Most citizens were still only skimming the surface of the Web for their information



YAHOO!

Google™

YAHOO!

Google™

YAHOO!

Google™

So through collaborative efforts with Yahoo!
and Google, OSTI has been working to
expose deep Web content to the surface Web

An initial step

Now the public can access over 80,000 DOE R&D full-text reports and more than 2 million bibliographic records of energy-related R&D results through commercial surface Web crawlers

Other potential searchable collections resulting from OSTI introductions to science agencies:



Now we begin another approach to
support deep Web searching

Why shouldn't DOE information
and that of other science agencies
be searchable through the NSDL
collection?

We believe it should!

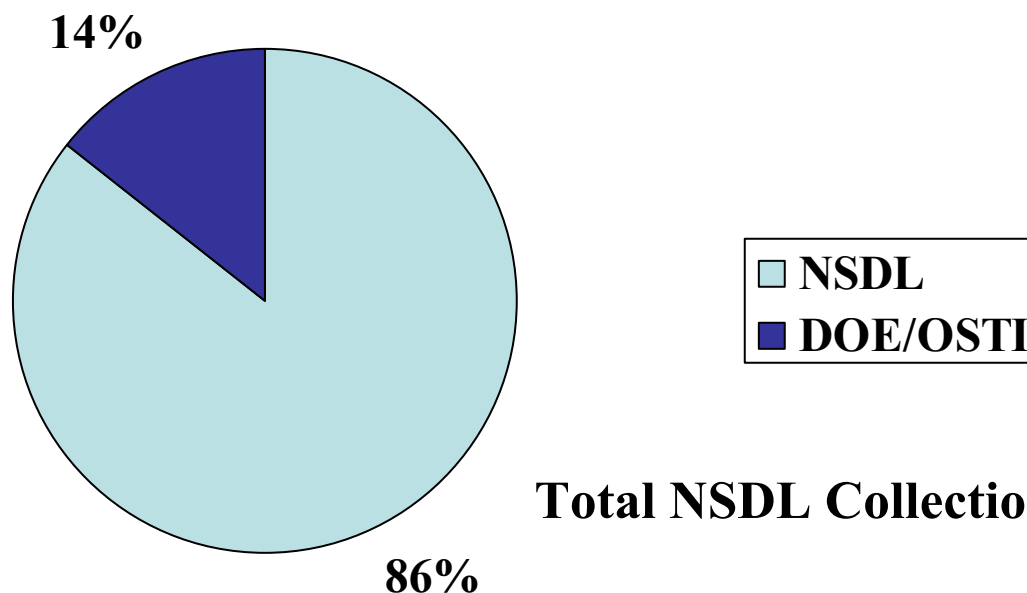
So – how do we get there?

- ❖ OSTI metadata was already in Dublin Core format (like that of the NSDL collection)
- ❖ OSTI established an OAI server so NSDL can harvest our records
- ❖ By becoming a Data Provider and adopting the OAI technical framework, greater exposure is gained to DOE's metadata and ultimately full-text scientific research information

OSTI's OAI compliance

- ❖ **Matches our mission of sharing science information**
- ❖ **Matches NSDL mission to enhance science, technology, engineering and mathematics education**
- ❖ **Introduces a large collection of government science information to NSDL resources**

Today we are pleased to announce the inclusion of 80,000 records of DOE R&D full text reports in the NSDL collection



Total NSDL Collection 557,000

We intend to share our efforts
with other federal agencies as
a model for encouraging even
broader exposure of science
databases to the NSDL
community

We look forward to continuing our role in facilitating greater access to the deep Web

- ❖ Continue our efforts in enhanced metasearch
- ❖ Continue exposure of deep Web content to commercial Web crawlers
- ❖ Continue to expose our deep Web content to other interested parties, such as NSDL



Yahoo! Search and Access to the Deep Web

November 15, 2004



Yahoo! as a Life Engine

Movie Clip

The background features a large, faded purple oval containing a white 'Y' logo, and a purple exclamation mark to its right.

Yahoo! and Search



Background on Yahoo!

- Started in 1994
- Over 150 million active registered users
- Global presence
 - 24 countries
 - 13 languages
- Strength in critical content verticals
 - #1 online Finance destination
 - #1 online News destination
 - #1 online Movies destination



Yahoo! Search Mission

- To deliver the highest quality search experience on the Internet
- Fundamental Principles of Yahoo! Search
 - Relevancy
 - Comprehensiveness
 - Freshness
 - Presentation



Powered by Yahoo! Search Technology (YST)

- On Feb. 18, Yahoo! rolled out its own search technology – Yahoo! Search Technology (YST)
- Powers nearly half of all searches conducted in the U.S.
- End-to-end search solution: generate, distribute, and monetize search results



Strong Response



“The new Yahoo search engine very, very, good...it is definitely a step up and an improvement for Yahoo...It is truly, for the first time in a long time, a viable alternative..”



“...One thing that has to be said is that Yahoo! has really hit this one out of the park...This is a really impressive set of search results.”



“We compared six searches on the two services...”

Score:

Yahoo!: 2

Google: 1

Tie: 3



“If Yahoo has not built a better mousetrap; it's at least developed one just as good.”



Conclusion: Web Search is a Parade

- Most popular online activity after e-mail
- But challenges remain
 - Traditional crawlers cannot penetrate the Deep Web
 - Hard to index unstructured content
 - Result: Half of current searchers are left unsatisfied within the industry
- We launched CAP to solve these problems
 - 99% of our search index will always be formed by crawling
 - Partnering directly with select content providers enables us to **deliver the highest quality search experience on the Internet**



**Building relationships to
explore the Deep Web**



What is the Content Acquisition Program (CAP)?

- Mission
 - To partner with content rights-holders and integrate top-tier content into search results, ultimately improving and differentiating the Yahoo! Search experience
- Progress
 - Since March, have established relationships with dozens of best-of-breed government, educational, and non-profit organizations
 - Billions of search queries have been impacted
 - Winner of Best Content Product award from the Charleston Advisor for an “aggressive approach to deep web searching”



Select Partners



WIKIPEDIA
The Free Encyclopedia

cdli

CUNEIFORM DIGITAL LIBRARY INITIATIVE

A joint project of the
University of California at Los Angeles and the
Max Planck Institute for the History of Science



The New York Public Library

OYEZ®

U.S. SUPREME COURT MULTIMEDIA





Digital Library of MIT Theses (OAster)

Yahoo! My Yahoo! Mail Welcome, **guest** | [Sign in](#) [Search Home](#) [Help](#)

YAHOO! search [Advanced Preferences](#)

[Web](#) [Images](#) [Directory](#) [Yellow Pages](#) [News](#) [Products](#)

Categories: • [Japan > Banking](#) • [Japan > Central Banking](#)

TOP 20 WEB RESULTS out of about 417,000 ([What's this?](#))

1. [Japanese Banking and Investment in the United States - Peter S. Rose Book. Read business & investing book reviews and co](#)
Yahoo! Shopping is the best place to comparison shop for **Japanese Banking and Investment** in the United States - Business & Investing Book. Compare products, compare prices, read reviews and ...
shopping.yahoo.com/shop?d=b&id=1966883494 - 28k - [Cached](#)
2. [NOMURA | Global Investment Banking](#)
... provide a broad range of **investment banking** services, including underwriting, financial advisory ... Announced Mergers and Acquisitions: Any **Japanese** involvement. League table based on ...
www.nomuraholdings.com/services/securities/wholesale/investment - 23k - [Cached](#)
3. [IT Investment Allocation in Japanese Banking Industry](#)
IT Investment Allocation in Japanese Banking Industry In this study, we investigate and discuss **IT investment** and related issues in the **Japanese banking** industry. And we insist that banks should ...
theses.mit.edu/Dienst/UI/2.0/Describe/0018.mit.etheses/2003-76
4. [Investment Banking in Japan Junichi Ujiie President and Chief Executive Officer The Nomura Securities Co., Ltd.](#) (PDF)
Investment Banking in Japan Junichi Ujiie President and Chief Executive Officer The Nomura Securities Co., Ltd. Good evening ladies and gentlemen. It is really a pleasure to see all of you here at the Columbia Business School. ... you an overview of **Japanese investment banking** strategies during the 1980's ...



News from NPR

Yahoo! My Yahoo! Mail Sports Entertainment Launch [edit] Welcome, **sumirmeghani** [Sign Out, My Account] Search Home Help

YAHOO! search info about movie shrek Yahoo! Search [Advanced Preferences](#)

Web [Images](#) [Directory](#) [Yellow Pages](#) [News](#) [Products](#) powered by

TOP 20 WEB RESULTS out of about 2,200,000 Search took 0.125 seconds. ([What's this?](#))

- [Shrek \(2001\)](#)
Shrek (2001) - Cast, Crew, Reviews, Plot Summary, Comments, Discussion, Taglines, Trailers, Posters, Photos, Showtimes, Link to Official Site, Fan Sites ... Summary: **Shrek** is a delight for all ages! **Shrek** is a movie like no other ... and creative story line. **Shrek** is a delightful comedy about a crude ogre ...
us.imdb.com/Title?0126029 - 46k - [Cached](#) - [More pages from this site](#)
- [Shrek](#)
official site from Dreamworks for the film featuring the voices of Cameron Diaz, Linda Hunt, John Lithgow, Eddie Murphy, and Mike Myers.
Category: [Movie Titles](#) > [Shrek](#)
www.shrek.com/ - 6k - [Cached](#) - [More pages from this site](#)
- [NPR Audio: Movie Review: 'Shrek 2'](#)
Film critic David Edelstein reviews the follow-up to the popular Disney animated feature **Shrek**. **movie**, **moviereview**, **shrek**, **disney**, **eddie murphy**, **eddy murphy**, **eddie murphey**, **mike mayers**, **mike myers**...
www.npr.org/features/feature.php?wfid=1905212 - [More pages from this site](#)
- [Buy Shrek Talking Plush Shrek at Wal-Mart Toys](#)
Shrek Talking Plush Shrek - Bring the hilarious characters from **Shrek** home! This talking plush **Shrek** is huggable and lovable, and he talks,... Wal-Mart: Always Low Prices.
www.walmart.com/catalog/product.gsp?product_id=2603643 - [More pages from this site](#)
- [Shrek Movie Posters at Shopping.com!](#)



Department of Energy

Yahoo! My Yahoo! Mail Welcome, **Guest** [Sign In] Search Home Help

YAHOO! search tetrachloroethylene, adsorption Yahoo! Search [Advanced Preferences](#)

Web [Images](#) [Directory](#) [Yellow Pages](#) [News](#) [Products](#) powered by

TOP 10 WEB RESULTS out of about 1,370. Search took 0.15 seconds. ([What's this?](#))

1. [DOE Document - Adsorption of low molecular weight halocarbons by montmorillonite](#)
... halocarbons studied were trichloroethylene, **tetrachloroethylene**, hexachloroethane, and dibromochloropropane. When the ... a maximum in halocarbon **adsorption** around pH 4, while untreated ...
www.osti.gov/energycitations/product.biblio.jsp?osti_id=5254882
2. [RKCL3981 PHOTOCATALYTIC DECOMPOSITION OF TETRACHLORO-ETHYLENE IN THE GAS PHASE WITH TITANIUM DIOXIDE AS CATALYST](#) (PDF)
... gas phase photocatalytic destruction of **tetrachloroethylene** (PCE) was ... mixture. Keywords: **Tetrachloroethylene**, photocatalysis, titanium dioxide, **adsorption**INTRODUCTIONChlorinated ...
www.jate.u-szeged.hu/~dombia/rkcl3981.pdf - 113k - [View as html](#)
3. [Activated Carbon for Liquid Phase Adsorption Systems, Design and Operation Factors: Activated Carbon from CPL Carbon Link](#)
Activated carbon from CPL Carbon Link for liquid and gas phase purification by **adsorption**. Activated carbons for all applications including chemical, water, air, solvent recovery, gold recovery, ... the design and operation of liquid phase **adsorption** systems. This text has been prepared to ... polar nature whilst, in comparison, **tetrachloroethylene** has a low solubility due to ...
www.activated-carbon.com/3-1.html - 8k - [Cached](#)
4.
... Batch **adsorption** experiments on naphthalene **adsorption** were conducted at various soil ... Batch **adsorption** experiments of PCE (**tetrachloroethylene**) and TCE (trichloroethylene) were ...
www.ce.udel.edu/~huang/annual96-97.html - 242k - [Cached](#)



NSDL and Yahoo! Search

- Conversations began last November, content began appearing in Yahoo! Search by February
- Utilized Yahoo! feed technology to optimize indexing of rich, technical content
 - Leveraging NSDL metadata
 - Understanding source of content



Yahoo!'s Competitive Advantages

- Working directly with content providers allows us access to previously uncrawlable content
- Leverage domain expertise of partner and editorial team at Yahoo!
- Growing array of personal search products that facilitate access to content
 - My Yahoo! Search (beta)
 - Toolbar
 - Verticals
 - RSS
 - Messenger



Taking it one step further



Introducing My Yahoo! Search: mysearch.yahoo.com

Welcome, **sumirmeghani**
[[Sign out](#), [My Account](#)]

 [Yahoo!](#)  [Mail](#)  [More Yahoo!](#)

MY **YAHOO!** SEARCH^{BETA}

[Search the Web](#) [Search My Web](#)

[My Web](#) - [Recently Visited](#)

Personal Search - search for what you want, the way you want.

- **Save** the results you like. [more](#)
- **Block** the ones you don't. [more](#)
- **Share** what you've found. [more](#)
- **Find** anything you've saved. [more](#)

Install the My Yahoo! Search button to save pages anywhere on the Web. [Learn more](#)



Save, Annotate, Share, and Block Search Results

Yahoo! My Yahoo! Mail Welcome, **sumirmeghani** [Sign Out, My Account] Search Home Help

MY **YAHOO!** SEARCH^{BETA} mitochondria aging Search the Web Search My Web

Visited Results My Web Blocking is ON Shortcuts Advanced Search Preferences

Search Results Results 1 - 10 of about 42,326 for **mitochondria aging** - 0.16 sec. (About this page)

- CERI: Feature Article: Mitochondria, Aging and Cognition**
Research Update: Mitochondrial Nutrition, **Aging** and Cognition. by Ward Dean, M.D., and Steven Wm. Fowkes ... many tiny organelles called **mitochondria** (see inset, Figure 1 below). These **mitochondria** produce most ... role in the **aging** of **mitochondria** is cardiolipin (diphosphatidylglycerol), a ...
[Save](#) - [Save with Note](#) - [Share](#) - [Block Site](#)
www.ceri.com/mito.htm - 26k - [Cached](#) - [More pages from this site](#)
- Mitochondria and Aging, Spring/Summer 1998, The Linus Pauling Institute, Oregon State University**
Continuing the pioneering efforts of Linus Pauling in orthomolecular medicine; basic and applied research of the application of nutritional factors to health promotion by intervention and disease prevention. ... **Mitochondria and Aging**. Tory Hagen, Ph.D. ... on understanding the role that **mitochondria** play in **aging**. We are keenly interested in the **mitochondria** because these ...
[Save](#) - [Save with Note](#) - [Share](#) - [Block Site](#)
www.orst.edu/dept/lpi/sp-su98/aging.html - 22k - [Cached](#) - [More pages from this site](#)
- Immlnst.org -> Mitochondria Aging**
... Immlnst.org " Focus " Biotech. **Mitochondria Aging**, Protecting the germ line ... The theory that damage to the **mitochondria** caused **aging** was first proposed in 1972, according to Sinclair ...
[Save](#) - [Save with Note](#) - [Share](#) - [Block Site](#)
imminst.org/forum/index.php?s=%26act=ST%26f=44%26t=3187%26view=pid%26pid=27383 - 99k - [Cached](#) - [More pages from this site](#)



Personal search experience

Yahoo! My Yahoo! Mail Welcome, **sumirmeghani** [Sign Out, My Account] Search Home Help

MY **YAHOO!** SEARCH^{BETA} mitochondria aging Search the Web Search My Web

Visited Results ▾ My Web Blocking is ON Shortcuts Advanced Search Preferences

Search Results Results 1 - 10 of about 42,326 for **mitochondria aging** - 0.16 sec. (About this page)

- CERI: Feature Article: Mitochondria, Aging and Cognition**
Research Update: Mitochondrial Nutrition, **Aging** and Cognition. by Ward Dean, M.D., and Steven Wm. Fowkes ... many tiny organelles called **mitochondria** (see inset, Figure 1 below). These **mitochondria** produce most ... role in the **aging** of **mitochondria** is cardiolipin (diphosphatidylglycerol), a ...
→ Saved to My Web - Add Note - Share
www.ceri.com/mito.htm - 26k - Cached - More pages from this site
This result has been saved to My Web. (Learn more...)
- Mitochondria and Aging, Spring/Summer 1998, The Linus Pauling Institute, Oregon State University**
Continuing the pioneering efforts of Linus Pauling in orthomolecular medicine; basic and applied research of the application of nutritional factors to health promotion by intervention and disease prevention. ... **Mitochondria and Aging**. Tory Hagen, Ph.D. ... on understanding the role that **mitochondria** play in **aging**. We are keenly interested in the **mitochondria** because these ...
Save - Save with Note - Share - Block Site
www.orst.edu/dept/lpi/sp-su98/aging.html - 22k - Cached - More pages from this site
- Immlnst.org -> Mitochondria Aging**
... Immlnst.org " Focus " Biotech. **Mitochondria Aging**, Protecting the germ line ... The theory that damage to the **mitochondria** caused **aging** was first proposed in 1972, according to Sinclair ...
Save - Save with Note - Share - Block Site
imminst.org/forum/index.php?s=%26act=ST%26f=44%26t=3187%26view=pid%26pid=27383 - 99k - Cached - More pages from this site



Create “content domains” to search across in the future

The screenshot shows the Yahoo! My Web interface. At the top, there's a navigation bar with 'Yahoo!', 'My Yahoo!', 'Mail', and a welcome message for 'sumirmeghani'. The main header is 'MY YAHOO! SEARCH BETA'. Below this, there are search buttons: 'Search the Web' and 'Search My Web' (highlighted with a red box). A sidebar on the left lists 'My Web' categories: 'All Saved Pages (3)', 'Uncategorized (0)', 'Biology Project (3)' (highlighted with a red box), 'Visited Results', and 'Blocked Sites (1)'. The main content area shows search results for 'mitochondria aging' (dated November 09, 2004). The top result is 'CERI: Feature Article: Mitochondria, Aging and Cognition' (highlighted with a red box), with a snippet about mitochondrial nutrition and aging. Below it is 'What role do mitochondria play in aging?' and 'MITOCHONDRIA AND AGING'. The interface includes a 'Share' button, a 'Categorize...' dropdown, and a 'Delete' button. A 'Sort By' menu is set to 'Title'. At the bottom left, there's a 'Publish 'Biology Project'' section with 'MY YAHOO!' and 'XML' buttons.



Search from anywhere


The screenshot shows a web browser window with the address bar containing the URL: `http://search.yahoo.com/search?fr=slv1-ocd&p=solar+eclipse&vs=worldcatlibraries.org&dups=1`. The search bar contains the text "solar eclipse". A dropdown menu is open, listing various search categories: Libraries, Images, Directory, Local, News, Shopping (Products), Maps, Yellow Pages, Dictionary, Stock Quotes (ticker symbol), Movie Showtimes (City or ZIP), and Address Book. The "Libraries" option is highlighted with a red box. The search results page shows a message: "Your search was restricted to 'worldcatlibraries.org/wcpa/ow/fbbc3...'. If you want to search the entire web, try searching [the Web](#) instead." The first search result is: "1. [Find in a Library: Solar eclips](#) See which libraries own this English [worldcatlibraries.org/wcpa/ow/fbbc3...](#)"



News Search

Yahoo! My Yahoo! Mail Welcome, **sumirmeghani** [Sign Out, My Account] News Home Search Home Help

[Web](#) | [Images](#) | [Directory](#) | [Local](#) NEW! | **News** | [Products](#)




YAHOO! News 

[Advanced Search](#) | [Preferences](#)

News Results [Got feedback?](#) | [Suggest a Site](#)

Also try: [total solar eclipse](#)


NEWS STORIES Results 1 - 20 of about 178 for solar eclipse. Sort Results by: Relevance | [Date](#)

- [Shock to the \(Solar\) System: Coronal Mass Ejection Tracked to Saturn](#) 
SPACE.com - Nov 05 4:02 AM
 In a dramatic proof that **solar** coronal mass ejection (CME) events affect even the outermost portions of the **Solar** System, scientists have traced an interplanetary shock from the Sun to Earth to Jupiter to Saturn.
- [Triple Eclipse on Jupiter](#) 
Universe Today - Nov 04 11:24 AM
 The Hubble Space Telescope took this rare picture of a triple **eclipse** on Jupiter, an event that only happens once or twice a decade. Io is near the middle, Ganymede on the planet's left edge, and Callisto is near the right edge. Astronomers tested a new technique with Hubble when taking this picture. They sped up Hubble's tracking system so that Jupiter passed through its field of view more
- [Space telescope spots triple eclipse on Jupiter](#) 
MSNBC - Nov 04 9:21 AM
 The Hubble Space Telescope captured a rare alignment of Jupiter's moons earlier this year in a picture that was released Thursday. Three of Jupiter's largest moons — Io, Ganymede, and Callisto — cast dark shadows on Jupiter in this near-infrared image

NEWS ALERTS

 Alert me when articles match all these words:
[solar eclipse](#)

ADD TO MY YAHOO! / RSS

Add your news search for [solar eclipse](#) to My Yahoo!:


View as RSS: [XML](#)

SPONSOR RESULTS

[Next Solar Eclipse](#)
 The first solar eclipse of 2002 is annular with a path that stretches the breadth of...
www.rainbowsymphony.com



Image Search


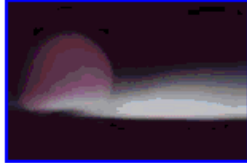

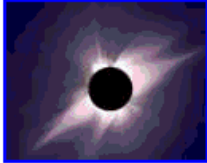
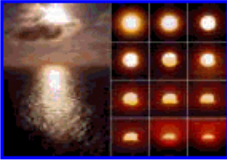



Yahoo! My Yahoo! Mail Welcome, **sumirmeghani** [Sign Out, My Account] Search Home Help

[Web](#) | [Images](#) | [Directory](#) | [Local](#) NEW! | [News](#) | [Products](#)

YAHOO! search

[SafeSearch is ON](#) | [Advanced Search](#) | [Preferences](#)

Image Results Results 1 - 20 of about 38,817 for **solar eclipse**. Search took 0.03 seconds.

 <p>ap12_eclipse.jpeg 374 x 330 pixels - 6.2kB cwx.prenhall.com/bookbind/pubbooks/.../deluxe-content.html</p>	 <p>Eclipse.jpg 1770 x 1170 pixels - 52.0kB www.nsas.ozau.net/gallery/kens%20eclipse.htm</p>	 <p>solar_eclipse.jpg 500 x 387 pixels - 69.2kB jeff.fassett.com/things/solar_eclipse.html</p>	 <p>solar_eclipse.jpg 504 x 403 pixels - 24.4kB physics.njit.edu/~dgary/202/Lecture2.html</p>
 <p>solar_eclipse_19741213.jpg 1000 x 692 pixels - 109.7kB www.astrosurf.com/re/eclipse.html</p>	 <p>solar_eclipse.jpg 725 x 590 pixels - 40.9kB www.aa6g.org/Astronomy/Eclipses/solar_eclipse.html</p>	 <p>S_eclipse.jpg 500 x 130 pixels - 19.4kB odin.physastro.mnsu.edu/~eskridge/ast102/week1.html</p>	 <p>12-solar_eclipse.jpg 335 x 226 pixels - 3.7kB www.worldbook.com/features/apollo11/html/apollo_12.htm</p>



RSS and My Yahoo!

The screenshot shows the Yahoo! Plus homepage. At the top left is the 'YAHOO! PLUS' logo. To the right are links for 'Answers, Solutions Yahoo! Plus Help' and 'MailPlus Included At No Extra Cost'. Below the logo is a navigation bar with 'Select Type:' and buttons for 'Web', 'Images', 'Yellow Pages', 'News', and 'Products'. A search bar is located below the navigation bar. The main content area is divided into several sections. A red box highlights the 'NYT Science' RSS feed, which contains five items: 'Flu Shots Skin Injection Might Stretch Supply of It', 'Sperm Stem Cells Are Grown Outside Body', 'Measure Passed, California Weighs Its Future as a Stem Cell Epicenter', 'Spawning of Interloper Fish in Potomac Worries Experts', and 'Drought Unearths a Buried Treasure'. Below this is the 'Science@NASA RSS Feed' with one item: 'Jupiter, Venus in Close Encounter'. Further down is the 'TIME Magazine Online: Top Rated Covers' section with three items. To the right of the RSS feeds is a 'Scoreboard' section for 'NBA' games, showing 'Utah vs. Golden State' and 'Detroit vs. Toronto'. Below the scoreboard is a 'Calendar' section for 'November 2004'.

YAHOO! PLUS

Answers, Solutions
Yahoo! Plus Help

MailPlus Included
At No Extra Cost

Select Type: **Web** Images Yellow Pages News Products

Search: Search

Welcome, Sumir! [Sign Out, My Account] 49° - 62°

[Add Content](#) [Change Layout](#) [Change Colors](#)

Select Page: [My New Page](#)

NYT Science [edit] [x]

- Flu Shots Skin Injection Might Stretch Supply of It - 8 hours ago
- Sperm Stem Cells Are Grown Outside Body - 8 hours ago
- Measure Passed, California Weighs Its Future as a Stem Cell Epicenter - 8 hours ago
- Spawning of Interloper Fish in Potomac Worries Experts - 8 hours ago
- Drought Unearths a Buried Treasure - 8 hours ago

Science@NASA RSS Feed [edit] [x]

- Jupiter, Venus in Close Encounter - 1 day ago

TIME Magazine Online: Top Rated Covers [edit] [x]

- TIME Magazine Cover: Charles Goren - 10 hours ago
- TIME Magazine Cover: Homosexuality - 10 hours ago
- TIME Magazine Cover: Albert Einstein - 10 hours ago

Scoreboard [edit] [x]

TODAY

NBA

Utah vs.	0	10:30
Golden State	0	ET
Detroit vs.	0	7:00
Toronto	0	ET

YESTERDAY

no games for selected teams

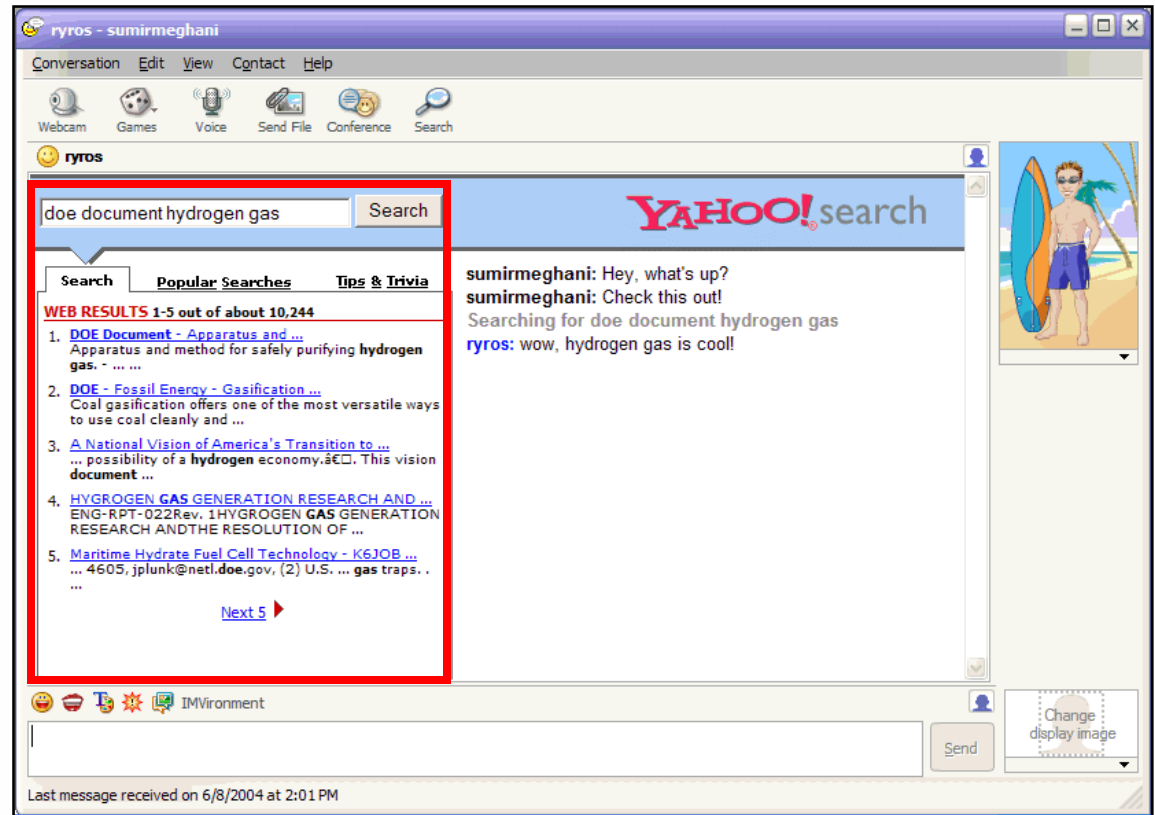
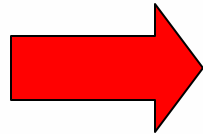
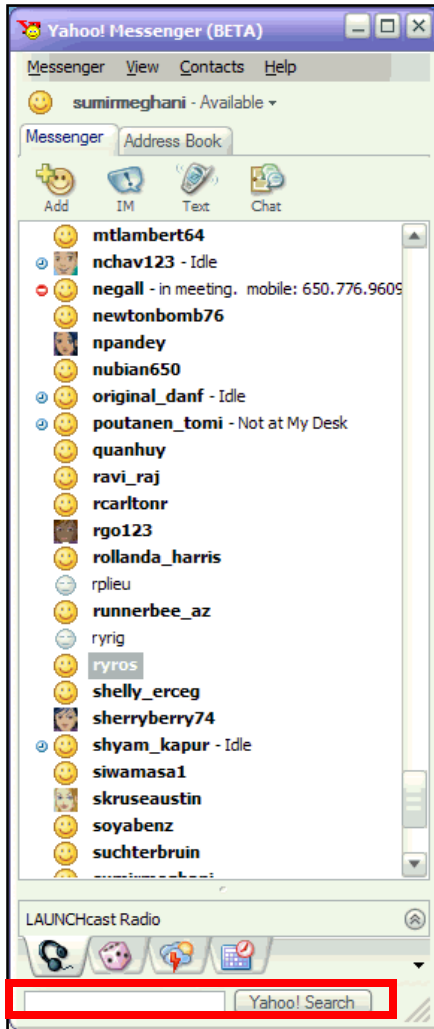
Calendar [edit] [x]

November 2004

Su	Mo	Tu	We	Th	Fr	Sa
29	1	2	3	4	5	6



Search with anyone





Questions and Discussion



A specialized search engine for scholarly searching

Presented by: Ammy Vogtlander, Sr. Product Manager Scirus
Event: NSDL Annual Meeting 2004 in Chicago
Date: Nov 2004



Agenda

- Introduction
- Defining 'The Deep Web'
- What a specialized web search engine can do
- In Conclusion

Introduction

- Libraries tend to focus on published content
- Users need also non-published information
 - People, Discussion forums, Conferences, Working papers, Project papers etc.
- Web search engines 1st choice, library portal 2nd
- 50% users find general search engines not satisfactory for their assignment
- Elsevier developed a specialised Web search engine: SCIRUS

Introduction of Scirus

- Free search engine focussed on scholarly Web
 - Over 150 million Web pages
 - Many valuable sources (preprint servers, patents government reports, journal sources (18 mil. articles))
- Offers scholarly-specific search functionality for the Web
- Valued by the scientific community
 - 1 million users (40% Europe, 30% Asia, 20% US); over 3 million searches per month; 19% brand awareness
 - Winner of WebAward 2004 for Directory or Search Engine and Search Engine Watch Award 2001 & 2002

Defining 'The Deep Web'

Reasons for not finding information:

1. Information cannot be indexed by (conventional) search engines
2. Information is missed by search engines
3. Results are buried

What a specialised search engine can do

Issue 1

Information cannot be indexed by search engines

What Scirus does

Different index methods for different sources

- Crawling
- Specialised crawling (e.g. AIP)
- OAI harvesting (e.g. ArXiv)
- Content loading (e.g. patents)

What a specialised search engine can do

Issue 2

Information is missed by search engines

What Scirus does

- Partner with content owner
 - Increase visibility of institute by branding results
 - Allow for searching on source via Advanced Search
 - Allow for Scirus to be used as search engine on site
- Use optimal index method for source
 - Use of site map
- Quality check

Scirus Advanced Search

SCIRUS

for scientific information only

[About Us](#)

[Newsroom](#)

[Advisory Board](#)

[Submit Web Site](#)

[Search Tips](#)

[Contact Us](#)

Advanced Search

[Basic Search](#) [Advanced Search](#) [Search Preferences](#)

AND

The complete document All of the words

The complete document All of the words

Published between

1930

and 2005

Information types

- All
- Abstracts
- Articles
- Books
- Company homepages

- Conferences
- Patents
- Preprints
- Scientist homepages

File formats

- All
- HTML

PDF

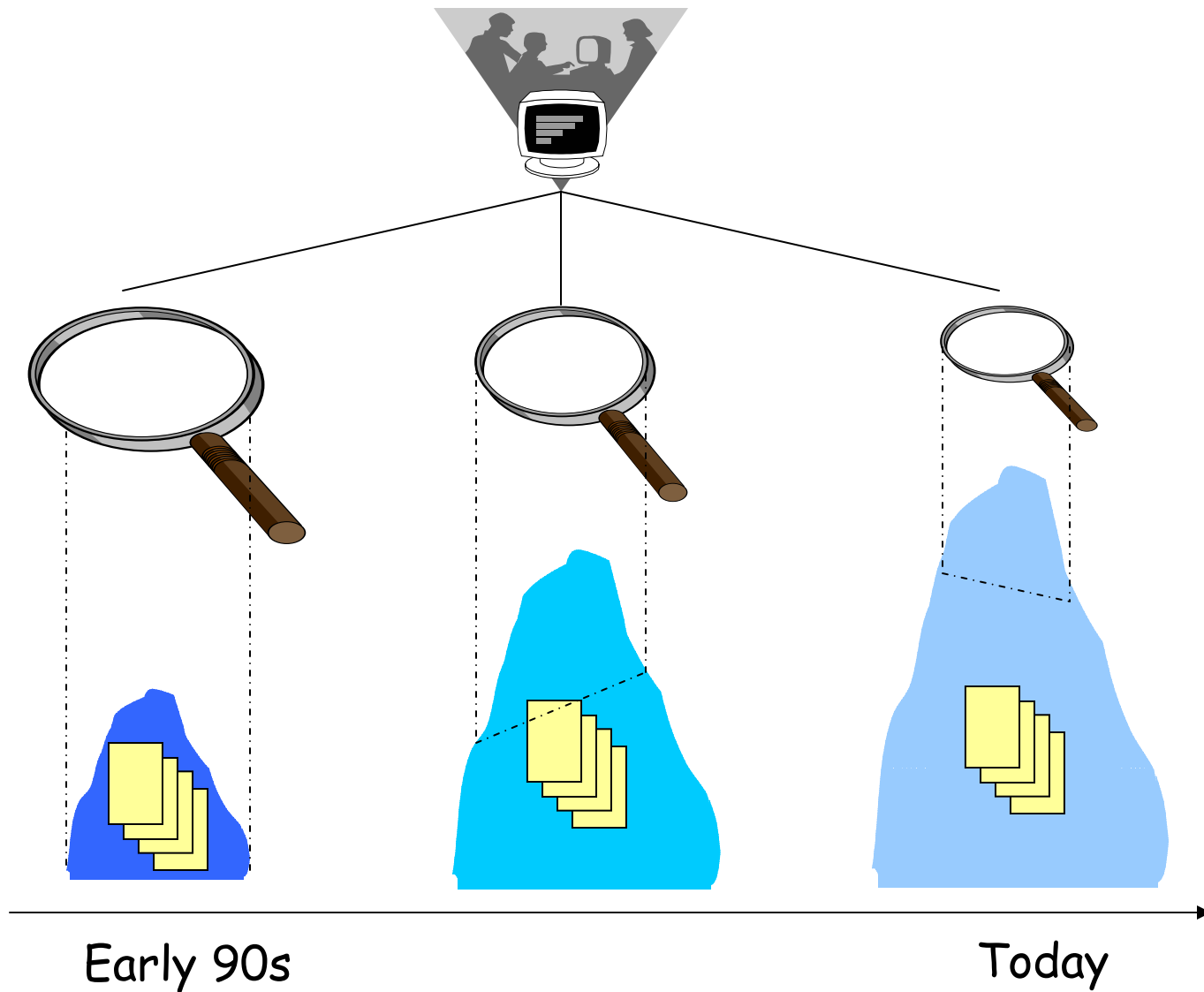
Content sources

- All Journal sources
- [BioMed Central](#)
- [MEDLINE](#)

- All Web sources
- [CogPrints](#)
- [E-Print ArXiv](#)

- The complete document
- Article title
- Journal title
- Author(s) name
- Author affiliation(s)
- Keyword(s)
- ISSN
- (Part of a) URL

Buried results: Today's search dilemma



What a specialised search engine can do

Issue 3

Results are buried

What Scirus does

- **Focussing on scholarly Web only**
- Analysing query; optimising results and ranking scholarly information higher
- Clearly distinguishing authoritative sources
- Offering scholarly-specific functionality

Search for Dolly in Google ... and get

[For God N Country](#)

Description: Fan page with exclusive **Dolly** news, photos, wallpaper, and album information.

Category: [Arts > Music > ... > Country > Bands and Artists > P > Parton, Dolly](#)

[www.dolly.net/](#) - 9k - [Cached](#) - [Similar pages](#)

[Dolly - Home Page](#)

Dolly magazine is for the fun loving Australian girl who wants to know the latest in everything ranging from fashion and beauty to entertainment news, movies ...

[g1.ninemsn.com.au/au01/455](#) - 33k - [Cached](#) - [Similar pages](#)

[Site officiel du groupe DOLLY](#)

[www.dolly-friends.com/](#) - 10k - 1 Oct 2003 - [Cached](#) - [Similar pages](#)

[Dollywood - Dolly Parton's Smoky Mountain Theme Park in Pigeon ...](#)

Escape to the unique world of **Dolly** Parton's Smoky Mountain theme park! ... Escape to the world of **Dolly** Parton's Smoky Mountain theme park! ...

Description: **Dolly's** theme park in Pigeon Forge, Tennessee.

Category: [Arts > Music > ... > Country > Bands and Artists > P > Parton, Dolly](#)

[www.dollywood.com/](#) - 11k - 1 Oct 2003 - [Cached](#) - [Similar pages](#)

[About the Poway Unified School District](#)

Projects. This page has been relocated. The new URL is: [www.powayschools.com/projects/dolly](#)

Please make a note of it! ~Thank you.

[powayusd.sdcoe.k12.ca.us/dolly/default.htm](#) - 13k - [Cached](#) - [Similar pages](#)

[ACME Dolly!](#)

Search for Dolly in Scirus ... and get

[Dolly](#)

Jan 1999

...MEAN (an opinion from Scott Gilbert) I. **Dolly** the sheep Individuals from two mammalian...cows--have recently been cloned.The sheep "**Dolly**" has been generated asexually by fusing...What does this mean? Scientifically, **Dolly** has not told us much. The cloning of an... [http://zygote.swarthmore.edu/gene4.html]

[similar results](#)

[AgBiotechNet - Topics - Animal Cloning: includes cloning of Dolly the sheep by Wilmut at Roslin and PPL](#)

Jan 2004

AgBiotechNet® - Topics - Animal Cloning: includes cloning of Dolly the sheep">
[more hits from](#) [http://www.agbiotechnet.com/topics/animal.asp]

[Scientists Cast Doubts on Dolly](#)

Aug 2003

Scientists Cast Doubts on **Dolly** by Nicholas Wade Copyright 1998 The New...announcement that Dr. Ian Wilmut had cloned **Dolly** the sheep from the adult cell of another...it is extraordinarily unlikely that **Dolly** is from a fetal cell." Since the ewe...

[more hits from](#) [http://www.junkscience.com/news/wilmut3.htm]

[similar results](#)

[NPR](#)

May 2004

...by Ken Rudin The First Mrs. JFK NPR Ombudsman by Jeffrey Dvorkin May 2004: Turning Point for Some Listeners Driveway Moments **Dolly** Parton's Stairway to Success This feature was suggested by Neil, who listens to WERN in Madison, Wisconsin. What is a 'Driveway...

[more hits from](#) [http://www.npr.org/]

[similar results](#)

[TPC NHC DOLLY 1996 PRELIMINARY REPORT](#)

Jun 2002

...RECONNAISSANCE Preliminary Report Hurricane **Dolly** 19-25 August 1996 Edward N. Rappaport National...Hurricane Bertha Hurricane Cesar Hurricane **Dolly** Hurricane Edouard Hurricane Fran Tropical...Hurricane Marco Visible image of Hurricane **Dolly** making landfall on the Yucatan Peninsula...

[more hits from](#) [http://www.nhc.noaa.gov/1996dolly.html]

using these keywords found in the results:

[cyclone](#)

[design](#)

[embryonic](#)

[embryonic stem](#)

[embryonic stem cells](#)

[embryos](#)

[land area](#)

[latching](#)

[miles per hour](#)

[stem cell research](#)

[strengthening](#)

[threaten](#)

[tropical cyclone](#)

[unnamed](#)

[weather forecasters](#)

[yellow pages](#)

Or refine using:

All of the words

refine

What a specialised search engine can do

Issue 3

Results are buried

What Scirus does

- Focussing on scholarly Web only
- **Analysing query; optimising results and ranking scholarly information higher**
- Clearly distinguishing authoritative sources
- Offering scholarly-specific functionality

Users want a basic search

SCIRUS
for scientific information only

[Latest Scientific News - from New Scientist](#)

[About Us](#)

[Newsroom](#)

[Advisory Board](#)

[Submit Web Site](#)

[Search Tips](#)

[Contact Us](#)

Basic Search

[Advanced Search](#) [Search Preferences](#)

Search

All journal sources All Web sources Exact phrase

Solve scientific challenges for financial rewards. [Click here!](#)

[Scirus Test Zone](#) | [Scirus Toolbar](#) | [Subscribe to Scirus News Updates](#)
[Advertising Information](#) | [Add Scirus to Your Website](#) | [Privacy Policy](#) | [Legal](#)

Powered by [FAST](#) © Elsevier 2004

Query analysis & refinement terms

Save checked results

Email checked results

- [1. Multilayer Neural Networks: Learning Models and Applications \(ResearchIndex\)](#)
Nov 2002
Multilayer neural networks: learning models and applications...Corrections) Sergio Gomez Jimenez **Multilayer neural networks:** learning models and... Home...networks 3 1.3
Multilayer neural networks...
[more hits from](#) [http://citeseer.nj.nec.com/269102.html]
[similar results](#)
- [2. Likelihood ratio of unidentifiable models and multilayer neural networks](#)
Fukumizu, Kenji, *The Annals of Statistics*, Jun 2003
This paper discusses the behavior of the maximum likelihood estimator (MLE), in the case that the true parameter cannot be identified uniquely. Among many statistical models with unidentifiability, neural network models are the main concern of this...
[http://ProjectEuclid.org/GetRecord?id=euclid.aos/10565...]
[similar results](#)
- [3. Likelihood ratio of unidentifiable models and multilayer neural networks](#)
Fukumizu, Kenji, *The Annals of Statistics*, Jun 2003
This paper discusses the behavior of the maximum likelihood estimator (MLE), in the case that the true parameter cannot be identified uniquely. Among many statistical models with unidentifiability, neural network models are the main concern of this...
[http://ProjectEuclid.org/getRecord?id=euclid.aos/10565...]
[similar results](#)
- [4. Adaptive Optical Multilayer Neural Networks](#)
Oct 1998
...project in Machine Learning : Adaptive Optical **Multilayer Neural Networks** Optical neural networks provide a very promising...Liquid-Crystal Light Valve Non-Linearities in Optical **Multilayer Neural Networks** , Applied Optics 35(26), 5301-5307 (1996). Georg...
[more hits from](#) [http://www.idiap.ch/~perry/opticalnn.html]
[similar results](#)
- [5. ON THE CAPACITY OF MULTILAYER NEURAL NETWORKS TRAINED WITH BACKPROPAGATION](#)
Oct 2000
...of Neural Systems, Vol. 10, No. 4 (2000) 281-285 © World Scientific Publishing Company ON THE CAPACITY OF **MULTILAYER NEURAL NETWORKS** TRAINED WITH BACKPROPAGATION E. N.

Your query was rewritten to: "multilayer neural networks"
We did this by adding quotes to common phrases, and by removing non-essential words.
[Repeat without rewrite](#)

Refine your search using these keywords found in the results:

- [activation function](#)
- [backpropagation](#)
- [feedforward](#)
- [feedforward networks](#)
- [feedforward neural networks](#)
- [generalisation](#)
- [generalization error](#)
- [gradient descent](#)
- [ieee transactions](#)
- [learning machine](#)
- [multilayer network](#)
- [on line learning](#)
- [training algorithm](#)
- [training examples](#)
- [two dimensional](#)

Or refine using:

All of the words

refine

What a specialised search engine can do

Issue 3

Results are buried

What Scirus does

- Focussing on scholarly Web only
- Analysing query; optimising results and ranking scholarly information higher
- **Clearly distinguishing authoritative sources**
- Offering scholarly-specific functionality

Special attention to authoritative sources

- 1. [Structure and method for forming a high efficiency electro-optics device](#)
Lin, Ming-Der / Wang, Kwang-Ru / Tsai, Chang-Da / Hsu, Jung-Kuei, Apr 2004
...structure of a high efficiency electro-**optics** device are disclosed. In the present invention...between the die carrier and the electro-**optics** cell is decreased to increase the light...emitting and absorbing regions of the electro-**optics** device. Thus, the operating efficiency...

Patent record available from the USPTO

[similar results](#)

- 1. [Results of soy-based meal replacement formula on weight, anthropometry, serum lipids & blood pressure during a 40-week...](#)
Fontaine, Kevin R. / Yang, Dongyan / Gadbury, Gary L. / Heshka, Stanley / Schwartz, Linda G. / Murugesan, Radha / Kraker, Jennifer L. / Heo, Moonseong / Heymsfield, Steven B. / Allison, David B., *Nutrition Journal*, Nov 2003
...involving a soy-based meal replacement **formula** is effective in lowering weight, fat mass...hypersensitivity to any of the ingredients of the **formula**, including but not limited to, soy protein...received the Scan Diet meal-replacement **formula**, instructions for its use, a single session...

Full text article available from  BioMed Central

Results from regular Web sites display the URL

- 1. ["Magneto-optic Kerr effect"](#)
Aug 1998
Magneto-optic Kerr Effect (MOKE) Contents Introduction Theory Experimental setup Typical...Vector Magnetometry Scanning **Kerr** Microscopy Introduction The **Magneto-optic Kerr effect (MOKE)** and the **Paraday effect** correspond to a change in the intensity...
[more hits from \[http://homer.phy.cam.ac.uk/Members/Cyrus_Daboo/MOKE/MO...\]](http://homer.phy.cam.ac.uk/Members/Cyrus_Daboo/MOKE/MO...)
[similar results](#)

What a specialised search engine can do

Issue 3

Results are buried

What Scirus does

- Focussing on scholarly Web only
- Analysing query; optimising results and ranking scholarly information higher
- Clearly distinguishing authoritative sources
- Offering scholarly-specific functionality

Scirus Advanced Search

SCIRUS

for scientific information only

[About Us](#)[Newsroom](#)[Advisory Board](#)[Submit Web Site](#)[Search Tips](#)[Contact Us](#)[Advanced Search](#)[Basic Search](#)[Advanced Search](#)[Search Preferences](#)

AND

The complete document

All of the words

The complete document

Article title

Journal title

Author(s) name

Author affiliation(s)

Keyword(s)

ISSN

(Part of a) URL

All of the words

Published between	<input type="text" value="1930"/> and <input type="text" value="2005"/>
Information types	<input checked="" type="checkbox"/> All <input type="checkbox"/> Abstracts <input type="checkbox"/> Articles <input type="checkbox"/> Books <input type="checkbox"/> Company homepages <input type="checkbox"/> Conferences <input type="checkbox"/> Patents <input type="checkbox"/> Preprints <input type="checkbox"/> Scientist homepages
File formats	<input checked="" type="checkbox"/> All <input type="checkbox"/> HTML <input type="checkbox"/> PDF
Content sources	<input checked="" type="checkbox"/> All Journal sources <input type="checkbox"/> BioMed Central <input type="checkbox"/> MEDLINE <input checked="" type="checkbox"/> All Web sources <input type="checkbox"/> CogPrints <input type="checkbox"/> E-Print ArXiv

Scirus Subject Classification

File formats	<input checked="" type="checkbox"/> All <input type="checkbox"/> HTML	<input type="checkbox"/> PDF
Content sources	<input checked="" type="checkbox"/> All Journal sources <input type="checkbox"/> BioMed Central <input type="checkbox"/> MEDLINE <input type="checkbox"/> ScienceDirect <input type="checkbox"/> Society for Industrial & App. Mathematics	<input checked="" type="checkbox"/> All Web sources <input type="checkbox"/> CoqPrints <input type="checkbox"/> E-Print ArXiv <input type="checkbox"/> Computer Science Preprint Server <input type="checkbox"/> Mathematics Preprint Server <input type="checkbox"/> NASA <input type="checkbox"/> US Patent Office <input type="checkbox"/> Other
Subject areas	<input checked="" type="checkbox"/> All <input type="checkbox"/> Agricultural and Biological Sciences <input type="checkbox"/> Astronomy <input type="checkbox"/> Chemistry and Chemical Engineering <input type="checkbox"/> Computer Science <input type="checkbox"/> Earth and Planetary Sciences <input type="checkbox"/> Economics, Business and Management <input type="checkbox"/> Engineering, Energy and Technology <input type="checkbox"/> Environmental Sciences <input type="checkbox"/> Languages and Linguistics <input type="checkbox"/> Law	<input type="checkbox"/> Life Sciences <input type="checkbox"/> Materials Science <input type="checkbox"/> Mathematics <input type="checkbox"/> Medicine <input type="checkbox"/> Neuroscience <input type="checkbox"/> Pharmacology <input type="checkbox"/> Physics <input type="checkbox"/> Psychology <input type="checkbox"/> Social and Behavioral Sciences <input type="checkbox"/> Sociology

Search

Conclusions

- It is not only about published content or particular sources
Other information sources are equally important in research
- Look for the right channels for you and your users to disseminate scholarly content
- Understand the issues surrounding general-purpose search engines and the sponsored advertising model

SCIRUS
www.scirus.com

Thank You

Contact details
a.vogtlander@elsevier.com

