

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~



UNFANGLING THE WEB

A Guide To Internet Research

The opinions expressed in this article are those of the author(s) and do not represent the official opinion of NSA/CSS.



Approved for Release by NSA on 04-19-2013, FOIA Case # 70381

(b) (3) - P.L. 86-36

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Untangling the Web: An Introduction to Internet Research

by [redacted] Center for Digital Content

Last Updated: February 28, 2007

Cover Design by [redacted]

(b) (3) - P.L. 86-36

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

~~This product has been produced for official use only. This product is not approved for public release. The information contained herein is for the exclusive use of the original recipient and is not for further distribution outside the recipient's agency or organization. The original recipient may make copies for distribution only within the recipient's agency or organization.~~

~~Ultimate responsibility for the protection of this product from public disclosure resides with the user, with severe penalties for non-compliance.~~

~~For additional information, please contact:~~

9800 Savage Road
Suite 6324
Fort Meade, MD 20755-6324

(b) (3) - P.L. 86-36

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

This Page Intentionally Left Blank

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Table of Contents

| | |
|--|-----|
| Preface: The Clew to the Labyrinth | 1 |
| “Every Angle of the Universe” | 5 |
| What Will I Learn? | 6 |
| Why Do I Need Help? | 7 |
| What’s New This Year | 8 |
| Introduction to Searching | 11 |
| Search Fundamentals | 11 |
| The Past, Present, and Future of Search | 12 |
| Understanding Search Engines | 18 |
| Search Engine Basics | 20 |
| A Word About Browsers: Internet Explorer and Mozilla Firefox | 22 |
| The Great Internet Search-Offs | 26 |
| Types of Search Tools | 28 |
| Web Directories/Subject Guides/Portals | 28 |
| Metasearch Sites | 30 |
| Megasearch Sites | 35 |
| Types of Searches and the Best Ways to Handle Them | 36 |
| Search Savvy—Mastering the Art of Search | 43 |
| Google | 47 |
| Google Hacks | 73 |
| Yahoo Search | 89 |
| Yahoo Hacks | 113 |
| Windows Live Search | 118 |
| Gigablast | 141 |

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Exalead..... 146

Ask..... 161

More Help: Internet Guides and Tutorials..... 173

Specialized Search Tools & Techniques..... 175

 “Google Hacking”..... 175

 Custom Search Engines..... 186

 Fagan Finder..... 193

 Wikipedia..... 202

 Maps and Mapping..... 215

Uncovering the “Invisible” Internet..... 239

 A9 Search..... 239

 Book Search..... 245

 Answers.com..... 260

 OAlster..... 264

 The Internet Archive & the Wayback Machine..... 267

 Other Invisible Web Resources..... 273

Casting a Wider Net—International Search, Language Tools..... 277

 International Search..... 277

 Online Dictionaries and Translators..... 288

 You Gotta Know When to Fold ‘Em..... 304

Beyond Search Engines—Specialized Research Tools..... 306

 Email Lookups..... 308

 Telephone and FAX Directories..... 311

 Online Videos and Video Search..... 317

 Online Audio, Podcasts, and Audio Search..... 344

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

| | |
|--|-----|
| Special Topics—News, Blogs, & Technology Search | 349 |
| Newsgroups, Forums, & Mailing Lists | 349 |
| Weblogs & RSS Feeds..... | 356 |
| General News Sources | 361 |
| News Sites & Search Engines..... | 362 |
| Technology News Sources..... | 377 |
| Telecommunications on the Web | 379 |
| Research How-Tos..... | 384 |
| Finding People | 384 |
| Using the Internet to Research Companies..... | 400 |
| How to Research a Specific Country..... | 411 |
| Finding Political Sites on the Web | 419 |
| Research Round-up: The Best Research Tips & Techniques | 424 |
| Researching & Understanding the Internet | 433 |
| A Plain English Guide to Internetworking | 433 |
| Researching Internet Statistics..... | 441 |
| Regional Registries and NICs | 443 |
| Domain Name Registries | 449 |
| Understanding Domain Name and Whois Lookup Tools..... | 451 |
| World Network Whois Databases: AfriNIC, APNIC, ARIN, LACNIC, & RIPE .. | 455 |
| Global Network Whois Search Tools..... | 456 |
| Domain Name Whois Lookups..... | 458 |
| Internet Toolkits..... | 471 |
| How to Research a Domain Name or IP Address | 474 |
| Traceroute | 483 |

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Geolocating Internet Addresses 497

Finding ISPs & Internet Access Points 503

Cybergeography, Topology, and Infrastructure 511

Internet Privacy and Security—Making Yourself Less Vulnerable in a Dangerous World 514

 Basics for Improving Your Internet Privacy and Security 518

 Increase Your Knowledge..... 521

 Browser Concerns 525

 Email Concerns 543

 Microsoft and Windows Concerns 560

 Handle with Care: More Privacy and Security Concerns 578

 General Security & Privacy Resources..... 605

Conclusion 606

Web Sites by Type..... 607

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Preface: The Clew to the Labyrinth

One of the most famous stories about libraries tells of the tenth century Grand Vizier of Persia, Abdul Kassem Ismael who, "in order not to part with his collection of 117,000 volumes when traveling, had them carried by a caravan of 400 camels trained to walk in alphabetical order."¹ However charming this tale may be, the actual event upon which it is based is subtly different. According to the original manuscript, now in the British Museum, the great scholar and literary patron Sahib Isma'il b. 'Abbad so loved his books that he excused himself from an invitation by King Nuh II to become his prime minister at least in part on the grounds that four hundred camels would be required for the transport of his library alone.²

A 21st Century version of the story might feature any number of portable electronic devices—a laptop, a PDA, or even a mobile phone—designed to overcome this difficulty. Today, 1000 years later, the Persian scholar/statesman would have to find a new excuse for declining the job offer. Abdul Kassem Ismael (aka Sahib Isma'il b. 'Abbad) would be hard pressed to explain why he couldn't just find what he needed on the Internet. The message seems to be that books are passé, replaced by ones and zeroes, the real world replaced by a virtual one, knowledge supplanted by information at best and chaotic data at worst. Have we shrunk the world or expanded it? Or have we in some way replaced it?

Untangling the Web for 2007 is the twelfth edition of a book that started as a small handout. After more than a decade of researching, reading about, using, and trying to understand the Internet, I have come to accept that it is indeed a Sisyphean task. Sometimes I feel that all I can do is to push the rock up to the top of that virtual hill, then stand back and watch as it rolls down again. The Internet—in all its glory of information and misinformation—is for all practical purposes limitless, which of course means we can never know it all, see it all, understand it all, or even imagine all it is and will be. The more we know about the Internet, the more acute is our

¹ Alberto Manguel, *A History of Reading*, New York: Penguin, 1997, 19. Manguel cites as his source Edward G. Browne's *A Literary History of Persia*, 4 vols., London: T. Fisher Unwin, 1902-24. I found the specific reference to this story on pages 374-375 of Vol. 1, Book IV, "Decline of the Caliphate." There is, sadly, no mention of the alphabetical arrangement of the library. This entire masterpiece is available online at The Packard Humanities Institute, Persian Texts in Translation, 23 February 2006, <<http://persian.packhum.org/persian/pf?file=90001011&ct=0>> (15 November 2006).

² Edward G. Browne. Vol. 1, Book IV, "Decline of the Caliphate," *A Literary History of Persia*, 4 vols., London: T. Fisher Unwin, 1902-24, 374-375. Available online at The Packard Humanities Institute, Persian Texts in Translation, 23 February 2006, <<http://persian.packhum.org/persian/pf?file=90001011&ct=0>> (15 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

awareness of what we do not know. The Internet emphasizes the depth of our ignorance because "our knowledge can only be finite, while our ignorance must necessarily be infinite."³ My hope is that *Untangling the Web* will add to our knowledge of the Internet and the world while recognizing that the rock will always roll back down the hill at the end of the day.

I will end this beginning with another story and a word of warning. "Tlön, Uqbar, Orbis Tertius" describes the discovery of an encyclopedia of an unknown planet. This unreal world is the creation of a secret society of scientists, and gradually, the imaginary world of Tlön replaces and obliterates the real world. Substitute "the Internet" for Tlön and listen. Does this sound familiar?

"Almost immediately, reality yielded on more than one account. The truth is that it longed to yield...The contact and the habit of Tlön have disintegrated this world. Enchanted by its rigor, humanity forgets over and again that it is a rigor of chess masters, not of angels...A scattered dynasty of solitary men has changed the face of the world. Their task continues. If our forecasts are not in error, a hundred [or a thousand] years from now someone will discover the hundred volumes of the Second Encyclopedia of Tlön. Then English and French and mere Spanish will disappear from the globe. The world will be Tlön."⁴

As we enjoy, employ, and embrace the Internet, it is vital we not succumb to the chauvinism of novelty, that is, the belief that somehow whatever is new is inherently good, is better than what came before, and is the best way to go or best tool to use. I am reminded of Freud's comment about the "added factor of disappointment" that has occurred despite mankind's extraordinary scientific and technical advances. Mankind, claims Freud, seems "to have observed that this newly-won power over space and time, this subjugation of the forces of nature, which is the fulfillment of a longing that goes back thousands of years, has not increased the amount of pleasurable satisfaction which they may expect from life and has not made them feel happier."⁵ Indeed, most of the satisfactions derived from technology are analogous to the "cheap enjoyment...obtained by putting a bare leg from under the bedclothes on a cold winter night and drawing it in again."⁶ What good is all this technology and information if, instead of improving our lot, it only adds to our confusion and suffering? We are continually tempted to treat all technology as an end in itself instead of a means to some end. The Internet is no exception: it has in large

³ Karl Popper, *Conjectures and Refutation: The Growth of Scientific Knowledge*, London & New York: Routledge, 2002, p. 38.

⁴ Jorge Luis Borges, "Tlön, Uqbar, Orbis Tertius," in *Labyrinths*, ed. Donald A. Yates and James E. Irby, New York: New Directions Books, 1962, 17-18.

⁵ Sigmund Freud, "Civilization and Its Discontents," tr. James Strachey, New York: Norton, 1962, 34-35.

⁶ Freud, 35.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

measure become the thing itself instead of a means of discovery, understanding, and knowledge.

Like Tlön, the Internet, "is surely a labyrinth, but it is a labyrinth devised by men, a labyrinth destined to be deciphered by men." We must avoid getting lost in the labyrinth without a clew. My hope is that *Untangling the Web* will be something akin to Ariadne's clew,⁷ so that as you unravel it, you can wind your way through the web while avoiding some of its dangers. Remember also that those who use the Internet to do harm, to spread fear, and to carry out crimes are like the mythical Minotaur who, as well as being the monster in the Minoan maze, was also its prisoner.



8

⁷Daedalus, the architect of the infamous labyrinth on Crete, purportedly gave King Minos' daughter Ariadne the clew, a ball of thread or yarn, to use to find a way out of the maze. Ariadne in turn gave the clew to Theseus, who slew the Minotaur and found his way out of the labyrinth. Theseus repaid Ariadne's kindness by leaving her on an island on their way back to Athens.

⁸ "Minotaurus," Wikimedia Commons, <<http://commons.wikimedia.org/wiki/Image:Minotaurus.gif>> (6 February 2007). This image is in the public domain because its copyright has expired.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Notes

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

“Every Angle of the Universe”

One wag has suggested that the Internet is an “electronic Boswell,” the chronicler of our age. It is that and more because the Internet chronicles not only a time and place but all times and all places, known and unknown, real and imaginary. The Internet is the closest thing to the fantastical “Aleph” imagined by the great Argentine story-teller Jorge Luis Borges, an object whose diameter is “little more than an inch” but which nonetheless contains all space, “actual and undiminished,” and in which one can see “every angle of the universe.”

While the comparison with the mythical Aleph may strike you as a bit whimsical, it is in fact not an altogether unfair metaphor. There has never been anything that approaches the Internet’s reach (to almost every part of the globe in less than thirty years), its size (estimated at 532,897 *terabytes* way back in 2003⁹), and its ability to link us together in a new kind of world community (words, pictures, sounds, ideas beyond imagining). But, as with all new technologies, it comes at a cost—many costs, in fact. ***We pay for the benefits of the Internet less in terms of money and more in terms of the currencies of our age: time, energy, and privacy.***

The goal of this book is to help you save some of each of these valuable resources: time, by making your searches more efficient; energy, by reducing the frustration using the Internet often entails; and privacy, by pointing out some simple measures to take to lower your cyber-profile and enhance your security.

I cannot emphasize strongly enough that this book was already out of date by the time it was published. Even though I have checked and rechecked every link in this book, some addresses are bound to have changed, some sites will have shut down, and some tips and techniques—such as search engine rules and syntax—will no longer be accurate. This is a testament to the changeable nature of the Internet and I must beg your forbearance for any such errors. Writing about the Internet is much like trying to catch Proteus¹⁰—as with the mythical prophet, it keeps changing and escaping our grasp.

⁹ School of Information Management and Systems, University of California at Berkeley, “How Much Information? 2003,” 27 October 2003, <<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/internet.htm>> (October 2005), Executive Summary.

¹⁰ “Proteus—i.e. full of shifts, aliases, disguises, etc. Proteus was Neptune’s herdsman, an old man and a prophet... There was no way of catching him but by stealing upon him during sleep and binding him; if not so captured, he would elude anyone who came to consult him by changing his shape, for he had the power of changing it in an instant into any form he chose.” “Proteus,” *Brewer’s Dictionary of Phrase and Fable*, 1898, <<http://www.bartleby.com/81/13723.html>> (14 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

“The Internet has often been called the world’s largest library with all of the books on the floor.”

Curtin, M., Ellison, G., Monroe, D., “What’s Related? Everything But Your Privacy,” 7 October 1998, Revision: 1.5,
<<http://www.interhack.net/pubs/whatsrelated/>>
(14 November 2006).

What Will I Learn?

To achieve these goals, this book will:

- help you understand how to use the Internet more efficiently to find useful information and, in so doing...
- make clear why the Internet is an invaluable resource.

This year I have reorganized the book to make it more logical and easier to use. The first part of the book still focuses on the ***ins and outs of searching***: how search engines work, types of search tools, how to handle different types of searches. The next section has expanded to offer in-depth tutorials on six major search engines. Next, the book covers ***specialized search tools and techniques***, including a new section devoted to Wikipedia. I have also moved the discussion of maps and mapping to this section. This is followed by ***“invisible” web research*** to include the changes to A9 and Amazon’s search inside the book option. Next is the ***international search and language tools*** section, followed by specialized research tools, including new sections on video, audio, and podcast searches. The next section covers ***specific topical research***, such as news, telecommunications, blogs, and RSS feeds. This is followed by a series of “how to” guides, culminating with ***tips and techniques*** for more effective searching. The book then delves into ***using the Internet to research the Internet***, with the final section still addressing crucial ***privacy and security issues***.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Why Do I Need Help?

There are no Internet research experts.

There are people who make a living using the Internet for research and who know more than others about what is on the Internet, how to find what they want on the Internet, and how to do this with relative efficiency. But no one knows what is truly "out there" for two fundamental reasons:

- The Internet changes constantly. By that I mean daily, hourly, minute-to-minute, incessantly.
- It's too darned big! If we can't accurately size the Internet (which we can't), you can be sure we don't know what is available via this resource with any degree of accuracy or completeness.

This doesn't mean you can't ever hope to find anything on the Internet. You often can find what you're looking for (and usually a lot more) with comparative ease, but no one should be deluded into believing he has a good grasp of the entire world of information available on the Internet. Realistically, the best search engines index **only a fraction of all webpages** and keyword searching is at best an art that routinely misses relevant sites while loading you down with dross.

Are you discouraged? Don't be...novices often have more luck finding something arcane than seasoned researchers because of the power of creative thinking and serendipity. I've learned never to underrate luck and intuition when doing Internet research, but I think the two most important tools for successful Internet research are:

- 1. a good set of bookmarks**
 - 2. other people with experience searching the Web**

Never assume others are already aware of some website, tool, or technique you find particularly useful. The sheer quantity of data, information, and knowledge associated with the Internet is so enormous that **no one** can know more than a fraction of what's on it. While we're talking size, let me mention an important distinction. The **Internet** and the **web** are not one and the same, though the web is what most people think of when you say "Internet."

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

In fact, as huge as it is, the ***Worldwide Web is actually a subset of the Internet.*** The Internet is the network of networks, all the world's servers connected by routers, to put it in semi-technical terms. The web is that portion of the Internet that uses a browser (typically Netscape or Firefox—browsers built upon Mozilla—or Microsoft's Internet Explorer) and some type of hypertext language (usually HTML) to move around. This book focuses primarily on the web because tackling the web by itself is a big enough challenge.

As you have no doubt guessed by now, the Worldwide Web does not come with an instruction manual or users guide, which means much if not most of what you learn about researching using the Internet will come from hard-won experience. On the up side, you probably will not be able to break anything on the Internet. More than likely, no matter how lost or hopelessly confused you become, you will only damage your own computer and/or network—and perhaps your good humor and sanity. However, because of the almost astronomical growth of malicious activity, the Internet has become a dangerous place, and users have discovered that they have inadvertently spread malicious software (malware) such as viruses, worms, and Trojan horses. That is why I have devoted the last section of the book to personal computer security and privacy. We are all at risk from the rising tide of bad and in some cases criminal behavior, so we must take responsibility for protecting ourselves and our computers from the ruses and attacks that grow in number and sophistication each year.

This book will expand on simple “rules” of Internet research, rules that are really more in the nature of friendly suggestions. These rules are the fruit of my own experiences as an Internet user and may prevent you from repeating all the mistakes I made that gave rise to the rules in the first place. Some of these suggestions may at first strike you as odd or inconsistent, but the rationale for each I hope will become clear as we go along.¹¹ The fact is that today we are drowning in information and starving for knowledge. The goal of *Untangling the Web* is to help rescue users from the ocean of information (and misinformation) by throwing them a virtual lifeline.

What's New This Year

Most people probably have not thought about or been very much affected by the changing search landscape because, as is only natural, most people have one or two sites they routinely use for search and research, regardless of the nature of the inquiry. However, virtually all search professionals will agree that ***knowing where to look for information is the key to successful searching.*** Yet few venture beyond

¹¹ If you are using the hypertext version of this book on line, the links in the paper may not load correctly. Try the refresh button, copy and paste the url, or type in the url directly.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

the comfortable confines of the familiar search engine. While the major search engines continue to improve each year, they are far from the be all and end all of search. The problem with general search tools is that they cannot provide targeted or tailored results, certainly not without a lot of work on the part of the user. For this reason, a large part of *Untangling the Web* is devoted to other ways to uncover information, be it subject guides, "deep web" resources, targeted search tools, or unusual tips and techniques for revealing what is hidden.

Again this year, I have included detailed information on how to use Google, Yahoo, Gigablast, and Live Search (formerly MSN Search) to find very specific data. I have also updated and expanded the section on Exalead and added Ask to the major search engines. However, unless you spend a fair amount of time using each of these search tools, you will probably find their many options too complicated and cumbersome for everyday use. A different approach is to use specialized search tools, which begs the question of how to find these tools. *Untangling the Web* maps a number of the Internet's less-traveled roads, i.e., excellent but unheralded specialty search tools such as Fagan Finder, Amazon's A9 multipurpose search, and ThomasGlobal's business search. Also, the section on international search is substantially larger than before.

In recognition of the growing importance and influence of collaborative websites, there are several new sections in this year's book. One is a separate section devoted to Wikipedia, contributed in part by my colleague Diane White. Video and audio search exploded during 2006, and this year's edition contains a new and extensive examination of video search sites as well as a new section on audio search and podcasting. Two other new sections are devoted to custom search engines and book search, neither of which is an entirely new technology but both of which spread in popularity and improved in quality in the past year. Custom search is fast becoming a replacement for web directories, which continue their slide into irrelevance.

The section on researching and understanding the Internet now begins with a new section on "internetworking." This tutorial is a response to a number of requests from people such as myself who need basic knowledge and understanding of how the Internet works without too much technical jargon or expertise. I hope you find that it falls in a comfortable middle ground between simplistic and abstruse.

Once again, the section on privacy and security grows in proportion to concerns about protecting our privacy and security on the Internet. Fortunately, as the problems increase and the malicious users become more enterprising, so do the ways and means for protecting ourselves. However, home computer security is a personal responsibility few people take seriously until it is too late. *Untangling the Web's* privacy and security information is designed to help users avoid becoming victims and instead take the offense in protecting themselves, their families, and by extension, the Internet community from the Internet's evil-doers. The 2007 edition includes new sections on clearing private data in Firefox, encrypting files in

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Windows, pretexting, protecting yourself from search engine leaks, whether or not you can really opt out of online directories, and a brief discussion of wireless Internet use.

I have also reorganized *Untangling the Web* to make it easier to use. The new section on “Specialized Search Tools & Techniques” brings together some already existing topics, such as Google hacking, with the new sections on Wikipedia and Custom Search Engines. I also moved maps up to this section because they have become integral to basic search. Specialized Research Tools now include the video and audio search sections as well as telephone and email search. Basically, all types of search comprise the first two-thirds of the book, while the remainder focuses on the Internet itself and privacy and security.

As was true of last year’s edition, I can again say with confidence that the 2007 *Untangling the Web* was already out of date before it reached your desk. Experienced Internet users understand the Internet is truly a river of information that is impossible to step into twice. And the basic concepts for using the Internet to research topics of interest to our community of readers are sound despite changes in websites, links, and technology.

Web Tip

Web links often change. In case of a bad link for a news article, use the site’s search facility and search by the headline, author, or date. In the case of a bad link inside a website, try going to the site’s homepage and working your way down to the page, which may still be there, only in a different location.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Introduction to Searching

Search Fundamentals

The September-October 1997 issue of *IEEE Internet Computing* estimated the Worldwide Web contained over 150 million pages of information. At the end of 1998, the web's size had grown to more than 500 million pages. By early 2000, the best estimates put the number over 1 billion and by mid-2000 there was a study showing that there are over **550 billion unique documents on the web**.¹² Netcraft, which has been running Internet surveys since 1995, reported in its November 2006 survey that there are now **more than 100 million websites**. "The 100 million site milestone caps an extraordinary year in which the Internet has already added 27.4 million sites, easily topping the previous full-year growth record of 17 million from 2005. The Internet has doubled in size since May 2004, when the survey hit 50 million."¹³ The major factors driving this boom are free blogging sites, small businesses, and the relative and lower cost of setting up a website. Another recent survey found:

- *The World Wide Web* contains about 170 terabytes of information on its surface; in volume this is seventeen times the size of the Library of Congress print collections.
- *Instant messaging* generates five billion messages a day (750GB), or 274 Terabytes a year.
- *Email* generates about 400,000 terabytes of new information each year worldwide.¹⁴

The numbers hardly matter anymore. The enormous size of the Internet means we simply must use search tools of some sort to find information. Otherwise, we are voyagers lost on a vast uncharted ocean.

¹² Michael K. Bergman, "The Deep Web: Surfacing Hidden Value," *BrightPlanet*, August 2001, <<http://www.brightplanet.com/technology/deepweb.asp>> (14 November 2006).

¹³ "November 2006 Web Server Survey," Netcraft.com, 1 November 2006, <http://news.netcraft.com/archives/2006/11/01/november_2006_web_server_survey.html> (15 November 2006).

¹⁴ School of Information Management and Systems, University of California at Berkeley, "How Much Information? 2003," 27 October 2003, <<http://www.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm#summary>> (14 November 2006) Executive Summary.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~**Consider this:**

When you do a search, you are going through more information in less than 30 seconds than a librarian probably could scan in an entire career 30 years ago.

All the major search engines now index well over a billion pages of information. The problem generally isn't lack of data but finding that one tiny needle in a virtual haystack of almost limitless size (much like looking for a needle in a stack of needles).

Any serious researcher needs to know more about search engines than the average person using the Net for fun or even for very specialized searches associated with a hobby or perhaps a certain topic, e.g., cancer research. How do you learn the ins and outs of search?

The Past, Present, and Future of Search

"Search has become the most hotly contested field in the world of technology."¹⁵

Remember Northern Light? How about Excite, Galaxy, Lycos, HotBot, Magellan, InfoSpace, Go, Webcrawler, iWon, Netfind, or Webtop? If so, you've been searching the Internet a long time because many of these search engines are long gone and forgotten. However many changes in search and search engines have taken place in recent years, nothing has been quite so dramatic as what has occurred in the past two years with the appearance of the new Yahoo and Live Search engines.

While many smaller, focused search tools still exist, the sad fact is that, in terms of large, powerful, world-encompassing search engines, Internet searchers at this moment have fewer major search engines from which to choose.¹⁶ What happened to get us to this point and what does the future portend?

¹⁵ Terry McCarthy, "On the Frontier of Search," *Time.com*, 28 August 2005, <<http://www.time.com/time/magazine/article/0,9171,1098955-1.00.html>> (14 November 2006).

¹⁶ Of course there are many non-US search engines beyond those run by Google, Yahoo, and Microsoft, but they generally target a particular part of the world and are not serious competitors with Google, Yahoo, or Live Search at this time.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

In the early years of the Internet, there was enormous competition in the search market among a large number of search engines vying not only for users but, more importantly, for investors. The “dot bomb” crash in mid-2000 began the shakeout of search companies that continues to this day. The biggest change wrought by the failure of so many Internet-based investments was the growth of pay-per-click advertising in search results. Pioneered by Overture, these so-called sponsored results began to show up at the top of search result lists: the more an advertiser was willing to pay, the higher his result on the list. Then, in 2002 the big search engine consolidation began: first, Yahoo purchased Inktomi, a little known but major player in the search engine world. In 2005, Overture bought AltaVista, one of the oldest and most venerable search engines on the Internet, then quickly acquired AlltheWeb, another major search engine. To top it off, in July 2003, Yahoo bought Overture, thus acquiring three huge search properties at one time.

All this was done publicly. The real revolution was what was happening behind the scenes: with a remarkable degree of secrecy, Yahoo gave the engineers it had acquired from AltaVista, AlltheWeb, and Inktomi a new task—create a whole new search engine to compete with Google. On February 18, 2004, Yahoo unveiled its new search engine, which has a database and search features to rival Google’s. Shortly thereafter, Yahoo began killing off the “parents” of its new progeny: first Inktomi, then AlltheWeb and AltaVista. While users can still go to the AlltheWeb and AltaVista websites and run searches, the results are pulled from the Yahoo database and many of the unique search options and features of both search engines are no longer available. However, Yahoo continues to add new features and options that are improving its capabilities.

During 2006, two major search engines unveiled major changes that make them serious contenders: Ask and Exalead. During 2006 Teoma and Ask Jeeves ceased to exist as separate search sites and merged under the Ask.com umbrella. The French search engine Exalead came out of beta for a new look and major overhaul during 2006 and continues to offer a number of important and unique search features. MSN Search became Live Search, which left beta status in September 2006 and increased the much-needed competition from a company that knows how to make successful (if imperfect) products. **Amazon.com** still offers its own search engine, A9, although during 2006, Amazon eliminated some of A9’s unique functions, switched from Google to Live Search to power web searches, and appeared to be if not abandoning A9 then certainly scaling it back.

All the major search sites are still trying to be the “Swiss army knife” of search engines. Google, Yahoo, Live Search, Ask, and Exalead all competed hotly with each other to roll out new, better, faster, fancier, more powerful tools to do everything from search the contents of your computer in a heartbeat to letting you “fly” around the world with a bird’s (or satellite’s) eye view of the planet. Among the new search engine-based tools and programs arriving this past year were vastly improved maps and mapping technologies, enhanced multimedia search, desktop

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

search utilities, toolbars integrated into the browser, and application programming interfaces (APIs) for use by individual developers.

If 2004 was the year of the new search engine and 2005 the year of tailored search, 2006 seems to have been the first year of Web 2.0. Interactive, participatory Internet activities such as blogging, podcasts, online video sharing, and wikis dominated the discourse.

Podcasting finally came into its own last year. Podcasting is recording and broadcasting any non-musical information—be it news, radio shows, sporting events, audio tours, or personal opinions—usually in MP3 format for playback using a digital audio player. Many websites now serve as directories to help users find podcasts of every variety anywhere in the world. Podcasting has caught on because it is easy, inexpensive, mobile, flexible, and powerful. Yahoo got out in front of the podcasting trend with its new Podcasts Search site after a study the search giant published with Ipsos Insight, which disclosed that most of the people who are using RSS do so without even knowing it.¹⁷ RSS, which either stands for Rich Site Summary or Really Simple Syndication, is an XML format for news and content syndication. News aggregators are programs designed to read RSS formatted content, which is very popular in the blogging community. Many if not most blogs make their content available in RSS.

Although there is no agreed upon definition of what Web 2.0 means, in general terms most people believe it involves at a minimum users collaborating to share information online, i.e., an interactive, participatory web in contrast to what is now being called the static web (or Web 1.0). I think the Wikipedia article on Web 2.0 sums the current state of affairs up nicely when it says "To some extent **Web 2.0** is a buzzword, incorporating whatever is newly popular on the Web (such as tags and podcasts), and its meaning is still in flux."¹⁸

Another important aspect of Web 2.0 is that it organizes information differently from traditional web and other news and knowledge models. So reports a *Time* article on the frontiers of search in its 5 September 2005 issue. There is good reason to believe this claim, given a major investment firm's assessment that "by 2010, search-engine advertising will be a \$22 billion industry worldwide, up from an estimated \$8 billion today."¹⁹

One casualty of Web 2.0 appears to be directories. Directories are hierarchical guides to a subset of what are presumably the best, most relevant (or at least most popular) websites on a specific topic. Yahoo was always the king of directories, but

¹⁷ Yahoo! and Ipsos Insight, "RSS: Crossing into the Mainstream," October 2005 [PDF], <http://publisher.yahoo.com/rss/RSS_whitePaper1004.pdf> (14 November 2006).

¹⁸ "Web 2.0," Wikipedia, <http://en.wikipedia.org/wiki/Web_2.0> (15 November 2006).

¹⁹ McCarthy.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

several years ago, I noted a marked decline in both the quantity and quality of the Yahoo directory. The other major directory was and remains the Open Directory Project, which has always powered the Google Directory and, ironically, now powers the Yahoo Directory. What distinguished the Open Directory from Yahoo was that, while Yahoo was heavily commercial, the Open Directory has always relied upon volunteers to populate and maintain it. Now that most of users' creative energy seems to have moved to wikis, the ODP is in what may be a permanent and ultimately fatal decline. Today, the most successful directories tend to be specialty directories such as NewsDirectory.com or yourDictionary.com, and vertical search engines, such as Business.com or MedlinePlus, which focus on a particular topic instead of trying to catalog the entire Internet.

Directories were almost always a part of the **portal concept**. Portals were all the rage for a few years, while search was considered the Internet boondocks—no one was terribly interested in the boring (and unprofitable) technology of search. So where are portals now—those one-stop handy-dandy Swiss army knife websites that tried to do and be all things to all people? Most of them are gone, thanks in large part to Google's ascendancy. With its clean, spare look, Google changed the face of Internet search by moving away from the portal concept to pure search. While it is true that Google offers a directory as well as other types of searches—Image, news, shopping, groups—Google's focus has always been on web search. Google's new look, which debuted in April 2004, included removing the directory tab from the Google home page, further evidence of the decreasing importance of directories. Although there is growing criticism of the "googlization" of websites, Google continues to be the standard by which most sites are judged.

The rapid and dramatic decline in web directories is only partially attributable to Google's success. The other explanation for the waning of directories is the Tristram Shandy paradox. *The Life and Opinions of Tristram Shandy, Gentleman* is a nine-volume 18th century novel in which Tristram Shandy tries to record every detail of his life but discovers his task is hopeless because it takes him one year to document only one day. As Shandy writes an additional day, it takes him an additional year to complete the events of that day. Such is the fate, to a somewhat lesser degree, of those who seek to compile an Internet directory. By the time the information in the directory is researched, compiled, and published, the Internet has changed and made much of that information obsolete.

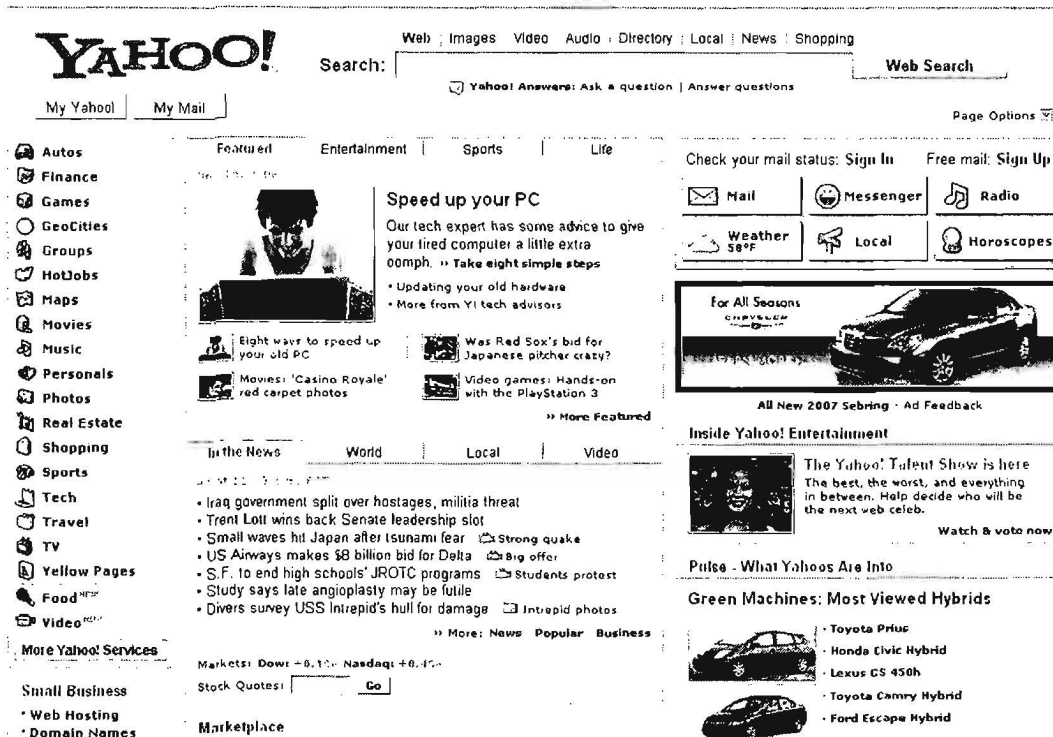
I believe Yahoo's decision to metamorphose from directory to search engine was in part a result of a tacit recognition of the Tristram Shandy paradox. Yahoo just couldn't keep up with the Internet's changes and it became too costly to try. ***Creating and maintaining a directory is an extremely manpower intensive endeavor, which flies in the face of the Internet model of relying on automation and technology.*** Undoubtedly, Yahoo's changes were largely driven by Google's enormous financial success. Yahoo sat by for years and watched as Google's popularity (and revenues) increased as Yahoo's stagnated. "By the late '90s much of [Yahoo's] focus was actually diametrically opposed to search, which is supposed to

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

send you to other sites. The Yahoo portal strategy was to keep the eyeballs on its turf, where they viewed more ad units, shopped, and bought premium services. Only when a third of online ad spending moved to search within a few short years did Yahoo decide to buy in big.”²⁰

Again in 2006 Yahoo changed the look of its homepage, but I believe Yahoo is making a fundamental error by still presenting its busy, messy portal face to the world. Although savvy Internet searchers know to go directly to <http://search.yahoo.com> in order to avoid the confusion and get a clean interface, most users are still going to the main Yahoo page where they are confronted with this:



Here's Yahoo's dilemma: how does it compete with Google for searchers seeking a simple, clean interface while simultaneously retaining and attracting users who want "one stop shopping"? Thus far, more searchers are still going to Google first rather than muddling their way through the kind of mess you see above. Where Yahoo excels—and in my opinion beats Google—is in shopping and in finding local information. This is a fact Yahoo not only recognizes but also embraces. Says Ted Meisel, head of Yahoo's Overture division, "We never claimed it [Yahoo] was a better approach for doing research on 18th century Spain. But if you are trying to buy a power washer for your back deck, it's a pretty good way to find what you

²⁰ Steve Smith, "Search Wars: Google vs. Yahoo!," *MediaPost.com*, April 2004 Issue, <http://www.mediapost.com/dtfs_dsp_mediomag.cfm?magID=245868> (registration required).

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

need.”²¹ That’s fine for personal searches, but it does not help the searcher who is using the Internet for work-related, academic, or other types of research.

The future of search seems to be in fewer but more experienced and more commercially driven hands now than a decade ago. Certainly both the quantity and quality of search results are much better today. And there are other trends in search that are going to have a major impact on users, love them or hate them. Among these are greater **personalization** of search, an area in which Google, Yahoo, and Live Search are all vying for your attention. Then there is the concept of **social networking**, through which Internet users with similar interests share their web knowledge and experience. **Social bookmarking** sites such as del.icio.us or [digg](http://digg.com) and sharing software such as [Stumbleupon](http://StumbleUpon.com) are growing in popularity as individual users seek ways to help each other discover and propagate information.

There has also been a strong impetus towards more **localized search** for shopping, news, map directions, services, telephone lookups, and more. Yahoo initially outpaced Google in this area because it already owns an enormous warehouse of information about where its users live and work, shop and play. However, Google, Yahoo, Ask, and Live Search all moved strongly into the local and personalized search arena during 2006. Add to the mix all the other services search companies offer or plan to offer, such as Google’s much ballyhooed and controversial foray into email with [Gmail](http://Gmail.com). The move toward greater personalization (likes and dislikes/interests/shopping/travel) and more services (especially email and tailored news) brings increased concerns about privacy and security. The more Yahoo, Google, Amazon, Microsoft, et al. know about us, the more they can serve up what we want.

But the more they know, the less control we have over our privacy and computer security. I am reminded of a scene from the film *Minority Report* in which the main character walks into a clothing store and, after his eye scan, the computer welcomes him by name, asks if he was happy with his previous purchase (which it details) and what he would like now. It doesn’t take a lot of imagination to see how this technology can be abused. Everyone wants convenience but it is a virtual axiom of technology that every increase in convenience brings with it some decrease in privacy and, most likely, security. Now more than ever, the future of search is one that appears to be heading towards more personalization, more features, more options and, inevitably it seems, less privacy, less security, and fewer companies with the will, technological know-how, and financial resources to build and maintain search engines.

²¹ Steven Levy, “All Eyes on Google,” *Newsweek*, 29 March 2005, p. 54, <http://www.msnbc.msn.com/id/4570868/> (14 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Understanding Search Engines

The best way to keep up to date with search engines in the US is to visit websites devoted to search and to read their newsletters. One of the oldest sites about search is Search Engine Watch. Although Search Engine Watch was originally designed for webmasters (by webmaster Danny Sullivan), it is a good resource for researchers who want and need in-depth information about the major English-language search services and some country specific engines. Search Engine Watch is also home to **Search Day**, noted search maven Chris Sherman's daily newsletter. While Search Day is kept current, Search Engine Watch now has many out of date pages.

Stepping into the breach is the superb **Pandia Search Central**, which offers current search news and an almost endless number of tips, tutorials, guides, and even its own search tools. Pandia has emerged as the premier site for news about and help with search.

A guide to internet searching and search engine marketing

Search the Web:

Select search engine: Pandia Powersearch

Ask for [More Space & Help](#) | [More Tools](#) | [Tutorial](#) | [Web Directory](#) | [News Search](#)

PANDIA

[Search Engine News](#)
[Search Engine Resources](#)
[SE Detective](#)
[Free Newsletter](#)
[RSS feed](#)

Search tools:
[Powersearch All-in-One](#)
[Plus Web Directory](#)
[Metasearch](#)
[Newsfinder](#)
[Radio Search](#)
[People Search](#)

On Web Searching:
[Search Tutorial](#)
[Svntax Q-cards](#)

On Search Ranking:
[SE Marketing Tutorial](#)
[SE Optimization Gateway](#)
[SE Submission](#)
[Pay Per Click SE](#)

On Pandia:
[Search this Site](#)
[Pandia FAQ](#)
[Store](#)

EVERYTHING SEARCH ENGINE

On Web searching and search engine marketing

Pandia has all you need to know about search engines, Internet searching and search engine marketing.

- Learn [how to search the Web](#) more efficiently.
- [Read about search engines](#) and sites devoted to searching.
- Gain easy access to all the [best search tools](#)
- Learn about [search engine marketing](#).
- Find the best online [search engine optimization resources](#)

Our latest search engine news and articles

- [Search Marketing ? More than Direct Response](#)
- [Seven elements for building an effective corporate blog strategy](#)
- [Search for images at the New York Public Library Digital Gallery](#)
- [Search Engine Marketing Inc., Driving Search Traffic to Your Company's Web Site](#)

Our search engine news column is updated several times a week, and includes more search engine news and in depth feature articles. If you subscribe to our [weekly newsletter](#), we will send you these stories by mail for free

How to get high search engine rankings

FREE!

Free newsletters!

Get Pandia Search World, our free weekly news update on Web searching and search engine optimization.

Subscribe also to The Pandia Post, our [free bimonthly newsletter](#), and get articles on Internet searching, search engines and ranking.

Pandia Search World (weekly)
 The Pandia Post (bimonthly)

Enter email address:

We will never give your address to any other company or organization
[Read our privacy policy](#)

Ads by Google

STORE

In depth search engine marketing tutorial with newsletter

The search engine landscape is changing dramatically. How are you to keep track of all these changes?

[The Unfair Advantage Book on Winning the Search Engine Wars](#) is the book the marketing professionals have been using to win the search engine optimization race for the past seven years. You get a new version every month, updated by people who know the industry from the inside.

Other good web search sites include John Battelle's Searchblog, Philipp Lenssen's Google Blogscoped (which covers much more than just Google), Gary Price's

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Resource Shelf, Phil Bradley's Weblog, Greg Notess's Search Engine Showdown, as well as Web Master World and Web Search Guide. Among the best search engine-specific blogs are the Yahoo Search Blog, the Official Google Blog, Google Operating System, and Live Search Weblog.

The only thing predictable about search engines is how quickly and frequently they change not only their content but also their features. Because there are websites devoted to keeping up with the myriad changes, they are your best bet for staying on top of the ever-changing world of search tools.

Search News and Blogs

| | |
|--|---|
| Google Operating System | http://googlesystem.blogspot.com/ |
| John Battelle's Searchblog | http://battellemedia.com/ |
| Live Search Weblog | http://blogs.msdn.com/msnsearch/default.aspx |
| Official Google Blog | http://googleblog.blogspot.com/ |
| Pandia Search Central | http://pandia.com/ |
| Philipp Lenssen's Google Blogoscoped | http://blog.outter-court.com/ |
| Phil Bradley's Weblog | http://philbradley.typepad.com/phil_bradleys_weblog/ |
| Research Buzz | http://www.researchbuzz.com/ |
| Resource Shelf | http://www.resourceshelf.com/ |
| Search Day | http://searchenginewatch.com/searchday/ |
| Search Engine Showdown | http://www.searchengineshowdown.com/ |
| Search Engine Showdown Reviews | http://www.searchengineshowdown.com/reviews/ |
| Search Engine Watch | http://searchenginewatch.com/ |
| Search Engine Watch Web Searching Tips | http://www.searchenginewatch.com/facts/index.html |
| Web Master World | http://www.webmasterworld.com/ |
| Web Search Guide | http://www.websearchguide.ca/ |
| Search Engine Watch Blog | http://blog.searchenginewatch.com/blog/ |
| Yahoo Search Blog | http://www.ysearchblog.com/ |

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

 **Web Tip**

Browsers assume the prefix “http://” unless you tell them otherwise, which means you do not need to type “http://”— just type the url (address).

Search Engine Basics

A search engine comprises three basic parts:

1. The **spider/robot/crawler** is software that “visits” sites on the Internet (each search engine does this differently). The spider reads what is there, follows links at the site, and ultimately brings all that data back to:
2. The **search engine index**, catalog, or database, where everything the spider found is stored;
3. The **search engine software** that actually sifts through everything in the index to find matches and then ranks or sorts them into a list of results or hits.

Important points to consider about search engines:

- Spiders are programmed to return to websites on a regular basis, but the ***time interval varies widely*** from engine to engine. Monthly or better is considered “fresh.”
- When you use a search engine, you are searching the index or database, not the web pages themselves. This is important to remember because ***no search engine operates in “real time.”***
- Spiders do not index all web pages they find, including pages that employ the “**Robots Exclusion Protocol**” or the “**Robots META tag**.” The first of these mechanisms is a special file website administrators use to indicate which parts of the site should not be visited by the robot or spider. The second is a special HTML metatag that may be inserted by a web page author to indicate if the page may be indexed or analyzed for links. Not every robot/spider

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

respects these mechanisms. Password protection, firewalls, and other measures will generally keep spiders from crawling a website and indexing it.

The Web Robots Pages ◀

Robots Exclusion

Sometimes people find they have been indexed by an indexing robot, or that a resource discovery robot has visited part of a site that for some reason shouldn't be visited by robots.

In recognition of this problem, many Web Robots offer facilities for Web site administrators and content providers to limit what the robot does. This is achieved through two mechanisms:

The Robots Exclusion Protocol A Web site administrator can indicate which parts of the site should not be visited by a robot, by providing a specially formatted file on their site, in `http://.../robots.txt`.

The Robots META tag A Web author can indicate if a page may or may not be indexed, or analysed for links, through the use of a special HTML META tag.

The remainder of this pages provides full details on these facilities.

Note that these methods rely on cooperation from the Robot, and are by no means guaranteed to work for every Robot. If you need stronger protection from robots and other agents, you should use alternative methods such as password protection.

Robots Exclusion Page

<http://www.robotstxt.org/wc/exclusion.html>

Not every search engine has its own proprietary search program but instead relies upon another company's search service for its results. Most of these **strategic alliances** now involve Yahoo, Google, and Windows Live Search. All these partnerships are subject to change without notice; for more on these strategic alliances, see:

Search Engine Alliances

<http://searchenginewatch.com/reports/alliances.html>

Knowing that Yahoo, for example, is the search tool behind a search engine can save you time because you can be pretty sure that using AltaVista will get you similar (although not identical) results to the other search engines also powered by Yahoo. It is critical to remember that **each service powered by a particular search engine produces different results** even though they may all use the same core database. Why is this? Because the search interfaces have their own algorithms that decide how queries are run, how results are returned, or even if they query the entire database (most do not). In short, go to the primary search engine—Google, Yahoo, or Live Search for best results.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

A Word About Browsers: Internet Explorer and Mozilla Firefox

Two years ago I declared that the "browser wars" were for all intents and purposes over and Microsoft's Internet Explorer (IE) had won. IE still commands more than 90 percent of the world's browser market, and AOL abandoned Netscape's Navigator/Communicator in mid-2003. However, during 2004, Mozilla browsers experienced a resurrection thanks largely to user frustration.

Caveat Browser

Alexa and Smart Browsing technology are very controversial because of their invasion of privacy implications. For more information, take a look at the article:
"What's Related? Everything But Your Privacy"

Curtin, M., Ellison, G., Monroe, D., "What's Related? Everything But Your Privacy," 7 Oct 1998, Revision: 1.5,
<<http://www.interhack.net/pubs/whatsrelated/>>
(24 October 2006)

Because of Internet Explorer's continued dominance of the browser market and, more importantly, because it is the standard browser for many *Untangling the Web* readers, I will focus my attention on Internet Explorer.

Nonetheless, all browsers have advantages and drawbacks. ***I still recommend you configure two browsers, both Internet Explorer and Mozilla Firefox.*** Both types of browsers allow you to make a number of decisions that affect your privacy and security while browsing. Also, both browsers have become much more customizable with each new release, allowing every user to select and save his own preferences for everything from fonts to what will appear on the toolbar. Be sure to familiarize yourself with the many evolving features of your browser(s). The Microsoft and Mozilla websites have extensive information and documentation about their browsers. At the Mozilla site you can download and install the highly regarded Firefox browser as well as other free software, such as the Thunderbird email program.

In October 2006, both Microsoft and Mozilla introduced new versions of their browsers: Internet Explorer 7 (IE7) and Firefox 2. Microsoft, which had owned upwards of 90 percent of the browser market until Firefox took off a couple of years ago, recognized it has a genuine competitor on its hands and made significant changes and improvements to its browser to try to bring some Firefox users back

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

into the fold. Will it work? *PC World* offered an excellent comparison of IE7 and Firefox 2.²² While Firefox 2's changes are mostly refinements of already existing features with no change in the browser's look and feel, IE7 marks a major overhaul since IE6 was released way back in 2001.

Among the changes to **Internet Explorer 7** are tabbed browsing, integrated searching, RSS newsfeed support, and an antiphishing tool. The most noticeable change is IE7's look and feel, which is designed to resemble Microsoft's new operating system, Vista. Probably the most obvious and popular addition to IE7 is tabbed browsing, something Firefox already offered. Also, IE7 has a built-in search box, which lets users search from anywhere without having to go to the search engine's home page. Google and other search engines had successfully lobbied Microsoft not to make Live Search the default search service, so you can pick your search engine.

The other major change is invisible: improved security features designed to cope with the almost endless number of vulnerabilities that have afflicted IE6.²³ The most prominent of these security upgrades is one shared with Firefox: an "antiphishing" tool that works by warning users that a website they are about to visit may be fake and redirects them away from the page unless they actively choose to go to it. The other major new IE7 security feature is something called Protected Mode, which prevents a website from changing a computer's files or settings. However, Protected Mode will not work with any Windows operating system except Vista, which is due out next year. Also, one of IE's major appeals had been its universality, that is, it would work with most websites. The security features in IE7 mean that some sites that could be viewed in earlier versions of IE cannot be viewed in IE7, undermining one reason many people still continued to use the Microsoft browser.

Firefox 2 is another in a long line of gradual updates. This version adds a spell checker, a system for suggesting popular search terms, and an option to pick up where you left off after a crash. Firefox 2 also upgrades the RSS newsfeed so that now, if you click on the feed itself, instead of seeing the usual XML gibberish, Firefox 2 will parse the raw feed into something readable and also subscribe to a feed using one of a numerous (but not all) newsreaders.

What is the bottom line? Firefox users should upgrade to version 2; it will be easy and pain free. IE6 users probably should wait a while before downloading IE7 to let

²² Erik Larkin, "Radically New IE7 or Updated Mozilla Firefox 2--Which Browser is Better?" *PC World*, 18 October 2006, <<http://www.pcworld.com/printable/article/id,127309/printable.html>> (24 October 2006).

²³ Not 24 hours after its release and the first vulnerability was detected in IE7. Of course, it also affects IE6, but this is embarrassing for Microsoft given that the company has touted the security of IE7 over its predecessor.

<[http://secunia.com/Internet Explorer Arbitrary Content Disclosure Vulnerability Test/](http://secunia.com/Internet_Explorer_Arbitrary_Content_Disclosure_Vulnerability_Test/)>

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

early adopters find the inevitable bugs that Microsoft will have to fix. Frankly, after five years, you would think Microsoft could do better than come up with a browser that basically mimics the best features of Firefox and its other (much smaller) competitors. This looks mostly like catch-up and very little like innovation.

If you are going to use Netscape, another Mozilla-based browser, I do not recommend using Netscape 8x because it has many reported problems. Stick with either Netscape 7.1x or 7.2x. Also, if you prefer a streamlined version of Netscape 7x without all the annoying "extras," I can recommend one from Sillydog (silly name, great tool). "Netscape 7.1 is based on Mozilla 1.4. Both applications share almost identical features, such as tabbed browsing, custom keywords, and Sidebar. Exceptions are additions of proprietary features such as the support for Netscape WebMail and AOL mail."²⁴ Netscape 7.2 is based on Mozilla 1.7.2. "In addition to the technologies that Netscape 7.2 shares with Mozilla 1.7.2, it includes additional features such as a number of installed plugins, support for Windows Media Player Active X control which are not available in Mozilla."²⁵

Microsoft Internet Explorer <http://www.microsoft.com/windows/ie/default.htm>

Mozilla Firefox <http://www.mozilla.com/firefox/>

Netscape 7.1 Streamline <http://sillydog.org/narchive/sd/71.html>

Netscape Archive (7.1 or 7.2) <http://browser.netscape.com/ns8/download/archive.jsp>

²⁴ Mozilla FAQ, <<http://www.mozilla.org/start/1.4/faq/general.html#ns7>> (14 November 2006).

²⁵ Sillydog.org Browser Archive, 31 October 2005, <<http://sillydog.org/narchive/full67.php>> (24 October 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

What the heck are “cookies”?

Cookies are text placed on your computer's hard disk (yes!) by a website in order to remember something about you. For example, a site may set a cookie that enables you to reenter without logging in or customize its pages based on the type of browser you're using. Cookies remain controversial (more later).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

The Great Internet Search-Offs

Over the last decade, the inevitable “search offs” have become commonplace (both Internet vs. traditional researcher and Internet researchers against each other). Some of the findings of these “contests” provide insight into how search engines work.

1. Most search-offs and wide-ranging studies continue to find surprisingly little overlap among major search engines, so **use more than one search engine** as a general rule.
2. The Internet is now being widely used for “serious” research, which means higher quality, more reliable information on the web. But, as with any research source, you must **weigh the validity, accuracy, currency, and overall quality** of the information before using it.
3. Search engines rely on statistical interfaces, concept-based search mechanisms, or link analysis to return and rank hits; using **boolean expressions**²⁶ usually interferes with or defeats these statistical approaches. In general, **do not use boolean queries** unless you know exactly what you are looking for and are very comfortable with that search engine’s boolean rules (no, they are not all the same; for example, you may have to use CAPS for all operators). Also, many search engines do not correctly process nested boolean queries (boolean searches with parentheses).
4. Be aware that search engines are giving **more weight to popular and/or pay-for-placement web pages**. In fact, most search engines use services to determine which are the most visited, and therefore most popular, websites and return them at the top of the results list. This is a strategic move away from the traditional “words on a page” ranking system. Trustworthy search engines will clearly indicate which hits are paid entries.

²⁶ The term “boolean,” often encountered when doing searches on the web (and frequently spelled “Boolean”), refers to a system of logical thought developed by the English mathematician and computer pioneer George Boole (1815-64). In boolean searching, an “and” operator between two words or other values (for example, “pear AND apple”) means one is searching for documents containing both the words or values, not just one of them. An “or” operator between two words or other values (for example, “pear OR apple”) means one is searching for documents containing either of the words. “Boolean,” SearchSMB.com, <http://searchsmb.techtarget.com/sDefinition/0,290660,sid44_gci211695,00.html> (14 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

5. **Learn the search syntax** of the search engines you use (never assume). Most search engines use double quotes (“”) to enclose a phrase and the plus + and minus - keys to indicate “must include” and “must exclude” respectively. But these are by no means universal rules (especially when using *international* or *metasearch* engines).
6. **The default operator for all major US search engines is now AND.** As of February 2002, no major search engine used OR as its default operator. However, most search engines will let you use an OR in the simple search box: Yahoo and Google permit OR searches in the simple search box, but you must capitalize the OR.
7. Keep in mind that because **HTML does not have a “date” tag**, “date” can mean many things: creation date; the last modified date for the page; or the date search engine found the page. *I do not recommend searching by date except when using weblog, news, or newsgroup search engines.*

Understanding **statistical interfaces** is important, especially for researchers used to boolean and other non-statistical query languages. Most search engines use statistical interfaces. The search engine assigns relative weights to each **search term**, depending on:

- its rarity in their database
- how frequently the term occurs on the webpage
- whether or not the term appears in the url
- how close to the top of the page the term appears
- (sometimes) whether or not the term appears in the metatags.

When you query the database, the search engine adds up all the weights that match your query terms and returns the documents with the highest weight first. **Each search engine has its own algorithm** for assigning weights, and they tweak these frequently. In general, rare, unusual terms are easier to find than common ones because of the weighting system.

However, remember that “popularity” measured by various means often trumps any statistical interface.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Types of Search Tools

Before delving into the intricacies of search engines, let's look at some other ways of finding information on the web. Search engines are not the only and often not even the best way to access information on the Internet.

Web Directories/Subject Guides/Portals

Web directories are organized subject catalogues that allow the web searcher to browse through lists of sites by subject in search of relevant information. **Yahoo**, **Galaxy**, **Google Directory**, **Lycos**, and the **Open Directory** are select lists of links to pertinent websites. Directories were once viewed as the future of the Internet because they could sift through the mountains of information and millions of websites to offer only the best and most relevant. However, directories have truly fallen by the wayside over the past several years with the rise of Google and, even more importantly, wikis in general and Wikipedia in particular. Directories continue to recede in importance and value to researchers as they are increasingly replaced by better alternatives, including Custom Search, by which a voluntary community of searchers shares expertise to create more focused searches with more relevant results. The reason for the decline of directories is obvious: directories are simply too manpower intensive and expensive to keep up with the ever-changing and expanding web. I would say at this point directories, while not dead, are probably moribund.

Directories rely on people to create their listings

Obviously, this is a much more labor-intensive business than operating a search engine robot. Websites indexed in a directory are either described/evaluated by editors/reviewers or rely on descriptions provided by web page owners who may pay for placement in a directory. When you search a directory, the only retrievals will come from those descriptions, so keep this in mind. Although directories give you a much more limited view of the web, directories do have their own utility. Most directories also have a backup search that provides responses to queries that don't match anything in the directory listings.

Directories may produce more relevant results

Subject guide databases are always smaller than those of search engines, which means that the number of hits returned tends to be smaller as well. On the bright

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

side, this means the results directories produce are often more relevant. For example, while a search engine typically indexes every page of a given website, a subject guide is more likely to provide a link only to the site's home page. For this reason, they *lend themselves best to searching for information about a general subject*, rather than for a specific piece of information.

Yahoo still has the best-known subject guide/directory and can be a good starting place for research, even on technical subjects. Yahoo used to list links alphabetically, but once Google came along with its ranked list of sites, Yahoo started offering most popular sites first before going to its alphabetical list. However, Yahoo's directory has suffered in recent years as the **Google Directory** has steadily improved. Google gets its directory data free in the form of the **Open Directory Project**.

You may not recognize the Open Directory Project by this name, but you have probably used it. The ODP is the directory behind the Google Directory, AOL Search, Yahoo Directory, and many others. The ODP "is the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a vast, global community of volunteer editors."

Galaxy is definitely worth a look because it was designed for and by "professionals," so it has a bent toward business, technology, and science that other directories lack. You may search either the Galaxy collection or the web using their proprietary search engine. **Best of the Web** started life in 1994 as a web awards site and is now a full-fledged directory.

Many more specialized directories are discussed under the "Invisible" Internet.

| | |
|------------------|---|
| Best of the Web | http://botw.org/default.aspx |
| Galaxy | http://www.galaxy.com/ |
| Google Directory | http://directory.google.com/ |
| Open Directory | http://dmoz.org/ |
| Yahoo Directory | http://dir.yahoo.com/ |

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Metasearch Sites

The growth in the number of search engines has led to the creation of “meta” search sites. These services allow you to **invoke several or even many search engines simultaneously**. These metasearchers may do a more thorough job of sifting through the net for your topic than any single search engine. If you are new to using search engines, these are a great way to do a very broad search, while familiarizing yourself with the popular engines and how they respond. But metasearch engines inevitably **lack the flexibility** of individual search tools.

It is important to note that many metasearch engines do not employ some of the best search engines, such as Google and Yahoo. Also, my biggest complaint about metasearch engines is that **they perform shallow searches**, usually only retrieving the top ten or so hits from a site, which is far too few to be comprehensive or truly representative of what is “out there.”

However, metasearch engines do serve a purpose. If you are unsure if a term will be found anywhere on the web, try a metasearch engine first to **“size” the problem**: you may get zero hits with a dozen search engines (you’ve got a problem) or you may get a half-dozen right-on-the-money hits right off the bat.

Clusty

<http://clusty.com/>

Vivisimo, in my opinion the best free metasearch tool available, opened a new search site—Clusty—in 2004 and then made Clusty its search home in 2006. Fundamentally, Vivisimo and Clusty are the same, but Clusty adds options for news, image, [Wikipedia](#), government, and [blog searches](#).

The Vivisimo technology behind Clusty is unique because it employs its own **clustering engine**, software that organizes unstructured information into hierarchical folders. Clusty offers clustered results of web, news, and certain specialty searches. The Clusty default is to search the web using Live Search, Gigablast, Ask, Wikipedia, and the [Open Directory](#).

Clusty is especially useful for searching ambiguous terms, such as *cardinal*, because it clusters them by logical categories, as shown below.

UNCLASSIFIED//FOR OFFICIAL USE ONLY

The screenshot shows the Clusty search interface. At the top, there are navigation links for 'web', 'news', 'images', 'wikipedia', 'blogs', 'jobs', and 'more'. A search bar contains the word 'cardinal' and a 'Search' button. Below the search bar, it indicates 'Top 244 results of at least 18,087,488 retrieved for the query cardinal (definition) (details)'. On the left side, there are tabs for 'clusters', 'sources', and 'sites'. Under 'clusters', a list of categories is shown with counts: St. Louis Cardinals (21), University (12), Catholic (12), Cardinal Health (17), Northern Cardinal (5), Arizona Cardinals (8), Designs (9), Cardinal Ratzinger (13), and Cardinal number (4). There is a 'find in clusters' search box and a 'Find' button. The main content area is titled 'Top News' and features a 'Find more news stories' link. A small image of a cardinal is shown next to a news snippet: 'Pope names Brazil cardinal to key post'. Below this, there are several search results listed:

- Cardinal Stritch Univ**: Info on Attending Cardinal Stritch Univ, Plus a Free Info Booklet - CardinalStritchUniv.com
- St. Louis Champs Gear**: 2006 Cards Champs Merchandise - T-Shirts, Hats, Mugs, & Souvenirs. - www.StLouisChamps.com
- Cardinal Health - Making healthcare safer and more productive**: Offers drug development services in the pharmaceutical and biotechnology industry. Contains information on company, investor relations and news. www.cardinal.com - [cache] - MSN, Open Directory, Wisenut, Ask, Gigablast
- Cardinality**: In mathematics, the cardinality of a set is a measure of the "number of elements of the set". There are two approaches to cardinality - one which compares sets directly using bijections, injections, and surjections, and another which uses cardinal numbers. en.wikipedia.org/wiki/Cardinality - [cache] - Wikipedia
- Cardinal Stritch University**: The Cardinal Stritch University Theatre Arts department will present ... In honor of the Feast of St. Francis, Cardinal Stritch University More. www.stitch.edu - [cache] - Gigablast, Open Directory, Wisenut, Ask
- Cardinal**: The word cardinal comes from the Latin *cardo* for "hinge" and usually refers to things of fundamental importance, as in

Also, Clusty lets users look at the sources of the search results and types of sites (e.g., .com, .gov). Clusty has a unique feature that allows users to search inside clusters. In this example, the original search was [iran] and the "find in clusters" search was [nuclear]. Here are the results of this recursive search looking at the sources of data:

Clusty [web](#) [news](#) [images](#) [wikipedia](#) [blogs](#) [jobs](#) [more »](#)

iran [advanced preferences](#)

clusters **sources** sites

24 of the top results retrieved for the query iran match the sub-query

All Results (254)

- Ask (35)
- Gigablast (90)
- MSN (84)
- NY Times (16)
- Open Directory (192)
- Sponsored Listings (4)
- Wikipedia (19)
- Wisemart (30)
- Yahoo! News (10)

find in clusters: [remove highlighting](#)

Font size:

- Iran may limit work with IAEA if sanctions imposed**

4 hours ago - Iran may limit its cooperation with the U.N. nuclear watchdog if sanctions are imposed over the Islamic Republic's atomic program, a former chief nuclear negotiator was quoted as saying on Wednesday. European states have prepared a draft sanctions resolution after... Iran rejected repeated U.N. demands to scrap uranium enrichment, which can be used to make material for power stations or warheads. Russia has expressed misgivings about the proposal. Hassan Rohani, a moderate politician who led Iran's nuclear negotiations with the European Union since 2003 until hardline President Mahmoud Ahmadinejad took office last year, warned about...
news.yahoo.com/s/irnm/20061101/ir_nm/nuclear_iran_iaea_dc - [cache] - Yahoo! News
- Iran to pay incentives to attract tourists**

8 hours ago - Iran will offer cash incentives to travel agencies to encourage Western tourists to visit the country, giving a premium for Americans, the official Islamic Republic News Agency reported. The Islamic republic's political leadership has been trying to reach out to... ordinary Americans to show that a standoff over Iran's nuclear ambitions is with the Bush administration not U.S. citizens. The latest initiative comes as the United Nations Security... Council deliberates a draft resolution that would impose sanctions on Iran for its disputed nuclear program. "Iran's tourism department will pay \$20 per person to those who attract European or...
news.yahoo.com/s/aph/061101/ir_nm/iran_us_tourism - [cache] - Yahoo! News
- NTI: Country Overviews. Iran Profile**

Introduction. Iran's chemical weapons and ballistic missiles, and possibly its nuclear weapon program and biological warfare capabilities, are meant to deter opponents and to gain influence in the...
www.nti.org/research/profiles/iran/index.html - [cache] - MSN
- Moscow may back U.N. sanctions on Iran**

13 hours ago - Two senior Russian officials indicated Tuesday that Moscow could back a draft U.N. Security Council resolution imposing sanctions on Iran, in an apparent sign of the Islamic Republic's growing isolation over its nuclear program. The comments came after months of Russian opposition to sanctions and suggested that Moscow could reluctantly support punishment as a lever of influence over a stubborn Iran. Just last Thursday, the foreign minister said Moscow was opposed to the resolution

For news, Clusty searches the *New York Times*, Associated Press, Reuters, and Yahoo News (which subsumes a huge number of sources). One of the best features of Clusty news search is the ability to toggle among clustered results, sources, and sections (such as business, health, tech, science).

Clusty [web](#) [news](#) [images](#) [wikipedia](#) [blogs](#) [jobs](#) [more »](#)

iraq [advanced preferences](#)

clusters **sources** sections

Top News World U.S. Business Sports Health Tech Science

Top 205 results of at least 686 retrieved for the query iraq (definition) (details)

All Results (205)

- AP (3)
- NY Times (65)
- Reuters (58)
- Yahoo! News (60)

find in clusters: [remove highlighting](#)

Font size:

Yahoo! News »

- Canada welcomes 63 Palestinians stranded between Iraq, Jordan (Yahoo! News) 14 mins ago
- Military chart shows Iraq sliding toward "chaos" (Yahoo! News) 15 mins ago
- Gunmen kidnap Iraqi coach for blind (Yahoo! News) 54 mins ago

Reuters »

- Results in key House races: Reuters poll (Reuters) 8 hours ago
- Kerry says sorry for "botched joke" (Reuters) 2 hours ago
- Iraqis divided over checkpoints order (Reuters) 3 hours ago

AP »

- Hezbollah threatens street protests (AP) 7 hours ago
- Iran says it needs \$100 billion in aid (AP) 14 hours ago
- Democrats increase hope of winning House (AP) 16 hours ago

NY Times »

- A Senate Newcomer, Helping Fellow Democrats on the Trail and Drawing Big Crowds (NY Times) 54 mins ago
- Democrats Discover New Political Frontier (NY Times) 54 mins ago

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Clusty also provides a number of advanced search options and preferences, including the option to add your own customized tabs to the main search page.

Clusty stands out as one of—if not indeed—the best metasearch tools available for free and without registration on the Internet. When clustering works (and the Vivisimo technology was independently rated as accurate 90 percent of the time), it offers advantages for automatically grouping huge amounts of information logically. Because there is no human intervention, Vivisimo's clustering algorithm "also helps in discovering new areas of subject development, avoiding the 'mummy's curse,' in which human catalogers have to recognize a term before approving it for usage and then leaving the earlier material using the term un-indexed and irretrievable by that term as an authorized descriptor or metatag."²⁷

Jux2

<http://www.jux2.com/>

Jux2 lets users query three search engines—Google, Yahoo, and Live Search (still referred to as MSN Search)—and then shows you:

1. The Best Results from all three search engines and the total hits for each.
2. What only Google found and what is missing from Google.
3. What only Yahoo found and what is missing from Yahoo.
4. What only Live/MSN Search found and what is missing from Live/MSN Search.

²⁷ Barbara Quint, "Vivisimo Clustering Chosen to Enhance Searching at Institute of Physics Publishing Site," *Infoday*, 25 March 2002, <<http://www.infoday.com/newsbreaks/nb020325-2.htm>> (14 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//FOR OFFICIAL USE ONLY

jux [advanced search preferences](#)
 A Better Way to Search? [Best Results](#) [Compare Google's Results](#) [Compare Yahoo's Results](#) [Compare MSN's Results](#)

What only **Google** found

ATROPOS: The deity from Greek Mythology
 Greek Mythology. Meet the Classical Gods of Ancient Greece. **ATROPOS**: Oldest of the three FATES. She's the one who cuts the Thread of Life with her fatal ...
<http://www.gardhearts.com/pantheon/greek-mythology.php?deity#ATROPOS> [# 6 on Google]

www.katsudon.net/gwfi/atropos.html
<http://www.katsudon.net/gwfi/atropos.html> [#7 on Google]

Amazon.com: Hornblower and the **Atropos** (Hornblower Saga (Paperback ...
 Amazon.com: Hornblower and the **Atropos** (Hornblower Saga (Paperback)): Books: CS Forester by CS Forester.
http://www.amazon.com/used/books?pf_rd_p=16316359297v=qlenx [# 9 on Google]

What's missing from **Google**

<http://en.wikipedia.org/wiki/Atropos>
 From Wikipedia, the free encyclopedia. **Atropos** is also a British entomological journal - see **Atropos** (journal). In Greek mythology, **Atropos** was the third of the Moirae. ... It was **Atropos** who chose the mechanism of death and ended the life of each mortal by cutting their thread ...
<http://en.wikipedia.org/wiki/Atropos>. [# 3 on Yahoo!, #3 on MSN]

Blogger: User Profile: **Atropos**
Atropos Gender: female Industry: Fashion Occupation: Model Location: Ohio : United States About Me I'm a 22 year old female who enjoys exploring deviant fashion as a form of self expression
<http://www.blogger.com/profile/4634126> [#5 on MSN]

I believe you will be as surprised as I was to see how little overlap there often is among the "big three" search engines.

Dogpile

<http://www.dogpile.com/>

Dogpile, despite its name, is a good metasearch engine. Dogpile includes Live Search results, along with those from Google, Yahoo, and Ask Jeeves. This is, of course, very good news because Dogpile is now drawing from all the major US-based search engines with the exception of Gigablast. It also searches smaller or lesser-known search engines and directories, including the MIVA (formally FindWhat), LookSmart, Ask, About, and more. Interestingly, the European version's name is Webfetch because of "unfortunate associations" between Dogpile and manure.

Mamma

<http://www.mamma.com/>

Mamma, the "Mother of All Search Engines," might just be exaggerating a wee bit. Mamma offers web, news, image, and yellow and white page search options.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Search engines queried are Ask, Wisenut, Gigablast, and Entireweb (a serious misnomer) and directories queried are Open Directory, About, Business.com, and two pay-per-click sources.

The Pandia Metasearch Engine <http://www.pandia.com/metasearch/index.html>

The famed search guide site, Pandia, offers its own excellent metasearch engine. The Pandia metasearch engine “collects and sorts the hits, takes out duplicates, and presents the end result in a simple format. “The first results you’ll see are from what Pandia describes as the “essential search engines and directories,” which include Google, Yahoo, HotBot and Wisenut. Strangely, Pandia continues to list AlltheWeb (Fast) and AltaVista as search engines while they acknowledge at other places on their site that Yahoo subsumed both engines. Still, this is a very good metasearch site.

More metasearch sites:

| | |
|---|---|
| Ithaki | http://www.ithaki.net/indexu.htm |
| IxQuick | http://www.ixquick.com/ |
| Metacrawler | http://www.metacrawler.com/ |
| Search.com | http://www.search.com/ |
| Surfwax | http://www.surfwax.com/ |
| Open Directory’s List of Metasearch Sites | http://dmoz.org/Computers/Internet/Searching/Metasearch/ |

Megasearch Sites

Megasearch sites simply store several search engines under one roof, but you have to do the searches one search engine at a time. They are becoming more sophisticated and better as time passes, serving as good entry points for finding and evaluating search engines. They are especially useful for **locating international search engines**.

| | |
|--------------------|---|
| All Search Engines | http://www.allsearchengines.com/ |
| Find It Quick | http://www.quickfindit.com/Search_Engines/ |
| Search—22 | http://www.search-22.com/ |
| SearchEzee | http://www.searchezee.com/search.shtml |

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Types of Searches and the Best Ways to Handle Them

The first thing to ask yourself is the one question a lot of people never consider: is the Internet the best place to start? In general, the Internet has become so good at **answering factual questions**—the kinds of things you find in an almanac, an encyclopedia, or a phone book—that it is now usually better in terms of speed, timeliness, and accuracy than other resources. For example, if I need to know the world's largest hydroelectric plants, I can open an almanac and look up this information or I can type [world's largest hydroelectric plants] into Google, Yahoo, or Live Search, where the first result links me to a page at **Information Please.com** that contains the answer to the question.

Still, compared to traditional library-type resources, the Internet may be:

- slower (though this is changing with new technologies).
- less reliable (large amounts of bad data in among the good).
- disorganized (a library with all the books on the floor).
- frustrating (lots of “broken” links).
- hard to use (generally poor search tools and too much data to sift through).
- risky because of growing privacy and security threats.

This being said, why do we need to use to the Internet? Because:

- it has almost **unlimited amount** of data (also a minus...too much of a good thing and way too much of the bad).
- the data tend to be **current**.
- it offers **multimedia** (video, audio, charts, tables, illustrations).
- it allows the individual to do much more of his **own research**.
- it is relatively **inexpensive** (at least in some countries).
- most importantly, it contains a vast amount of **unique** information.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

You've thought through your research question and decided to use the Internet to find information either because you've already tried traditional sources without success or you believe the Internet is your best option. You're sitting in front of your terminal, you've logged onto the Internet and you're staring at a blank screen. Now what? Let's start with a (relatively) easy type of search. You need to find **general information about a fairly broad topic**.

Let's say you need to research a broad topic unfamiliar to you, for example, Java. The best approach may not be to type *java* into a search engine. Why? Because you'll probably get millions of hits, and the first ones may be to commercial sites trying to sell you something relating to Java and will undoubtedly also include other meanings of Java, such as Indonesia and coffee. If you are looking for general information on a topic, **wikis, specialized (vertical) search engines, and virtual libraries** are often better starting points for researching general or broad topics than big search engines.

The single biggest mistake searchers make is using the wrong search tool. For example, search engines are generally not the best tools for finding current news (use a news search engine), for researching broad topics (use a specialty directory or virtual library), or for performing specialized searches such as scientific research (use a specialty search engine). That's why the number one rule for web research is:

Rule One

Use the right tool for the job.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Let's go back to the Java example where you want to find general information on the web about Java programming. Start with the **Yahoo** directory and see what categories it offers on Java. You can ignore the sponsored results and the categories about Indonesia, classic arcade games, and commercial Java services. Instead, your best bet is Programming Languages > Java:

YAHOO! SEARCH Directory

Search: [the Web](#) | [the Directory](#) | [this category](#)

Programming Languages > Java [Email this page](#) [Suggest a Site](#) [Advanced Search](#)

[Directory](#) > [Computers and Internet](#) > [Programming and Development](#) > [Languages](#) > [Java](#)

CATEGORIES

Industry Categories

- ◆ [Developers@](#)

Additional Categories

- ◆ [Applets](#) (45)
- ◆ [Books@](#)
- ◆ [Classes](#) (3)
- ◆ [Compilers](#) (9)
- ◆ [Events](#) (2)
- ◆ [FAQs](#) (5)
- ◆ [Games](#) (65)
- ◆ [Guides, Tutorials and Documentation](#) (27)
- ◆ [Java 2 Platform, Enterprise Edition \(J2EE\)](#) (5)
- ◆ [Java 2, Micro Edition \(J2ME\)](#) (8)
- ◆ [Java Advanced Imaging \(JAI\)@](#)
- ◆ [Java Virtual Machines \(JVMs\)](#) (5)
- ◆ [Jini@](#)
- ◆ [Magazines](#) (4)
- ◆ [Mailing Lists](#) (1)
- ◆ [People](#) (3)
- ◆ [Porting Projects](#) (4)
- ◆ [Security](#) (21)
- ◆ [User Groups](#) (18)
- ◆ [Utilities](#) (5)
- ◆ [Web Directories](#) (6)

SITE LISTINGS [By Popularity](#) | [Alphabetical](#) | [\(What's This?\)](#)

SPONSOR RESULTS

[CocoBase O/R Persists Data for Java](#)
Get CMP, BMP, Session, Transparent Persistence for Java applications. ...
[www.thoughtinc.com](#)

[Hands-on Java Training in 75 Cities](#)
Instructor-led Java courses in more than 75 cities across the U.S. ...
[www.traininghott.com](#)

[Java Code Generation](#)
Generate java classes from XML schemes or UML models, or generate ...
[www.altova.com](#)

[Complete Java Training](#)
Hands on instructor-led complete java training.

[Computers_and_Internet/Programming_and_Development/Languages/Java/](#)

Right there on one page is a wealth of promising links to documentation, reference, tutorials, news, downloads, articles, etc., and to the most lucrative resource of all, the **metaguide**. In this case, take a look at Java Boutique, which is a collection of useful Java information, news, forums, and more collected in one convenient location.

UNCLASSIFIED//FOR OFFICIAL USE ONLY

FIND THE RIGHT IBM RATIONAL ANALYSIS DESIGN AND CONSTRUCTION TOOLS FOR YOUR PROJECT

- Colorado Tech Online: IT Services, Education
- LookSmart Search Marketing & Online Advertising
- Managing Data Through Virtualization
- Generate Database-Driven Web Apps for .NET in Minutes
- Best Buy For Business. Find Database Software and more.
- Best Buy For Business. Find Computer Security Hardware.
- Web Design Services: Aplus.Net
- Best Buy For Business. Find Software Applications and more.

Thanks to thousands of individuals, corporations, and organizations, the Internet offers countless such metaguide sites on a huge variety of subjects. Which brings us directly to...

Rule Two

Let other people do as much work for you as possible (use their metaguides, their FAQs, their expertise to your advantage).

Directories are not the only good sources of general information. A number of **virtual libraries** and **reference desks** have sprung up on the web and they tend to be terrific starting places for all types of general information because they have thousands of pre-selected links to sources of data the researchers know to be good.

Let's continue with the Java example. If we go to the **Intute Science, Engineering, and Technology** page (formerly EEVL, the Internet Guide to Engineering,

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Mathematics and Computing) and search on *java*, we get back a list of highly relevant and carefully evaluated websites:

The screenshot shows the Intute search results page for the term 'java'. The page is titled 'Intute : science, engineering & technology' and includes navigation links for Home, Contacts, Help desk, and Site map. The search results are displayed in a table format with columns for 'Search term(s)', 'Results', and 'Harvester results'. The search term is 'java' and the results are 773. The harvester results are 4068. The page also includes a sidebar with various subject links and a 'Records marked: 0' indicator.

| Search term(s) | Results | Harvester results |
|------------------------|---------|-----------------------|
| java [subject(s): All] | 773 | 4068 (initials this?) |

Now showing: 1 - 25 of 773 records
Page: 1 2 3 4 5 ... 31
Order by: Relevance | Alphabetical

Petroleum Systems of the Northwest Java Province, Java and Offshore Southeast Sumatra, Indonesia
Details

Focus on Java
Details

The site includes a weekly article on Java, as well as links of various sorts for JUGs, JDKs, and Usenet groups. There is a news section updated daily, tutorials, documentation and source code on all aspects of Java scripting and programming. The site also includes free tools and utilities as well as links to Java resource sites on the Web such as source code banks, web directories and online publications.

In addition to the obvious SUN sites about Java, there are many others, such as links to Java FAQs, news, tutorials, course notes, seminar slides, articles, development tools, users' groups, mailing lists, books, conferences, links to web-based courses, and other resources.

Now you have a new resource for future Java-related research. Naturally, the first thing to do is bookmark the page.

Rule Three

Bookmark constantly, organize your bookmarks, and back them up as though your life depends on it.

One of the biggest and most influential entries into the reference/research world on the Internet is **Wikipedia**, a self-described free encyclopedia that anyone can edit. Because of its growth and importance, Wikipedia has earned a separate section in this year's edition. According to the Wikipedia, the term "wiki" describes "a group of

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Web pages that allows users to add content, as on an Internet forum, but also allows others (often completely unrestricted) to edit the content. The term *wiki* also refers to the collaborative software (wiki engine) used to create such a website (see *wiki software*). In essence, the wiki is a vast simplification of the process of creating HTML pages, and thus is a very effective way to exchange information through collaborative effort. **Wiki** is sometimes interpreted as the acronym for 'what I know, is', which describes the knowledge contribution, storage and exchange up to some point.²⁸ The most obvious potential problem with an encyclopedia that "anyone can edit" is quality control, and in fact, one of the Wikipedia's co-founders admitted serious problems with the quality and accuracy of some (perhaps a lot) of the Wikipedia content.²⁹ While there is a tremendous amount of good information in Wikipedia, it should not be relied upon as a sole source. Neither should it be ignored as this example of a "disambiguation" page on "java" shows:

[Create account](#) [log in](#)
[article](#) [discussion](#) [edit this page](#) [history](#)
Java
 From Wikipedia, the free encyclopedia.
 The term **Java** can refer to:
In geography:

- Java (island), the most populous island in Indonesia
 - Javanese language, a language widely spoken on the island of Java
 - Java coffee, a variety of coffee plant which originated on the island of Java, or a slang word for coffee
- Java Trench, a subduction zone trench off of the island of Java
- Java, Georgia (Republic of Georgia)
- Java, New York (United States of America)
- Java, South Dakota (United States of America)

In computer science:

- **Java** is a technology developed by Sun Microsystems for machine independent software, which encompasses:
 - Java programming language, an object-oriented high-level programming language
 - Java virtual machine, the virtual machine that runs Java byte code
 - Java platform, the Java virtual machine plus associated libraries
 - JavaScript, a scripting language (unrelated to the Java programming language).

Java may also mean:

Wikipedia also has the advantage of offering a free encyclopedia in a number of languages besides English, including French, Polish, Portuguese, Spanish, Dutch, Swedish, Italian, German, and Japanese.

²⁸ "Wiki," *Wikipedia*, Wikipedia, 2005. Answers.com <<http://www.answers.com/topic/wiki>> (14 November 2006).

²⁹ Andrew Orlowski, "Wikipedia Founder Admits to Serious Quality Problems," *The Register*, 18 October 2005, <http://www.theregister.co.uk/2005/10/18/wikipedia_quality_problem/> (14 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

To review, the best starting places for general information on broad topics are web directories/subject guides, virtual libraries, and reference desks. There are hundreds of such websites, but I've selected a few of the best.

| | |
|---------------------------------------|---|
| About | http://www.about.com/ |
| Encyclopedia.com | http://www.encyclopedia.com/ |
| Encyclopedia Britannica ³⁰ | http://www.britannica.com/ |
| Hotsheet | http://www.hotsheet.com/ |
| INFOMINE | http://infomine.ucr.edu/ |
| Information Please | http://www.infoplease.com/ |
| Internet Library for Librarians | http://www.itcompany.com/inforetriever/index.htm |
| Intute (formerly RDN) | http://www.intute.ac.uk/ |
| The Internet Public Library | http://www.ipl.org/ |
| Librarians' Index to the Internet | http://lii.org/ |
| The Library Spot | http://www.libraryspot.com/ |
| Martindale's The Reference Desk | http://www.martindalecenter.com/ |
| My Virtual Reference Desk | http://www.refdesk.com/ |
| Pinakes Subject Gateway ³¹ | http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html |
| Wikipedia | http://en.wikipedia.org/ |
| WWW Virtual Library | http://vlib.org/Overview.html |
| Yahoo Reference | http://education.yahoo.com/reference/ |



Web Tip

Think of search engine databases as *huge warehouses in which everything from diamonds to debris is stored*. Your job is to find the jewels amid the muck.

³⁰ Although full-text articles require a paid subscription to *Encyclopedia Britannica*, the site is still a useful starting place for research and includes free access to the *Britannica Concise Encyclopedia*.

³¹ Pinakes is the gateway to [Intute](http://www.intute.ac.uk/) and dozens of other equally valuable specialized research sites.

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

Search Savvy—Mastering the Art of Search

While directories and virtual libraries contain information selected by people, search engine databases are mostly unfiltered, that is, no human being is looking at the data being indexed to determine its value, authenticity, and reliability. Search engines are where the researcher's experience, knowledge, judgment, and intuition really come into play. Because of their vast scope and size, search engines are the heart and soul of Internet search and research. No other resource reaches as far or wide or quickly as a search engine. A researcher must learn to use search engines to their fullest extent despite their limitations.

Individual search engines have some very important advantages over directories, metasearch, and megasearch sites. Foremost among these is the fact that they have much larger databases of indexed sites. However, ***no single search engine is best***. Each has its own advantages and drawbacks. Furthermore, ***there is a remarkable lack of overlap among search engines databases***, so it is vital that you train yourself to use more than one search engine.

Greg Notess ran an interesting little experiment that demonstrated the need to use more than one search engine.³² He was looking for the real name for an AOL screen name, a piece of information that is often hard to find. One only search engine—in his example, Yahoo, found the name—while Google, Live, Gigablast, Ask, and Exalead all failed to locate the information. It could have been any search engine, not just Yahoo, that provided the data, but the point is clear: you must try multiple search engines, especially when looking for obscure or hard to find information.

On a larger scale, the metasearch engine Dogpile touted the results of a 2005 study they did in collaboration with researchers from the University of Pittsburgh and Pennsylvania State University showing a lack of duplication in the top results of the major search engines.

“When the researchers ran 12,570 different queries through search engines at Yahoo, Google, MSN and Ask Jeeves, they found that only 1.1 percent of the results appeared on all four engines, while 84.9 percent of the top results were

³² Greg Notess, “Overlap Showdown: Only 1 of 6,” Search Engine Showdown, 28 December 2006, <[http://www.searchengineshowdown.com/blog/2006/12/overlap showdown only at 1 of 1.shtml](http://www.searchengineshowdown.com/blog/2006/12/overlap%20showdown%20only%20at%201%20of%201.shtml)> (16 January 2007).

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

unique to one engine. Only 2.6 percent of the results were shared by three search providers, and 11.4 percent were delivered by two search engines.”³³

I am not surprised by the results, although I doubt the lack of overlap is quite as significant as the study indicates. The researchers used a relatively small sample, they only looked at the top ten results, they included paid results, and—probably most significantly—Dogpile sponsored the study. If a study to test metasearch engine results compared to individual search engine results concludes that metasearch engines do a better job, it is hardly surprising (and not necessarily convincing).

You can read the Dogpile/University of Pittsburgh and Penn State University study and take a look at Dogpile’s “Compare Search Engines” page to see how the comparison works.

Dogpile’s Compare Search Engines <http://comparesearchengines.dogpile.com/>

“Different Engines, Different Results”

<http://comparesearchengines.dogpile.com/OverlapAnalysis.pdf>

Rule Four

Use more than one search engine.

All search engines have their own way of doing things, which means there is no set of rules or guidelines that users can apply to them all. It helps, however, to familiarize yourself with the kinds of features available so you will at least know what to look for.

Often research involves the **search for specific information**: a telephone number, a name or title, a specific company or product, a piece of equipment, etc. Even researching a general subject may require a broader data set than a virtual library or subject guide offers if you need to find out as much as possible about a subject. For example, if I need to know everything available on the web about Mexico and NAFTA, I cannot limit myself to someone else’s edited list. Besides, there won’t be

³³ Dogpile.com in collaboration with researchers from the University of Pittsburgh and Pennsylvania State University, “Different Engines, Different Results, A Research Study,” July 2005, <<http://comparesearchengines.dogpile.com/OverlapAnalysis.pdf>> (14 November 2006). I have serious doubts about the accuracy of this claim, but the general conclusion of the study that there is a lack of overlap among search engine results is valid, if exaggerated.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

much, if any information because I am looking for specialized information (Mexico) within a big topic (NAFTA).

Compared to directories and metasearch services, individual search engines offer much greater flexibility and many more options for searching, not the least of which is the ability to search using *boolean expressions*. Search engine companies have concluded (probably rightly) that boolean searches are beyond the ken of most users, although you may find the boolean queries permitted by the best search engines are inferior to what you've used before.

One of the hottest areas of contention surrounding search engines has always been and continues to be *search engine index size*. I recommend you take size claims with a grain of salt. Search engine index sizes are self-reported and not validated by any objective third party. This old contest came to a head in 2005. First Yahoo claimed to have indexed over 20 billion "items" in its index. These items included "just over 19.2 billion web documents, 1.6 billion images, and over 50 million audio and video files."³⁴ Yahoo's claim at first appeared to mark the beginning of another competition to retain the "honor" of having the biggest search engine database, something Google had prided itself on for years. This time, however, instead of fighting back with bigger number counts on its homepage, Google dropped those numbers entirely as part of its seventh birthday celebration in September 2005. At the same time, Google announced a "newly expanded web search index that is 1,000 times the size of our original index...which makes Google more than 3 times larger than any other search engine."³⁵ Google did not offer any specific number but insisted it offers the most comprehensive collection of websites and documents on the Internet. Yahoo makes a similar claim. The answer? There is no one "best" search engine or site; researchers need a good toolkit of many resources when looking for rare information.

Determining search engine database size is something more akin to alchemy than arithmetic, so I suggest you take all such estimates of size with a large dose of skepticism. Besides, numbers are one thing and good search results are quite another. What good do 20 billion web documents do if not one of them provides the results you are seeking? **Relevant results** are the best measure of a search engine's value, but from my experience, having a larger pool in which to fish for these answers really does make it more likely that a search engine will retrieve the results users seek in the case of obscure information, which is after all the kind of information we are often seeking. Search engine size wars are almost always a good thing for researchers because it keeps the big players on their toes and

³⁴ "Our Blog is Growing Up—And So Has *[sic]* Our Index," *Yahoo! Search Blog*, 8 August 2005, <<http://www.ysearchblog.com/archives/000172.html>> (15 November 2006).

³⁵ "We Wanted Something Special for Our Birthday," *Google Blogspot*, 26 September 2005, <<http://googleblog.blogspot.com/2005/09/we-wanted-something-special-for-our.html>> (15 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

motivates them to improve their services. This past year's competition was no exception.

Another important fact to remember is that most search engines do not index entire websites or documents. It is no longer clear exactly how much of a webpage the major search engines index. For example, Google used to only index approximately the first 100KB of HTML, and reportedly the first megabyte of PDF documents, but in October 2005, Google dramatically increased the size of its cache limit. Yahoo indexes at least the first 500KB of HTML and PDF documents. As for Microsoft files types, my experimentation with them indicates that, in most cases, Yahoo indexes virtually the entire file, even in the case of very large documents.

The following is an overview of the major search engines in terms of their features, how to use them effectively, and what makes each one distinctive. It is important to remember there is no such thing as a perfect search engine. Each one has its advantages and drawbacks. The only way to fully exploit a search engine is to take the time to learn to use it, which means you must read the instructions.

Rule Five

Read the instructions.

Google

Google first gained fame and widespread use because of its single-minded focus on search, exemplified by its “clean” interface, and its **PageRank™** “weighted link popularity.” In simple terms, Google gives each webpage a rank based on the number of other pages linking to it and the “importance” of those pages, where importance is derived from an overall link count. While PageRank is imperfect, it works better than most other approaches to ranking search results and, indeed, is one of the primary reasons for Google’s success.

Some of Google’s features that helped to create this very successful and powerful search tool are:

- **cached versions** of webpages; Google was the first search engine to offer this option, which let users peek into its vast database.
- **automatic conversion of non-HTML filetypes** to HTML is available; Google was not the first to do this, but certainly has been the most successful.
- **backlinks** (the link: syntax); unfortunately, Google now limits the number of backlinks it shows, greatly reducing the utility of this option.
- Google seems to have **increased its limits on the size of indexed pages**. I found an indexed PDF document over 764K, a text file over 1000K, and a webpage over 366K. Very few webpages are larger than 500K. Google does not offer HTML versions of very large PDF or Word documents, e.g., the complete 9/11 Commission Report, but exactly what their cut-off size is, I do not know.
- Google **refreshes its index** continuously, not on a schedule (this is a good thing); Google’s Matt Cutts explains Google’s refresh rate: “It’s true that when an event happens on the web, our index can often pick it up in 1-2 days, and usually even faster. But a typical page in Google’s main web index is updated every 2-3 weeks or faster; it’s not the case that the entire main web index is updated every 2-3 days.”³⁶
- Google stopped advertising the **size of its database** in 2005, but Google is one of the largest if not the largest search database.

In determining the overall size of its index, Google also *includes urls of pages that it has not crawled and for which it has not indexed the text*. These “orphan”

³⁶ Matt Cutts, “Google Update Speed,” Google Blogoscoped, 26 July 2006, <<http://blog.outercourt.com/archive/2006-07-26.html#n28>> (14 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

pages may be any number of things, including pages with robots.txt command or tag. Unindexed pages are identifiable by what they lack: no summary, no page size, and no cached copy.

www.stat.vcu.edu/robots.txt

[Similar pages](#)

javangelist.snipsnap.org/space/SnipSnap/config/rob...

[Similar pages](#)

www.atmos.washington.edu/robots.txt

[Similar pages](#)

[fichier indiquant aux robots les endroits interdits ## voir http ...](#) - [[Translate this page](#)]

fichier indiquant aux robots les endroits interdits ## voir

<http://info.webcrawler.com/mak/projects/robots/norobots.html> User-agent: * Disallow: ...

www.ann.jussieu.fr/robots.txt - 1k - [Cached](#) - [Similar pages](#)

www.pamarys.ku.lt/robots.txt

[Similar pages](#)

sagittarius.student.utwente.nl/robots.txt

[Similar pages](#)

Google Orphans—no cached copy, no summary, no page size

Indexed page—cached copy, summary, page size

Google no longer displays the number of pages searched on its home page, but a search on **[the]** returns an estimated 4.8 billion pages, so that represents the *minimum* number of pages in the Google database (in fact, it probably is far larger). Remember, all size claims are “self validated,” so take them with a grain of salt. Still, for the types of Internet research we perform, bigger really is better because we have a much better chance of finding obscure information in billions of webpages than in millions.

Customizing Google Preferences

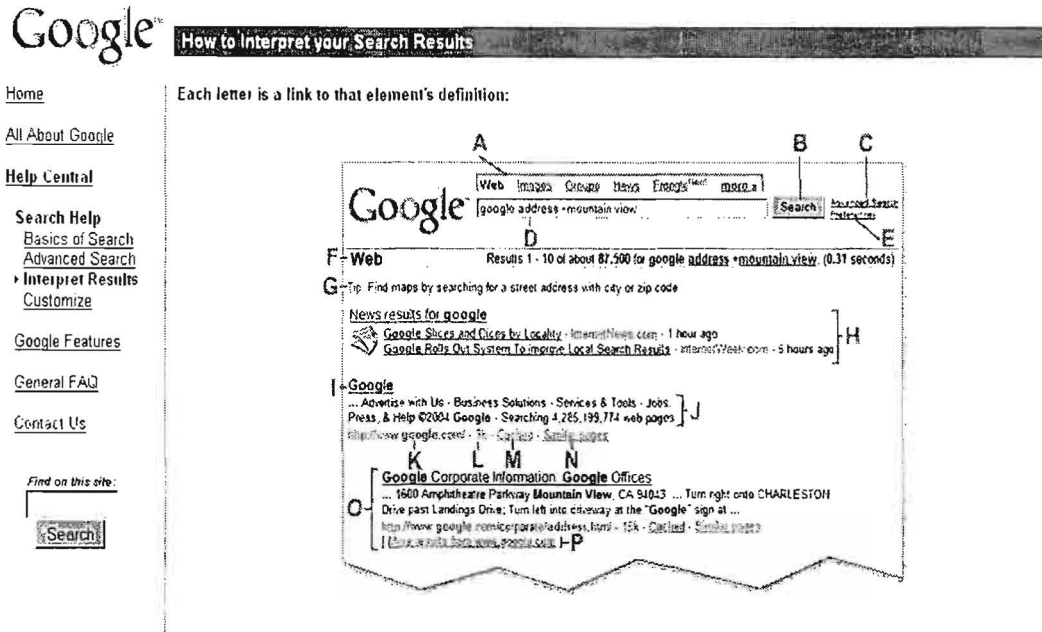
Google offers five basic **Preferences** settings:

1. Interface language: if you are more comfortable working in another language, Google can display in dozens.
2. Search language: generally, most searchers choose to search in any language, but there are occasions when it makes sense to limit your search to one or more specific languages. Google supports 35 languages, including non-Latin languages such as Arabic, Simplified and Traditional Chinese, Greek, Hebrew, Japanese, Korean, Russian, and Turkish.
3. Filtering of pages containing explicit sexual content.
4. Number of results: a purely individual preference, but 10 results per page is simply frustrating; Google lets you see up to 100 results per page.
5. Results window: opens results in a new browser window.

UNCLASSIFIED//FOR OFFICIAL USE ONLY

The Google Results Page

Once you've entered your search terms and selected the *Google Search* button, Google will present you with a list of results (hits). For each result returned you may see:



- the **statistics bar (F)** describes the type of search, e.g., web, and shows the number of results returned as well as the amount of time it took to complete your search
- a **tip (G)** may appear here, but not all queries generate tips. A typical tip might be "Try removing quotes from your search to get more results."
- "**OneBox Results**" (H) typically include news, stock quotes, maps, weather and local websites related to your search
- the **title (I)** of the webpage found
- an **excerpt (J)** from the webpage with your search terms **bolded**
- the **url (K)** of the webpage
- you will see **Supplemental Result** only if Google retrieved the result from its supplemental and not its main index
- the **size (L)** of the text portion of the webpage (omitted if page is not yet indexed)
- a **cached (M)** link to the version of the site stored by Google if it is indexed
- a **similar pages (N)** link for pages related to this result

UNCLASSIFIED//FOR OFFICIAL USE ONLY

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

- **indented results (O)** indicate Google has found more than one result from the same website; the most relevant page is listed first
- **a more results from (P)** link if there are more than two results from the same site

Google Basic Search

<http://www.google.com/>

Google assumes as its default that multiple search terms are joined by the **AND** operator, so that a search on the keywords [windows explorer]³⁷ will find all the webpages that contain **both** search terms. Furthermore, Google will first try to find all the webpages that contain the **phrase** ["windows explorer"]. Google will search:

- first, for **phrases** (keywords as one long phrase)
- second, for webpages containing all the keywords with the **greatest adjacency** (closest together),
- third, for webpages containing **all the keywords**, regardless of where they appear on the webpage

Google **will not return any results** if there is no webpage containing all the search terms. Try this query to see what I mean:

[kong spektioneer synecdoche]

There is an exception to this rule. Google often returns results when a keyword is not actually on a webpage but is in a **link to a website**, usually as text in a link anchor.

³⁷ Matt Cutts, one of Google's software engineers who also writes a blog mostly about Google, let his readers in on a little bit of Google insider information. "At Google, we use [and] to mark the beginning and end of queries. So ["scorpio submarine"] means to do a phrase search for "scorpio submarine", while [scorpio submarine] means just to type in those words without the quotes—you leave the brackets out when you actually do the search." That's an interesting and useful bit of trivia. I have tried various schemes to distinguish queries and ended up using italics (not a very elegant solution). From now on, I will be using brackets to set off queries in UTW. *Matt Cutt's Blog*, 11 August 2005, <<http://www.mattcutts.com/blog/writing-google-queries/>> (14 November 2006).



UNCLASSIFIED//FOR OFFICIAL USE ONLY

This is Google's cache of <http://www.athensams.net/myathens/> as retrieved on May 3, 2005 02:17:43 GMT. Google's cache is the snapshot that we took of the page as we crawled the web. The page may have changed since that time. [Click here for the current page](#) without highlighting. This cached page may reference images which are no longer available. [Click here for the cached text](#) only. To link to or bookmark this page, use the following url:
<http://www.google.com/search?q=cache:5mmyPb4Tq00J:www.athensams.net/myathens/+username+login+&22click+here&22shl=en&20target=&>

Google is not affiliated with the authors of this page nor responsible for its content.

These search terms have been highlighted: **username login**
 These terms only appear in links pointing to this page: [click here](#)


eduserv athens Contact us | Eduserv Athens | Help

 MyAthens 

Tuesday, 3 May 2005

To access your organisation's electronic resources, log in to MyAthens with your Athens **username** and password. Please note that the password is case sensitive.

Username:
 Password:
 » Login

 This is a secure login.

Forgotten your Athens account details? Don't have an Athens account?
 Check the [Athens organisation list](#) to contact your Athens administrator or to find out if your organisation uses Athens. If you are not a member of one of the organisations listed, then you are not entitled to use Athens.

Google *limits the number of search terms to 32 keywords*. It ignores any terms beyond that number. However, there are ways to force Google to search for more than 32 keywords.

Google is *not case sensitive*. There does not appear to be anything you can do about this.

In late 2003 Google introduced automatic *word stemming* or *truncation*, i.e., searching for variations of search terms. Normally, word stemming involves searching for plurals and verb conjugations such as *drink, drank, drunk*. However, Google's word stemming is not consistent and somewhat confusing. For example, stemming does not work either with single word or phrase searches, i.e., a search on [child] will not find *children*. Yet a search on [child health] will find *child, childhood, children, and children's*. Google will also find some variations of verbs, e.g., a search on [drink water] will find *drinking water*. Users should still search on all variations of a term, including plurals. There is a Google hack for disabling word stemming.

Google automatically *clusters search results*. Multiple hits from the same site are indented and there is usually an option to see more results from a specific site.

Google *permits the use of the OR operator* in simple search. The OR must be capitalized.

UNCLASSIFIED//FOR OFFICIAL USE ONLY

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

Beyond the use of the OR operator in its simple search, *Google does not support boolean search*.

While Google *assumes that multiple keywords are a phrase*, searchers can delimit phrases using double-quotes. For example, if I search on:

[the last king of france]

without double-quotes, Google will ignore the “the” and the “of” in its search. The results I get include many irrelevant hits, such as music from a group called “The Last King” and an article about Lance Armstrong. However, if I enclose the same query in *double-quotes*, Google will search on exactly the phrase [“the last king of france”], and return a result with the name of the last king of France. Enclosing searches in double-quotes is much more effective for finding precise results than relying on automatic phrase searching.

Google no longer routinely ignores *stop words* outside double quotes. Each of these searches will now return different results:

[the last king of france] [last king france] [“the last king of france”]

Stop words are English words that are so commonplace they are not included in a search unless the searcher forces Google to do so. The stop words Google recognizes include: *a, an, about, and, are, as, at, be, by, com, from, how, I, in, is, it, of, on, or, that, the, this, to, we, what, when, where, which, with, why*. There probably are others!

However, Google’s handling of stop words is inconsistent. For example, in the query [to be or not to be], Google ignores OR because it may be a logical operator, and it also appears to ignore TO and BE, only searching for NOT. Therefore, you may need to force Google to search for a stop word on occasion. There is a Google hack for forcing Google to search for stop words.

It is unnecessary to use the plus sign (+) with any terms except stop words because by default Google searches for all keywords. However, there are many times when searchers need to exclude certain terms that are commonly associated with a keyword but irrelevant to their search. That’s where the *minus sign (-)* comes in. Using the minus sign in front of a keyword ensures that Google excludes that term from the search. For example, the results for the search [“pearl harbor” –movie] are very different from the results for [“pearl harbor”].

Google’s handling of words with diacritical marks such as accents or umlauts is inconsistent. *By default, Google will search for terms matching those with and without the diacritic*. As Google’s Vanessa Fox explains, “When a searcher enters a query that includes a word with accented characters, our algorithms consider web pages that contain versions of that word both with and without the accent. For

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

instance, if a searcher enters [México], we'll return results for pages about both "Mexico" and "México."³⁸

For example, a search on [façade] will return pages containing both facade and façade. To force Google to search only for the term with the diacritic, put a **plus sign in front of the term: [+façade]**. You may see a few pages that do not appear to have the diacritic, but that is probably because that term appears in anchor text or an inbound link that is pointing to the page but not actually on the page in question.

However, Fox goes on to explain that results also vary depending upon whether you are searching at Google.com or a Google international site (e.g., Google.fr), whether your preferred language at Google is English or another language, and from where you are coming to the Google site as indicated by your IP address. If Google detects that your IP address geolocates to Peru, your search results will be different from those provided to someone coming to Google from Norway, regardless of the preferred language or the site you search. Also, users who have registered with Google and set up personalized search will find that their results are affected by their previous searches. In other words, while there are ways to *manipulate* the results Google provides, there is no way to *control* them.

Google treats most **punctuation marks** the same way, as links in a search string. For example, Google handles a search for [c-span], [c.span], ["c span"], and [c?span] basically the same way. However, a search for [cspan] with no space or mark is treated differently.

³⁸Vanessa Fox, "How search results may differ based on accented characters and interface languages," Official Google Blog, 32 August 31 2006, <<http://googlewebmastercentral.blogspot.com/2006/08/how-search-results-may-differ-based-on.html>> (November 27, 2006).

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Google will search for several punctuation marks and special characters:

- the *ampersand* [**&**]: Google will search for [barnes&noble] or [barnes & noble]
- the *underscore* [**_**]: Google will search for a phrase such as [public_records.doc] or even more specifically [public_records.html]
- the dollar sign [**\$**], used with a number, for example [\$100]
- the sharp [**#**], for example [F#]
- the slash, but only when used in the search [I/O]
- While Google will not actually search on a plus sign, the search engine does recognize the difference between searches for [c], [c+], and [c++]

Google Advanced Search

Google has a number of “query modifiers” to restrict searches and make them more effective in many cases. These query modifiers can be used in simple search in the following syntax or on the advanced web search page using the appropriate menu options. The query modifiers Google supports are:

- **site:** restricts results to websites in a given domain. *This syntax no longer requires you to add a keyword.* Google’s site: syntax will also **search within folders**, e.g., [site:jpl.nasa.gov/technology]. Remember you *can* add keywords to the site: search. [site:jpl.nasa.gov/technology “jpl spacecraft”]

Advanced Web Search > Domains

Examples of how to use the **site:** command:

[shuttle site:www.nasa.gov] finds pages about the space shuttle at the NASA website.

[site:info] finds all the pages in the Google database in the .info top-level domain


[“bulletin officiel” site:fr] finds pages in the French top-level domain about official bulletins

[cirrus -site:mastercard.com] finds pages about the keyword cirrus that are not at the Mastercard.com site

[site:jpl.nasa.gov/technology]

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

[site:jpl.nasa.gov/technology "jpl spacecraft"]



[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [Desktop](#) [more »](#)
 site:jpl.nasa.gov/technology "jpl spacecraft" [Advanced Search](#)
[Preferences](#)

Web Results 1 - 3 of 3 from jpl.nasa.gov/technology for "jpl spacecraft"

Tip. Try removing quotes from your search to get more results.

CASE STUDIES OF TECHNOLOGY IN THE DSN
 Mariner 10 was the first **JPL spacecraft** to transmit full resolution images in real time from planetary distances.; to photograph Venus.; to encounter and ...
deepspace.jpl.nasa.gov/technology/95_20/case.htm - 30k - [Cached](#) - [Similar pages](#)

[PDF] The Evolution of Technology in the Deep Space Network: A History ...
 File Format: PDF/Adobe Acrobat - [View as HTML](#)
 Mariner 10 was the first **JPL spacecraft**. (1) To transmit full-resolution images in real time from planetary distances. (2) To photograph Venus. ...
deepspace.jpl.nasa.gov/technology/95_20/95-20.pdf - [Similar pages](#)

Technology - Images & Videos
JPL spacecraft have visited all known planets except Pluto (a Pluto mission is currently under study for the early part of the next decade).
www2.jpl.nasa.gov/technology/images_videos/rv_pages/P37079.html - 17k - [Cached](#) - [Similar pages](#)

- **intitle:** restricts the results to documents containing the keyword in the title.

Advanced Web Search > Occurrences

Examples of how to use the **intitle:** command:

[intitle:amazon] finds all pages that include the word *amazon* in their title

[intitle:amazon "rain forest"] finds all pages that include the word *amazon* in their title and mention the phrase "*rain forest*" anywhere in the document (title or text or anywhere in the document)

- **allintitle:** restricts the results to documents containing all the keywords in the title of the document.

Advanced Web Search > Occurrences

Example of how to use the **allintitle:** command:

[allintitle:amazon jungle "rain forest"] finds pages that include all the words in the title (not the text) of the document, e.g.,

```
<title>Amazon Rain Forest Jungle Tours</title>
```

- **inurl:** restricts the results to documents containing the keyword in the url.

Advanced Web Search > Occurrences

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Examples of how to use the **inurl:** command:

[inurl:nasa] finds all pages that include *nasa* anywhere in the url (address)

[inurl:nasa -site:gov] finds all pages that include *nasa* anywhere in the url of sites that are *not* in the .gov top-level domain

[inurl:nasa shuttle] finds all pages that include *nasa* anywhere in the url of the site and *shuttle* anywhere in the document (url or anywhere else).

- **allinurl:** restricts the results to documents containing all the keywords in the url.

Advanced Web Search > Occurrences

Example of how to use the **allinurl:** command:

[allinurl:nasa shuttle] finds all pages that include *both* nasa and shuttle in the url of the site.

- **link:** restricts the results to documents that have links to a specific webpage. ³⁹
Cannot use with keyword search terms.

Advanced Web Search > Page Specific Search > Links

Example of how to use the **link:** command:

[link:www.noaa.gov] finds all pages linking to the NOAA homepage.

[link:www.noaa.gov/wx.html] finds all pages linking to a specific page at the NOAA site.

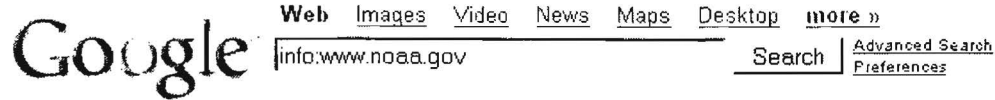
- **info:** presents information Google has about a webpage. This option is only available via the main Google search.

Example of how to use the **info:** command:

[info:www.noaa.gov] provides links to Google's cache of the page, pages that are similar to *www.noaa.gov*, pages that link to *www.noaa.gov*, and pages that contain the term *www.noaa.gov*.

³⁹ The Google *link:* command no longer shows all links as it once did in order to cut down on the amount of webspam created by hidden links on webpages. Therefore, the Google *link:* command is not nearly as useful as it used to be.

UNCLASSIFIED//FOR OFFICIAL USE ONLY



Web

NOAA Home Page

an agency of the US Department of Commerce. Conducts environmental research.
www.noaa.gov/

Google can show you the following information for this URL:

- Show Google's cache of www.noaa.gov
- Find web pages that are similar to www.noaa.gov
- Find web pages that link to www.noaa.gov
- Find web pages from the site www.noaa.gov
- Find web pages that contain the term "www.noaa.gov"

- **related:** restricts the results to documents Google has determined are similar to a specific webpage.

Advanced Web Search > Page Specific Search > Similar

Example of how to use the **related:** command:

[related:www.nasa.gov] finds other US government homepages.

- **cache:** presents the version of the webpage Google has stored. This option can also be accessed by clicking on the Cached link on the main results page.

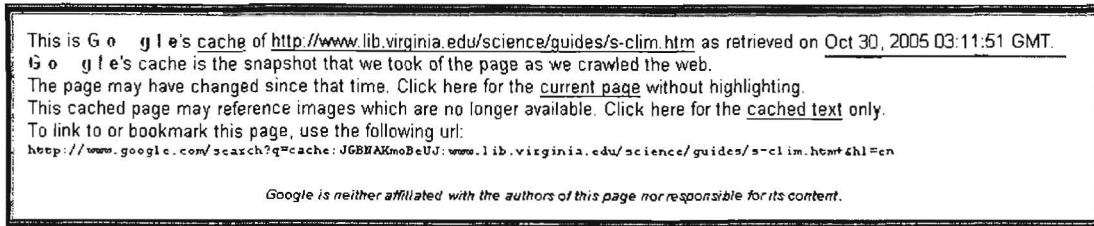
Examples of how to use the **cache:** command:

[cache:www.noaa.gov] shows the stored version of the NASA homepage

[cache:www.noaa.gov hurricane] shows the stored version of the NOAA homepage with the keyword *hurricane* highlighted.

With no fanfare, Google once again began showing the date and time when a webpage was cached. This is not a new feature in Google; the Google cache option showed date/time until mid-2000 when date and time unceremoniously disappeared. Now it's back. Who knows why? Who cares? It's a good thing. Here's what you'll see:

UNCLASSIFIED//FOR OFFICIAL USE ONLY

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

Important: there is a Google hack that lets you view the cached text only version without first viewing the cached page containing images and other non-text data that could send information back to the original website. Gigablast also offers a "stripped" cache option.

- **filetype**: Google will search the content of many file types and must be used with keyword(s). However, there is a Google hack that lets you get around the keyword requirement. **Microsoft filetypes are potentially dangerous** to open in their native formats. Please follow these instructions for handling Microsoft files on the Internet safely.

Warning: use Google option to "view as html" instead of opening certain file types (mainly Microsoft Word and Excel) that could contain macro viruses.

Google will search the content of these file types:

HTML

Corel WordPerfect (wp)

Lotus 1-2-3 (wk1, wk2, wk3, wk4, wk5, wki, wks, wku)

Lotus WordPro (lwp)

MacWrite (mw)

Microsoft Excel (xls)

Microsoft PowerPoint (ppt)

Microsoft Word (doc)

Microsoft Works (wks, wps, wdb)

Microsoft Write (wri)

Portable Document Format (pdf)

Postscript (ps)

Rich Text Format (rtf)

Text (ans, txt)

Macromedia Shockwave Flash (swf)

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Example of a filetype: search:

[filetype:doc bulletin] will find MSWord documents containing the keyword *bulletin*

For details on how to use the Google filetype option, please refer to the Google Filetype help FAQ. Also, there are a number of undocumented file type searches available using Google.

Google Filetype FAQ

http://www.google.com/help/faq_filetypes.html

Google Special Search Features

Spell Checker: Google has a very good spell check option. When you input a query, Google checks to see if you are using the most common spelling of the keyword. If not, Google nicely asks, "Did you mean: x?" where x is the most common spelling. I really love this because Google doesn't presume. Sometimes you are intentionally misspelling a term. The classic example is [http referrer]. This computer term is almost always misspelled, so searching on [http referrer] won't yield nearly as many results as searching on the misspelled term. Google's dictionary also includes *proper names*.

Calculator: the Google calculator will evaluate basic and complicated mathematical expressions as well as convert units of measurement and physical constants. Soopie makes the Google calculator extremely easy to use. For detailed help in using the calculator, see:

Google Calculator Help

<http://www.google.com/help/calculator.html>

Dictionary: Google has integrated dictionary definitions into its search options. Nothing could be easier to use. Underlined keywords appearing at the top of the results page are linked to Answers.com.

Define: new to Google is the *define* option. To use it, enter *define* then a word or phrase. This feature augments the dictionary option by searching a wider variety of sources. For example, the query [define blog] will return a web definition as the first result. *The advantage of the define option is that the definition appears at the top of*

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

- **FCC equipment Ids:** syntax is [fcc EJM386S303]

Weather: The Google weather search is for US locations only. Simply enter the keyword *weather* followed by a city and/or state or even just a zip code (which works just fine by itself) and Google will present you with an attractive, succinct weather chart:

Google Web Images Groups News Froogle Local^{Hot} Desktop more »
 weather 94101 Search Advanced Search Preferences

Web Results 1 - 100 of about 555 for weather 94101. (0.35 seconds)

Weather for San Francisco, CA

54°F
 Mostly Cloudy
 Wind: SW at 12 mph
 Humidity: 82%

| Wed | Thu | Fri | Sat |
|-----------|-----------|-----------|-----------|
| | | | |
| 63° 49° | 66° 49° | 65° 51° | 64° 50° |

Sponsored Links
 US Local Weather Service
 Current weather reports & forecasts for your area with realtime updates
www.localweather.com

Airport Delays and Weather: To see delays and weather conditions at a US airport, enter the airport's three-letter code and the word *airport*. For example, to see delays and weather conditions at Baltimore-Washington International, enter [bwi airport]. At the top of the Google results page you will see the image of an airplane followed by a link to "View conditions at Baltimore-Washington International Airport (BWI), Baltimore, Maryland." The link takes you directly to the FAA's Air Traffic Control System Command Center's real-time status information page for BWI.

Phonebook, Street Maps, and Stock Quotes: US residential and business phonebook lookups, US addresses, and US stock exchange data. Please see Google Help for information on using these features:

Google Help <http://www.google.com/help/features.html>

Google Guides: Did you know that Google publishes "a variety of reviewer's guides to selected Google products on the Google Press Center" designed for journalists who are reviewing these products? However, the guides are very well done and include a lot of useful set-up and user instructions.

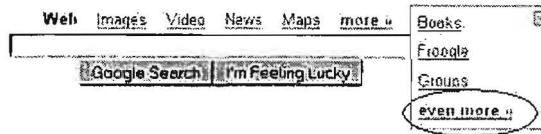
Google Guides <http://www.google.com/press/guides.html>

Google Services

Google has many services hidden behind its spare homepage. By selecting "more" and then "even more" on the homepage...

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//FOR OFFICIAL USE ONLY



[Advertising Programs](#) · [Business Solutions](#) · [About Google](#)

You will see the many products and services Google offers that are not reflected on the “googlized” homepage.

- [Froogle](#)
Shop for items to buy online and at local stores
- [Images](#)
Search for images on the web
- [Local](#)
Find local businesses and get directions
- [Maps](#)
View maps and get directions
- [News](#) - now with [archive search](#)^{NEW}
Search thousands of news stories
- [Scholar](#)
Search scholarly papers
- [Specialized Searches](#)
Search within specific topics
- [Toolbar](#)
Add a search box to your browser
- [Video](#)
Search TV programs and videos
- [Web Search](#)
Search over billions of web pages
- [Web Search Features](#)
Find movies, music, stocks, books, and more

Search

- [Alerts](#)
Get email updates on the topics of your choice
- [Blog Search](#)
Find blogs on your favorite topics
- [Book Search](#)
Search the full text of books
- [Catalogs](#)
Search and browse mail-order catalogs
- [Checkout](#)
Complete online purchases more quickly and securely
- [Desktop](#)
Search and personalize your computer
- [Directory](#)
Browse the web by topic
- [Earth](#)
Explore the world from your PC
- [Finance](#)
Business info, news, and interactive charts

- [Picasa](#)
Find, edit and share your photos
- [SketchUp](#)
Create 3D models for Google Earth
- [Talk](#)
IM and call your friends through your computer
- [Translate](#)
View web pages in other languages

Go mobile

- [Maps for mobile](#)
View maps and get directions on your phone
- [Mobile](#)
Use Google on your mobile phone
- [SMS](#)
Use text messaging for quick info

Make your computer work better

- [Pack](#)
A free collection of essential software
- [Web Accelerator](#)
Speed up the web

Explore and innovate

- [Code](#)
Download APIs and open source code
- [Co-op](#) - now with [Custom Search Engine](#)^{NEW}
Contribute your expertise and customize the web search experience
- [Labs](#)
Explore Google's technology playground

Communicate, show & share

- [Blogger](#)
Share your life online with a blog -- it's fast, easy, and free
- [Calendar](#)
Organize your schedule and share events with friends
- [Docs & Spreadsheets](#)
Create and share your projects online and access them from anywhere
- [Email](#)
Fast, searchable email with less spam
- [Groups](#)
Create mailing lists and discussion groups

News: Google News headlines are entirely generated using a computer algorithm that scours more than 4500 worldwide news sources. Google News also offers *international editions* for France, Germany, India, Italy, Spain, and several other countries. For details on Google News, see the [news search engine](#) section below.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Images: Google Images indexes more than a billion images in JPEG, GIF, and PNG (Portable Network Graphics) formats. The Advanced Image Search lets users limit searches by filetype, size of image, coloration, and site or domain.

Google Image Search

<http://images.google.com/>

Book Search: To use the book search, simply preface any search with the keyword *book* or *books*. The first three results, if there are any matches, will come from [Google Book Search](#).

Groups: Google acquired Deja, the last remaining Usenet newsgroup search engine, in 2001. Even before that, Google began its own Usenet archive in August 2000. The complete Usenet archive, more than a billion messages dating back to 1995, is searchable via Google Groups. In 2004 Google introduced a new version of Google Groups that includes a mailing list and discussion forum creation option to rival Yahoo Groups' similar service. Also new is the ability for registered users to keep track of their favorite topics using the star (☆) feature. By clicking the star next to a favorite topic, that topic is added to the user's "My starred topics" page. Postings now appear in minutes in Google Groups rather than the hours it used to take.

Google Groups still offers both a simple and advanced interface to search the newsgroup postings. Both interfaces are extremely easy to use. Google Groups not only returns results (sorted by relevance or date), it also shows you the most relevant groups for your topic. So a search on the term *oceanography* suggests I might want to take a look at the related groups [sci.geo.oceanography](#) and [bionet.biology.deepsea](#). Newsgroup searching in general and [Google Groups](#) are discussed in greater detail in a later section.

Google Groups

<http://groups.google.com/>

Mobile SMS Search: This service is different from the SMS text messaging that has been available at AOL, Yahoo, Live Search, etc., for some time. The new Google SMS permits queries using mobile technology. Google's SMS service offers similar services but with different shortcuts. It is open to all US subscribers using a "major" US cell phone provider and also to most UK mobile subscribers. The US number is 46645 (GOOGL on most phones) and for the UK it is 64664 (6GOOG on most phones). Google explains how to use the SMS search service and offers a number of sample queries at its new SMS webpage.

Google SMS

<http://www.google.com/sms/>

Patent Search (beta): New for 2007, Google Patent Search now has its own discrete page. Users could always search for US patents by number, but Google decided to create a separate page for these searches. The new site offers many advanced search options, including options to search by patent number, title, inventor's name, assignee's name, US and international classifications, and issue or filing date range. Even more valuable than the search options are the view choices.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

The new Google patent search has the ability to show you the patent itself, complete with zoom, page scrolling, drawings, internal search, and other patents that reference the current one. Here is the drawing of H. K. Markey (aka the actress Hedy Lamarr) and George Antheil's 1941 patent for a "secret communication system" that used a new concept: frequency hopping.

Google | inventor:Hedy | inventor:Kiasler | inventor: | Search Patents | Sign in

SECRET COMMUNICATION SYSTEM Hedy Kiasler Markey

Page 1 Full screen

Aug. 11, 1942. H. K. MARKEY ET AL 2,292,387
SECRET COMMUNICATION SYSTEM
Filed June 10, 1941 2 Sheets-Sheet 1

Fig. 1. 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000.

Fig. 2. 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000.

Patent summary
Patent number: 2292387
Filing date: Jun 10, 1941
Issue date: Aug 11, 1942
More about this patent

Patent sections
Abstract
Drawing
Description
Claims

Search within this patent
Go

Google Patent Search

<http://www.google.com/patents>

Blog Search: Google Blog Search is a direct competitor with Technorati, until now the big dog on the blog search block. Despite all the chatter about them, blogs are still kind of mysterious and confusing in part because there are lots of things called blogs that don't fit the earlier definition. Blogs originally referred to on-line personal journals, often updated daily, but now everything using RSS or Atom (XML formats for distributing newsfeeds) seems to be considered a blog. Therefore, blog search engines generally restrict themselves to indexing and searching for anything that uses a site feed. Google Blog Search is no exception. The new Google Blog Search FAQ says, "The goal of Blog Search is to *include every blog that publishes a site feed* (either RSS or Atom)." [emphasis added] This means that Google Blog Search defines "blog" as any site with an XML site feed, and that is fine as long as we know what we're getting. However, Google Blog Search is apparently excluding feeds from news sources to try to prevent overlap between Google News and Blog Search.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Keep in mind that **Google Blog Search only indexes the site feed, not the full content at the website** that originated the feed.

Google Blog Search indexes feeds dating back to January 2000. Also, one of the big advantages of the XML format is that it, unlike HTML, includes date/time data, which means you can use Google Blog Search to find information from a specific day or a range of dates. Google Blog Search will also enable users to search entire blogs or specific posts.

Some of the Google Search operators work in Google Blog Search and it has its own unique operators, too, as the About Google Blog Search page explains:

All of the standard Google Search operators are supported in Blog Search. These include:

- link: [very useful in finding who's linking to whom]
- site:
- intitle:

Additionally, Blog Search supports the following new operators of its own:

- inblogtitle:
- inposttitle:
- inpostauthor:
- blogurl:

For example, a search such as [mandolin inpostauthor:Graham] will show you posts about mandolins written by people named Graham. Note that you can also use the Advanced Search option to achieve the same effect.

In addition, you can restrict your results to any one or any combination of 35 languages using the Advanced Search option. Google Blog Search will also give users the option of subscribing to the blogs in the news aggregator of your choice.

The main drawback of Google Blog Search seems to be that it indexes only the content of *feeds* and often what is syndicated in a newsfeed is very sparse. Technorati wins hands down on this point because it does do full text searching. Google may eventually decide it needs to do so as well. After all, this is a beta version.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

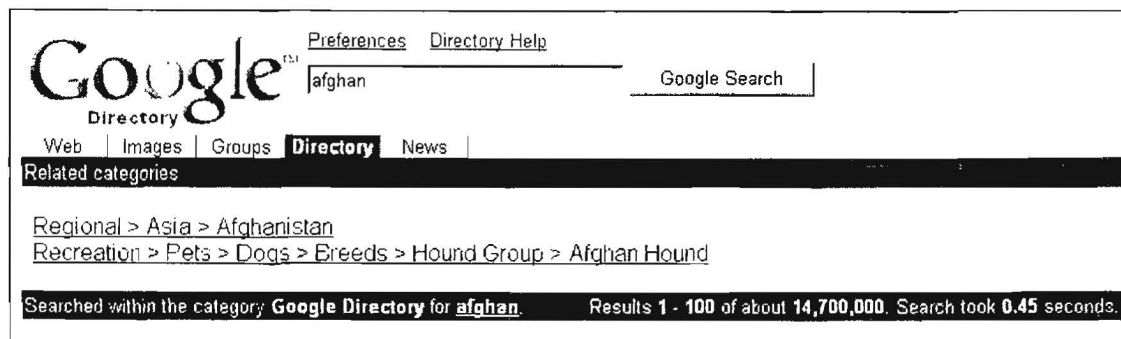
Google Blog Search

<http://blogsearch.google.com/>

About Google Blog Search

http://www.google.com/help/about_blogsearch.html

Directory: Google's web directory uses the [Open Directory Project's](#) collection and its own search technology to rank the sites based upon "importance" (which usually means popularity). Google Directory lets users limit searches to a specific directory category. For example, if I search the directory for the keyword [afghan], Google presents two categories:



By selecting the *Regional > Asia > Afghanistan* category, I have the option to search only in this category, thus focusing my search and avoiding irrelevant results. The Google Directory contains over 1.5 million urls.

Google Directory

<http://directory.google.com/>

Video Search: As of the first of the year 2007, Google Video began to include results from [YouTube](#), which it purchased in October 2006. For now, when users click on the YouTube results, they are taken to the YouTube website. YouTube videos do not appear on the Google video homepage, only in search results where they are recognizable from their address.

Google Video Search is a way to **search and view** TV shows, including news, entertainment, and more, for free. The search includes not only national networks, such as ABC, CBS, NBC, and CNN, but also local programming and shows from around the world. How does this work? Google Video indexes the closed captioning of TV shows so that when you search for a keyword, it finds that word in the captioning transcripts and displays a list of the shows with that keyword. In most cases, you can only see still shots of the show, but in a few instances, you will be able to view the entire broadcast. This is a very new Google option, and it is sure to expand and improve over time.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

One of the best features of Google Video is that you do not need any special software to view the videos, only Macromedia Flash Player, which is a free browser plug-in that most users already have on their computers.

Google Video offers users the opportunity to rank video using a system of one to five stars for users to rate videos as well as the ability to add labels (tags to describe a video) and comments.

The screenshot shows a Google Video search result for the keyword "iran". The search bar contains "iran" and the search button is labeled "Search". Below the search bar are navigation links: "Top 100", "Comedy", "Music videos", "Movies", "Sports", "Animation", "TV shows", and "Google Picks". The main video player displays a scene with two men in white clothing, one of whom is holding a rifle. The video title is "One hour video about Iran", with a duration of 58 min 23 sec and a date of Apr 19, 2005. The video has 16,265 views and a 3.8 rating. Below the video player are controls for "Add label", "Download" (for Windows/Mac), and "This is a one hour video all about Iran". To the right of the video player is a "Playlist" section with four video thumbnails and titles: "One hour video about Iran" (58 min), "Let's Bomb Iran!" (4hrs Productions, 3 min), "Women Police in Iran" (1 min), and "Bomb Iran unknown" (3 min). The video player interface includes a play/pause button, a progress bar, and a volume control icon.

There are some of the operators that work in Google Video:

title: enter the name of a TV show on one of the stations Google Video indexes, with or without keywords. For example:

[title:nature] or [title:newshour robotics]

genre: enter news, comedy, music, animation with or without a keyword; I recommend referring to the complete list of genres on the Advanced Search page.

type: enter sports, music_video, movie with or without a keyword.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

duration: short (< 4 min.), medium (4-20 min.), or long (> 20 min.) with or without a keyword.

is:free or **is:for sale** determines whether or not the search finds free or for purchase videos.

language: limits the search to videos in a specific language using the same digraphs as main Google search.

Google Video

<http://video.google.com/>

Google Scholar

In November 2004, Google introduced a new tool called Google Scholar. Here's how the Google site describes it: "Google Scholar enables you to search specifically for scholarly literature, including peer-reviewed papers, theses, books, preprints, abstracts and technical reports from all broad areas of research. Use Google Scholar to find articles from a wide variety of academic publishers, professional societies, preprint repositories and universities, as well as scholarly articles available across the web." Moreover, "Google Scholar...automatically analyzes and extracts citations and presents them as separate results, even if the documents they refer to are not online. This means your search results may include citations of older works and seminal articles that appear only in books or other offline publications." Google Scholar not only indexes journal articles, dissertations, and technical reports, it also indexes books, which means you can use Google's new Library Search (OCLC's WorldCat search) to locate the book in a local library or find a place to purchase the book online.

Although a number of scholarly search sites and tools already exist—e.g., CiteSeer, DOAJ, ArXiv, and even Google's own partnership with IEEE—the fact that the premier search engine has branched off into scholarly search is obviously significant. Google Scholar searches across a far wider range of sources than any other publicly available scholarly search tool currently available. Users should be able to read at least an abstract of articles that require registration and access the full text if they or their institution have a subscription for the content. The best thing about Google Scholar is that it gives users the range, power, and flexibility of Google. As far as I can tell, all the types of Google syntax—*site:*, *inurl:*, *filetype:*, etc.—work with Google Scholar. You can limit your search to file type using either the *filetype:* or *ext:* syntax, e.g., [*ext:pdf*] (*filetype:* and *ext:* work interchangeably). The most useful addition to Google Scholar is probably the new *author:* syntax (which, by the way, already existed in Google Groups search).

As you can see from this query, Google Scholar searches and retrieves scholarly references from many types of sources and also provides a handy "Cited by" link that shows all the pages referring to the original work.

UNCLASSIFIED//FOR OFFICIAL USE ONLY



Scholar Results 1 - 14 of 14 for author:candy "artificial intelligence". (0.09 seconds)

[BOOK] **Artificial Intelligence: A Modern Approach** - [Library Search](#) - [Web Search](#)
S Russell, P Norvig, JF **Candy**, JM Malik, DD ... - [Cited by 1956](#)
Englewood Clis, NJ: Prentice Hall, 1995

[CITATION] Creative design of the Lotus bicycle: implications for knowledge support systems research - [Web Search](#)
L **Candy**, E Edmonds - [Cited by 20](#)
Design Studies, 1996

[Support for collaborative design: Agents and emergence](#)
EA Edmonds, L **Candy**, R Jones, B Soufi - [Cited by 38](#)
... 10 Jones, RM and Edmonds, EA A framework for negotiation. In CSCW and
Artificial intelligence, J. Connolly and EA Edmonds, Eds. ...
Communications of the ACM, 1994 - [portal.acm.org](#) - [portal.acm.org](#)

[Issues in the Design of Expert Systems for Business](#)
E Edmonds, L **Candy**, P Slatter, S Lunn - [Cited by 1](#)
... INDEX TERMS Primary Classification: I. Computing Methodologies I.2 ARTIFICIAL
INTELLIGENCE I.2.1 Applications and Expert Systems Subjects: Office automation. ...
Expert Systems: Human Issues - [portal.acm.org](#)

[PDF] [Introducing creativity to cognition](#)
L **Candy**, E Edmonds - [Cited by 3](#)
... fortunate to welcome. To anyone who has even heard of **Artificial Intelligence**,
Marvin Minsky needs no introduction. In his forthcoming ...
Proceedings of the third conference on Creativity & ..., 1999 - [portal.acm.org](#) - [portal.acm.org](#)

Google Scholar also offers an **advanced search option**. It certainly simplifies searching for articles by author, articles published in a specific publication, and words in the articles' title. However, as with most date searches, forget it. I searched for articles about chemistry published in the year 2020 and found three. Either Google knows something about the future that we don't or their software is misreading some number as a year. The advanced Google Scholar search also let users limit their search by publication. This is somewhat misleading because the "publication" can be a citation, article, or book, although there is no way to tell Google Scholar to distinguish among these choices. Also, the publication searches are imperfect; a search limited to the publication *Nature* also returns results from *Nature Medicine*, for example.

During 2006, Google Scholar added a new feature that "will make it easier for researchers to keep up with recent research...It's not just a plain sort by date, but rather we try to rank recent papers the way researchers do, by looking at the prominence of the author's and journal's previous papers, how many citations it already has, when it was written, and so on. Look for the new link on the upper right for 'Recent articles'—or switch to 'All articles' for the full list."⁴⁰

⁴⁰ Dejan Perkovic, "Keeping up with recent research," Google Blogspot, 20 April 2006, <<http://googleblog.blogspot.com/2006/04/keeping-up-with-recent-research.html>> (31 October 2006).

UNCLASSIFIED//FOR OFFICIAL USE ONLY

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Also new for 2006 was a related search option: "For every Google Scholar search result, we try to automatically determine which articles in our repository are most closely related to it. You can see a list of these articles by clicking the 'Related Articles' link that appears next to each result. The list of related articles is ranked primarily by how similar these articles are to the original result, but also takes into account the relevance of each paper."⁴¹

Péter Jacsó has called Google Scholar's quality into question in his excellent and thorough analysis of Google Scholar's citation ability. Jacsó, Professor of Library and Information Science, University of Hawaii, concluded that "Google Scholar (GS) does a really horrible job matching cited and citing references."⁴² There are numerous other scholarly citation search options (CiteSeer, ISI Highly Cited, and Scirus) that, for now at least, are superior to Google Scholar.

However, I would not count Google Scholar out in the long run. Google Scholar is yet another example of what are called "vertical search engines," that is, search services that focus on indexing and searching specialized data sources. Vertical search has fundamentally replaced the portal concept as a more targeted, less manpower-intensive, and more cost effective means of getting the right information to the right people at the right time.

Google Scholar

<http://scholar.google.com/>

Advanced Google Scholar Search

http://scholar.google.com/advanced_scholar_search

Google Trends

Google unveiled Google Trends in May 2006 and set a lot of people thinking about its potential utility. Google Trends is a new technology that lets users see how many searches have been performed on one to five terms and where those searches originate.

"Google Trends analyzes a portion of Google web searches to compute how many searches have been done for the terms you enter relative to the total number of searches done on Google over time. We then show you a graph with the results—our search-volume graph—plotted on a linear scale.

⁴¹ Luiz Barroso, Distinguished Engineer, "Exploring the scholarly neighborhood," Google Blogspot, 22 August 2006, <<http://googleblog.blogspot.com/2006/08/exploring-scholarly-neighborhood.html>> (10 October 2006).

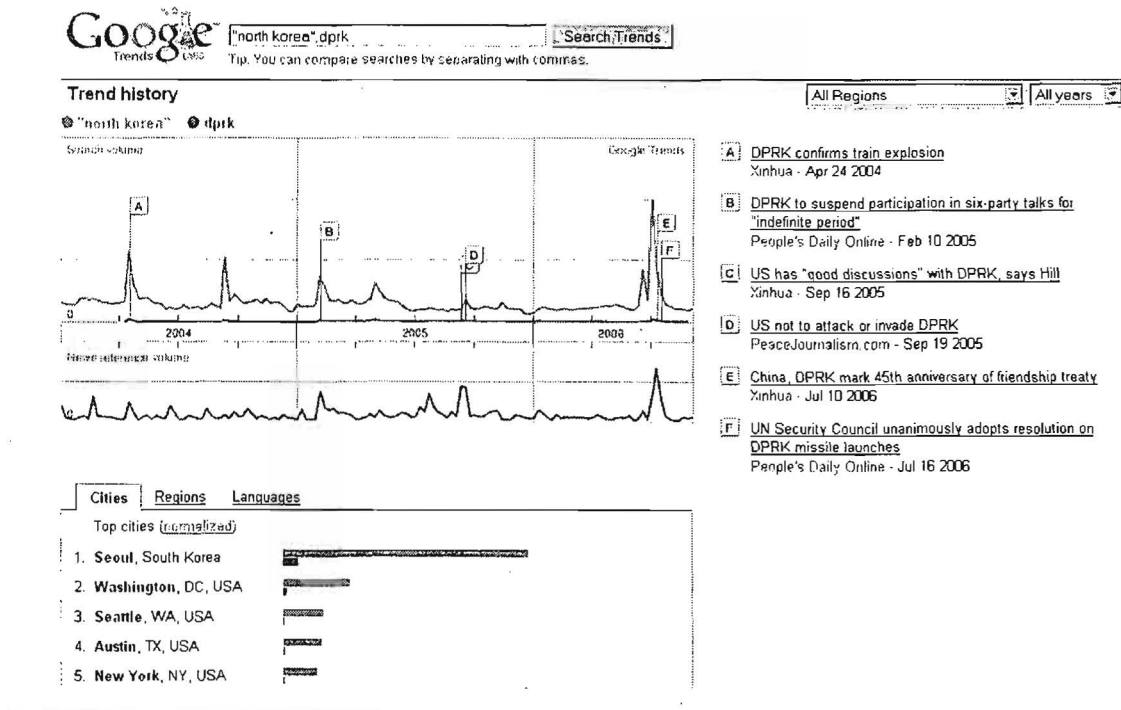
⁴² Péter Jacsó, "Google Scholar and *The Scientist*," Péter Jacsó's Review Extras, October 2005, <<http://www2.hawaii.edu/~jacso/extra/qs/>> (31 October 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Located just beneath our search-volume graph is our news-reference-volume graph. This graph shows you the number of times your topic appeared in Google News stories. When Google Trends detects a spike in the volume of news stories for a particular term; it labels the graph and displays the headline of an automatically selected Google News story written near the time of that spike. Currently, only English-language headlines are displayed, but we hope to support non-English headlines in the future. Below the search and news volume graphs, Google Trends displays the top cities, regions, and languages for the first term you entered.⁴³

There are some very important limitations to Google Trends, however. First of all, the feedback provided by Google Trends is based on *a portion of Google's searches*, not all of them. Google Trends seeks to provide "insights into broad search patterns," not detailed and verifiable data about searches. Second, "as a Google Labs product, it is still in the early stages of development," meaning it is prone to error because "several approximations are used when computing your results," but Google does not say what these are.⁴⁴

Here's a look at Google Trends' results for the query comparing search terms ["north korea",dprk] for all regions and all years. Note that in Google Trends, you can *compare terms by using a comma to separate them*.



⁴³ "About Google Trends," Google Trends, 2006, <<http://www.google.com/intl/en/trends/about.html>> (31 October 2006).

⁴⁴ Google Trends.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Despite all its limitations, I am intrigued by the potential this tool offers. For example, if I were looking back over the past couple of years for a very obscure term, I would definitely use Google Trends to see if I could discern anything useful from this data.

Google Trends

<http://www.google.com/trends>

Google Guides

Did you know that Google publishes "a variety of reviewer's guides to selected Google products on the Google Press Center" designed for journalists who are reviewing these products? However, the guides are very well done and include a lot of useful set-up and user instructions.

Google Guides

<http://www.google.com/press/guides.html>

Web Tip

Everything isn't on the Internet (or it's not free)!

Contrary to popular opinion, everything is not on the Internet. In fact, much of the kind of information you are used to working with is not and never will be on the Internet. Unrealistic expectations about the kinds of information you may find on the Internet can lead to frustration and wasted time and effort. A general rule of thumb: the more sensitive, rare, or expensive the information, the less likely it is to be on the Internet. Also, much valuable data on the Internet requires payment.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Google Hacks

Google hacks—a term usually associated with the book of the same name published by O'Reilly Publishing—are tips, tricks, techniques, and scripts that make Google more powerful and useful. Some are extraordinarily simple, such as being conscious of word order, while others are either so complicated (“scraping the code”) or trivial (Googlisms) that I doubt many of you will ever need them.

Because there is so much interest in Google hacks, I thought I would catalog links to the best sites on the web for finding more information about Google hacks as well as bring the best Google hack techniques together in one place.

First, a word about **Google APIs**, which are used to create many Google hacks. API stands for Application Programming Interface. Google offers its own free APIs developers' kit, which provides documentation and example code for using the Google Web APIs service.⁴⁵ That's fine if you are in a situation that permits downloading, installing, and running code from the web, but that is not always an option. However, many Google hacks either do not require an API key or, if they do, are available, thanks to the kindness of strangers, on websites.

Best Google Hack Websites

The Official (but not the best): Google hacks from O'Reilly. Taken directly from the book of the same name by Tara Calishain and Rael Dornfest, the complete list of 100 hacks is here, but only a few have details about how to use them.

Google Hacks <http://hacks.oreilly.com/pub/ht/2>

Researchbuzz.com: Google Hacks Archive. Much better source of Google hacks from Tara Calishain's website. Google hacks are listed by date and you can search the site.

Researchbuzz http://www.researchbuzz.org/archives/google_hacks.shtml

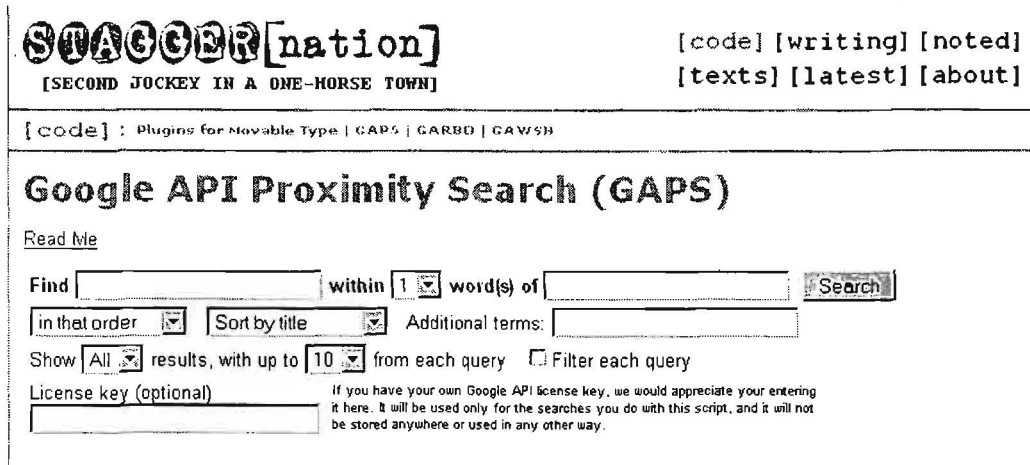
Google API FAQs. Everything you need to know about Google's API service. Remember, this process involves registration, downloading software, and other interaction with Google, so it's not for everyone.

Google API FAQs http://www.google.com/apis/api_faq.html

⁴⁵ "As of December 5, 2006, we are no longer issuing new API keys for the SOAP Search API. Developers with existing SOAP Search API keys will not be affected." Google SOAP Search API <<http://code.google.com/apis/soapsearch/>> (20 February 2007).

Staggernation: Three Google APIs.

- API Proximity Search (GAPS) "uses the Google API to search Google for two search terms that appear within a certain distance from each other on a page. It does this by using a seldom-discussed Google feature: within a quoted phrase, * can be used as a wildcard meaning 'any word.'" ***This is a very useful tool***; it gives users the option of searching for two terms within one, two, or three words of each other in any order or a specific order.
<http://www.staggernation.com/cgi-bin/gaps.cgi>



- API Web Search By Host (GAWSH) "uses the Google API to search Google for a query string, and returns a list of the web hosts found in the set of results. You can then expand any of these hosts and display only the results from that host...Clicking on the triangle to the left of a host will perform the same query again, but restricted to that host (using Google's "site:www.foo.com" query syntax), and expand the listing to display the first 10 results." I have to admit I find this particular script more confusing than useful. I prefer to use Google's **site:** syntax, but others may find this API to their liking.
<http://www.staggernation.com/gawsh/>
- API Relation Browsing Outliner (GARBO) uses the Google API to search Google for pages that are either related to (using the related: keyword) or linked to (using the link: keyword) that URL. ***Does not add a great deal to what Google can do already***, beyond offering the option to view the results as snippets or urls.
<http://www.staggernation.com/garbo/>


Google Rankings. This site contains a number of different options created as search engine optimization tools for website creators and maintainers. However, some of these Google tools are what you might call "dual use technologies."

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

- **Keyword Density.** Probably not something everyone needs everyday, but a pretty neat tool. It lets you enter a url, then see which are the words and phrases that address uses most in the form of a detailed report on their numbers and density. Although it was designed as a search engine optimization tool, I can see its utility as a rudimentary traffic analysis tool.

Google Rankings

<http://www.googlerankings.com/kdindex.php>


GOOGLE RANKINGS
Keyword Density
 Enter the domain or URL of your website to see which are the words and phrases it uses most, see a detailed report on their numbers and density
 URL of your website:

Ignore words of characters in length or less. | Minimum occurrences:

| | |
|---|---|
| Include the following: <input type="checkbox"/> Include Page Title <input type="checkbox"/> Include Meta tag Keywords <input type="checkbox"/> Include Meta tag Description | Word Filters <input type="checkbox"/> Use our default <u>stop word list</u> <input checked="" type="checkbox"/> Ignore the following word list <input type="checkbox"/> Include only the following word list Separate words by a space or put one word per line. |
|---|---|

- **Mass Keyword Search:** This tool lets you enter from two to ten keywords and a url to see where (if anywhere) the site ranks in Google *vis-à-vis* those keywords. The tool only searches up to the top 1000 Google sites. This could be useful to see if an unusual term is found at a specific site. <http://www.googlerankings.com/mkindex.php>
- **Mass Domain Search:** With this tool, you can enter up to 10 different urls (domains) for the same keyword and check the position of the websites in Google up to the top 1000 sites. This could be useful to see where an unusual term ranks in a number of different sites at once. <http://www.googlerankings.com/mdindex.php>

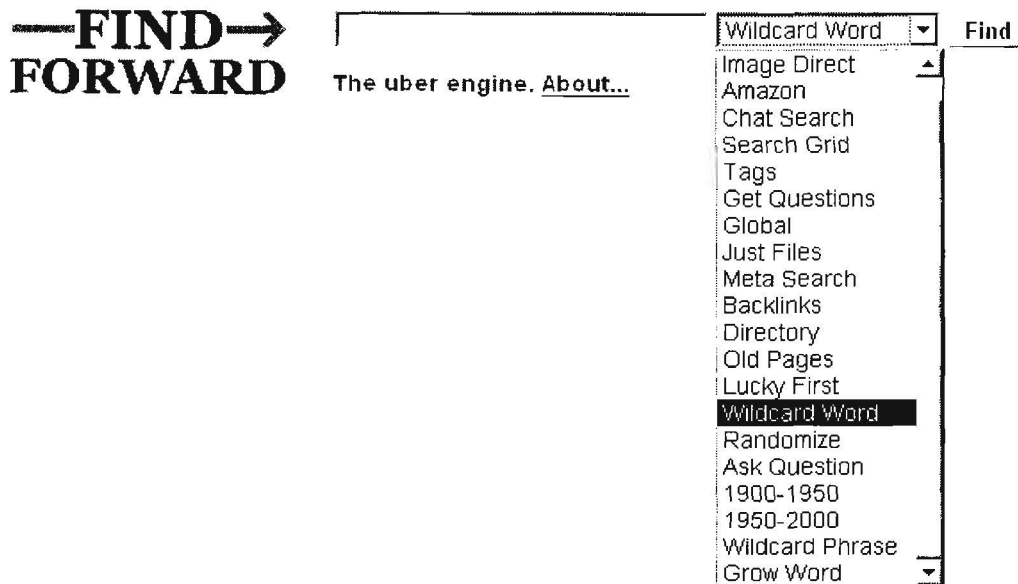
FindForward. If you love Google, you will probably find Philipp Lenssen's creation very useful. Lenssen has done Google one better by creating a simple way to power search Google with the help of the Google Web APIs. All users have to do is to enter

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

a query (FindForward supports all basic and advanced Google search options and syntax) and select the type of search desired from the pulldown list. As you can see from the list below, there are many possible types of searches available. Try the **Search Grid** option, in which you can enter up to five keywords to create a grid showing inter-relationships of the terms. Try [iran korea nuclear terror], for example. Then there is the potentially dangerous **Just Files** option for searching website directories that were (probably) not meant to be browsed.

Find Forward











<http://www.findforward.com/>

Fagan Finder's Google Ultimate Interface. A useful and friendly Google interface that allows users to maximize Google power without learning the syntax that it deserves to be on this list.

<http://www.faganfinder.com/google.html> (for Internet Explorer)

<http://www.faganfinder.com/google2.html> (for Mozilla browsers)

Soople. A Google interface that makes Google's advanced features so easy to use it's amazing. The main page offers lots of query boxes that enable users to run complex Google queries without knowing anything about Google. But I think the best Soople tool is the calculator page. Did you know Google has a very powerful calculator function as a basic part of the search engine? Most people don't, and even if they did, they wouldn't use it because it is not what I would call intuitive. Therefore, the Soople calculator interface to the Google calculator is a godsend for the mathematically impaired such as yours truly.

| | |
|--|--|
| SEARCH GOOGLE ABOUT SOOPLE | |
| Main Page -Calculator- Translate Phone & Location Superfilter | |
|  Smart calculator (explain) <input type="text"/> <input type="button" value="Calculate!"/> <small>Tip: Convert many measures/units Like ' 15 yards in meters '</small> |  Quick unit converter (explain) Number: <input type="text"/> <input type="button" value="Choose unit"/> in: <input type="text"/> <input type="button" value="Choose unit"/> <input type="button" value="Convert"/> |
|  Add, subtract, divide & multiply (explain) <input type="text"/> <input type="button" value="+"/> <input type="text"/> <input type="button" value="Calculate"/> |  Calculate the percentage of... (explain) <input type="text"/> percent of <input type="text"/> <input type="button" value="Calculate"/> |
|  Trigonometric functions (explain) <input type="button" value="Sinus"/> <input type="text"/> <input type="button" value="Calculate"/> <small>Note: numbers are assumed to be radians otherwise type 'degrees' after the number</small> |  # root of... (explain) <input type="button" value="Choose root"/> <input type="text"/> <input type="button" value="Calculate"/> |
|  ln (logarithm base e) (explain) <input type="text"/> <input type="button" value="Calculate"/> |  log (logarithm base 10) (explain) <input type="text"/> <input type="button" value="Calculate"/> |
|  Find the remainder after division (explain) Divide: <input type="text"/> by: <input type="text"/> <input type="button" value="Find the remainder"/> |  Choose: determine the number of ways of choosing (explain) <input type="text"/> choose <input type="text"/> <input type="button" value="Calculate"/> |

Soople Calculator

<http://www.soople.com/index.php?sub=calculator>

Compare Google Results From Different Countries. This is a valuable and much needed new tool. This site lets you run one query simultaneously against two different Google locations. What is more, it also lets you choose between two different Google datacenters. For example, if you select "www.google.com" as a datacenter, the query could go to a number of different access points. It is probably better to pick a specific Google datacenter. The site also lists many but not all local Google domains. If you select a local domain, the search and results' language will default to that country's language. However, if you specify a local Google domain, the tool may use any existing datacenter.

It is not really as hard to use this tool as I've made it sound. You can easily play around with this tool and see what works for you. Here's an example of a search for a Spanish term in the US using the generic "www.google.com" and the local Google and datacenter for Spain. The results are quite different.

Google Geotargeted World-Wide Search across Datacenters and Countries

[[Back to oy.oy.eu](#) | [More information](#) | [Discussion](#) | [Contact me](#)]

Query:

Location 1: Location 2:

Datcenter 1: Datcenter 2:

Google

Web Results 1 - 20 of about 31,700,000 for alvarez (0.30 seconds)

Alvarez - Wikipedia
Acoustic and electric guitars, Banjos and mandolins

Alvarez & Marsal
Alvarez & Marsal is one of the world's premier professional services firms specializing in financial restructuring, creditor advisory, crisis and interim...

A Science Column: People and Progression: Luis Alvarez
Luis Alvarez was a physicist with wide-ranging interests. Alvarez's colleagues sometimes called him the "pinto adductor" because of the huge range...

Luis Alvarez - Wikipedia, the free encyclopedia
His aunt was California artist and collector Mabel Alvarez. Associated bibliography for Luis Alvarez from the Flood Digital Library for Nuclear Issues...

Victor Alvarez - Wikipedia, the free encyclopedia
His grandfather is the famed physician Walter C. Alvarez and his grandmother... T. Rex and the Crater of Doom by Walter Alvarez ISBN 0-375-70210-5

Luis Alvarez - Biography
Luis Alvarez, Luis "Lu" Alvarez, was born in San Francisco, Calif., on June 13, 1911. Dr. Alvarez joined the Radiation Laboratory of the University of...

Google

Cerca: a tota la xarxa pàgines en català Espanya

Web Resultats 1 - 20 d'aproximadament 30.500.000 per a alvarez (0,19 seg)

a-alvarez.com - Home
Alvarez Primera empresa española en venta por Internet de caza, pesca, golf, náutica, equitación, buceo y esquí. Todos los productos al mejor precio.

Tarjeta V.P. - www.alvarez.com/.../php?pt=&aj=12

su País Inicio - www.alvarez.com/.../php?pt=&aj=10

Nuevos catálogos - www.alvarez.com/.../php?pt=&aj=11

Añadir a Favoritos - www.alvarez.com/.../php?pt=&aj=15

Alvarez Guitars
Acoustic and electric guitars, Banjos and mandolins

ISMAEL ALVAREZ.COM
www.ismaelalvarez.com/...

ISMAEL ALVAREZ COM * D'ARIO *
Mi hermano Pitingo en el telediaro de TVE! Aquí mi hermano Antonio Vélez EL PITINGO en el telediaro de television española...

Luis Alvarez - Home Page

Using Google to search for terms using non-Latin character and/or diacritical marks remains problematic. I recommend that you **put a plus sign in front of any non-English term** when using this tool. While the Google geographic comparison tool is imperfect, it is a useful and interesting way to see how Google handles queries differently depending upon geographic region.

Compare Google results for different geographic locations

<http://oy-oy.eu/google/world/>

Simply Google. On one page you can see and use pretty much every Google search, find most Google sites, link to all the Google blogs, and even locate some other search blogs and sites. This really is a useful site. Notice the live bookmarks for all the blogs. Google search syntax works fine here (it's just an interface to the "real thing").

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Google Sites

- [Google Analytics](#)
- [Google Accounts](#)
- [Google Alerts](#)
- [Google Mail](#)
- [Personalised Home](#)
- [Google Suggest](#)
- [Google Sets](#)
- [Google Moon](#)
- [Google Mars](#)
- [Google Store](#)
- [Google Current TV](#)
- [Google Zeitgeist](#)

Google Fun

- [Mentaflex](#)
- [Pigeon Rank](#)
- [Mocrbase Google](#)
- [Google Gulp](#)
- [Google Romance](#)

Downloads

- [Desktop](#)
- [Earth](#)
- [Talk](#)
- [Toolbar](#)
- [Video Player](#)
- [Picasa](#)
- [Pack](#)

Google Searches

| | | |
|-------------|---------------------------------|--------|
| Web | Search | |
| Images | Search | |
| Groups | Search | |
| News | Search | |
| Froogle | Search | |
| Local | Search | |
| Blogs | Search | |
| Books | Search | |
| Maps | Search | |
| Scholar | Search | |
| Video | Search | |
| Music | Search | |
| Feeds | Search | |
| Base | Search | |
| Directory | Search | |
| Finance | Search | |
| Catalogs | Search | |
| Answers | Search | |
| History | Search | |
| Bookmarks | Search | |
| Linux | Search | |
| Apple | Search | |
| Microsoft | Search | |
| Movies | movies | Search |
| Definitions | define | Search |
| Weather | weather | Search |
| bbc.co.uk | site:bbc.co.uk | Search |
| The Answer | the answer to life. the univers | Search |

Google Blogs

- [Official Google Blog](#)
- [Adwords API](#)
- [Blogger Buzz](#)
- [Google Base](#)
- [Google Code](#)
- [Google Enterprise](#)
- [Google Maps API](#)
- [Google Reader](#)
- [Google Research](#)
- [Google Talk](#)
- [Google Video](#)
- [Inside AdSense](#)
- [Inside AdWords](#)
- [Inside Desktop](#)
- [Inside Sitemap](#)

Other Blogs

- [Google Blogoscoped](#)
- [Search Engine Watch](#)
- [Inside Google](#)
- [MSN Search](#)
- [Yahoo Search](#)
- [John Battelle's Searchblog](#)

One thing that I found really interesting behind the concept of this page is that it has been “de-googled.” What does that mean? Well, “googlizing” a home page means stripping it down basically to one thing, in Google’s case, that one thing is search. The critics of googlization point out that while that worked great for Google, it rarely works most of the time for the simple reason that most websites are trying to do something more than just one thing. In fact, “the experience of using Google is not the experience of using any other site. People go to Google to search the entire web.”⁴⁶ People go to other sites for other reasons and to do other things, sometimes many other things. Furthermore, Google is not simple! It just hides its complexity behind that plain homepage.

Simply Google

http://www.usabilityviews.com/simply_google.htm

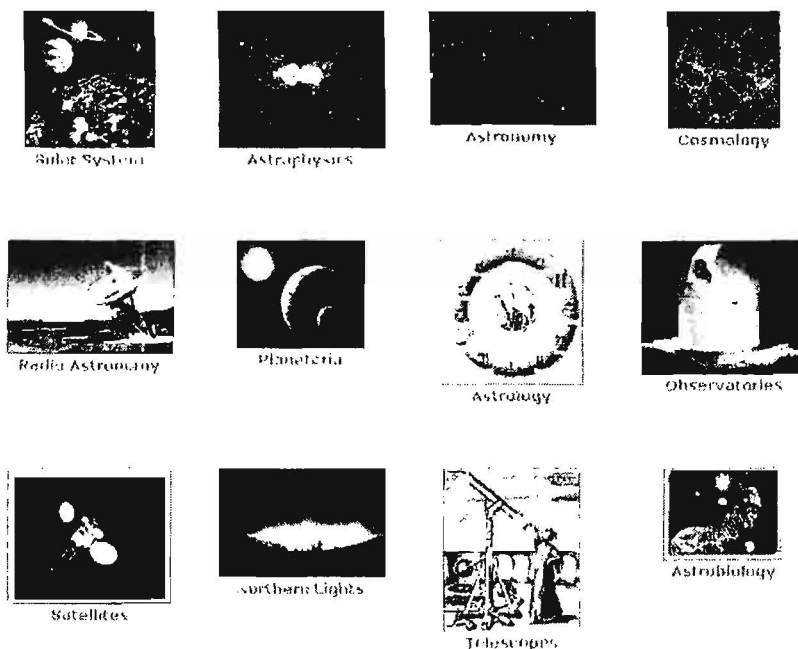
⁴⁶ Jared Spool, “Home Page Googlization,” User Interface Engineering Brainsparks.com, April 6, 2006, <<http://www.uie.com/brainsparks/2006/04/05/home-page-googlization/>> (30 October 2006).

Google Related Image Search. The ever-creative Philipp Lenssen has developed a new tool that looks for related images. It is a small PHP5 script that is "first screenscraping Google Sets, and then it's screenscraping Google Images (once per term found in the set, which is why this takes some seconds)." <<http://blog.outercourt.com/forum/68971.html>> I have to say it is mostly a novelty at this point, but the results are interesting and could potentially be useful. Here are the results for the query [solar system, milky way]:

Find Related Images

Term (like *bmw* or *brad pitt*):

Results for *solar system, milky way*:



Notice that you can search for up to two terms separated by a comma. That means that you can search for [solar system, milky way] but a search for three single terms, e.g., [mercury, pluto, mars] may return strange results at this time. In any event, this is an interesting and unique tool, so if you need to search for related images, this is the place to start.

Google Find Related Images

<http://blog.outercourt.com/related/>

The Best Individual Google Hacks

While not the only useful and interesting Google hacks, these are among the most valuable and less well-known techniques every serious searcher needs to know.

How Google Parses Queries. Google is sensitive to the number and order of query terms. If you enter the query [windows explorer] without quotation marks, Google will first try to find all the webpages that contain the phrase ["windows explorer"] Google will search:

- First, for keywords as one long phrase.
- Second, for webpages containing all the keywords with the *greatest adjacency* (closest together).
- Third, for webpages containing *all the keywords*, regardless of where they appear on the webpage.

Word Order Matters. Google gives more weight to the first term in a query, so ***put the most important search term(s) first.*** Try these two queries and you'll see how different the results are: [new york city] vs. [city york new]

Repetition, Repetition, Repetition. If you keep getting irrelevant hits, you can try repeating a keyword that will be emphasized by Google, such as [java coffee coffee coffee], which cuts down considerably on the number of results about the programming language.

Boilerplate Words or Phrases Yield Gold. Used in combination with keywords, standardized words or phrases can produce very useful results from Google. Whether it's "company proprietary," "not for distribution," or a copyright disclaimer, these are the kinds of identifying query terms that searchers need to look for. Tara Calishain sites the example of using "copyright * the new york times company" plus keywords to locate not only articles at *The New York Times* website but those reprinted elsewhere.

Disabling Word Stemming. The problem with Google's word stemming is that Google does not give users the option to turn it off, which can frustrate users trying to perform precise searches. However, ***if you put a plus (+) sign in front of a term, this will disable word stemming.***

Searching on Stop Words. There are two ways to force Google to search on stop words such as *the, a, an, I*. The first is to ***include stop words in phrases enclosed by double-quotes***, e.g., ["to be or not to be"]. The second way to force Google to search for stop words is to put a ***plus sign (+)*** in front of them, e.g., [+who +what +when].

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Google Wildcard. Google has tinkered with its wildcard and it is more useful but still woefully inadequate. The Google wildcard (*) can only be used to replace a term or terms in a query. It cannot be used to truncate search terms, e.g., [child*] to find children, childhood, etc., and it cannot be used to find alternate spellings of terms, e.g., [kazak*stan]. That being said, the Google wildcard is not useless. The wildcard search is very good at helping you do "fill in the blank" searches, and it works great not only in English but in other languages as well. Here are a few examples of "fill in the blank" searches using the wildcard (there is no need to capitalize because Google ignores uppercase):

Google [Web](#) [Images](#) [Groups](#) [News](#) [Finance](#) [Local](#) [more »](#)

[Advanced Search](#)
[Feedback](#)

Web Results 1 - 8 of about 11 for "Baltimore is the * city of Maryland" with SafeSearch on. (0.65 seconds)

Tip: Find maps by searching for a street address with city or zip code

[RIM: 2nd Annual Research Symposium: OVERVIEW & GENERAL INFORMATION](#)

Baltimore is the Charm City of Maryland While attending the conference, you will be staying in the heart of downtown, Baltimore's Inner Harbor. ...

[Maryland Mortgage Specialists.com - Real Estate - Baltimore](#)

The city of **Baltimore is the largest city in all of Maryland**. Baltimore is one of the leading cities in trade and commerce in Maryland as well. ...

[Travel - compare prices, reviews and buy at NexTag - Price - Review](#)

Comparison Shopping for Products, Mortgages, Travel, Cars & More.. All Categories, Magazines, Travel, Automotive, Baby, Books ...

[Maryland Vacation Rentals | Hotels](#)

Baltimore is by far the most exciting city of Maryland, it provides the Orioles and Ravens for spring entertainment and there is Harborplace which is ...

[MD Mortgage Home Loans - MD Real Estate - Baltimore](#)

The city of **Baltimore is the largest city in all of Maryland**. Baltimore is the leading city in trade and commerce in Maryland as well. ...

You can use more than one wildcard in a search, and the wildcard will find more than one missing term.

Google [Web](#) [Images](#) [Groups](#) [News](#) [Finance](#) [Local](#) [more »](#)

[Advanced Search](#)
[Feedback](#)

Web Results 1 - 100 of about 717 for "* é o presidente do Brasil". (0.28 seconds)

[User talk:Patrick.br - Wikimedia Commons - \[?\] - \[?\] - \[?\]](#)

Lula é o presidente do Brasil, Wade é o presidente do Senegal e Gorée é o nome do lugar em que eles se encontraram (ilha Gorée) ...

[planeta terra el amando - \[?\] - \[?\] - \[?\]](#)

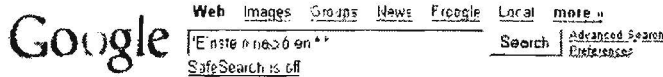
O Lula é o Presidente do Brasil, não só de brasileiros que comb eu votaram nela, mas de todos nós. Nós esperamos que ele se manifeste, que deixe de fazer ...

[Fórum dos Leitores - Estadão.com.br - \[?\] - \[?\] - \[?\]](#)

DEUS ajuda a quem quer ajudar Bernardo Lopes Lula presidente **LULA é o presidente do Brasil**. Eleito legitimamente pelo povo que quer mudanças. ...

The wildcard can be used at the beginning of a phrase.

UNCLASSIFIED//FOR OFFICIAL USE ONLY



Web Results 1 - 85 of about 250 for "Einstein nació en " (0.09 seconds)

In Search for English results only. You can specify your search language in Preferences

Yahoo! Groups clubhumanistalaino Messages: 1-100 of 2368 ...
Francis Albert Einstein nació en Ulm, Alemania, en March 14, 1879 , de padres J
humanist_a Off re Send Email, May 8, 2000 ...

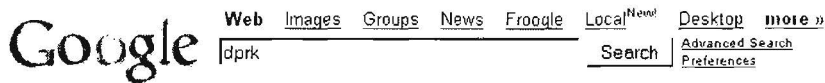
eListas.net - Mis elistas: virtual.mensa.es - [...]
Einstein nació en Ulm, Alemania, en 1879 y emigró a Estados Unidos en 1933 donde
falaciendo en Princeton (Nueva Jersey) en 1955 ...

Grupo Fórmula - Artículos - [...]
'El físico alemán -americano Albert Einstein, nació en Alemania, el 14 de marzo
de 1879 y murió en Princeton, NJ, el 18 de abril de 1955, contribuyó más ...

The wildcard can be used at the end of a phrase.

Stripping the Cached Copy: while Google offers the option to view the cached text version of a webpage from the full cached page, there is a Google hack that lets you view the cached text version only without having to open the cached page that contains images or other non-text data that might be sent back to the original website or that might redirect you to another page simply by adding &strip=1 onto the end of the url the cached page as follows:

- 1. Right-click on the Cached link [in red below] and select Copy link location in Mozilla or Copy Shortcut in Internet Explorer.
2. Paste that location (url) into your browser's address bar [do not hit return yet!].
3. Add &strip=1 to the end of the url and hit return.



Web Results 1 - 100 of about 1,720,000 for dprk with Safesearch on. (0.21 seconds)

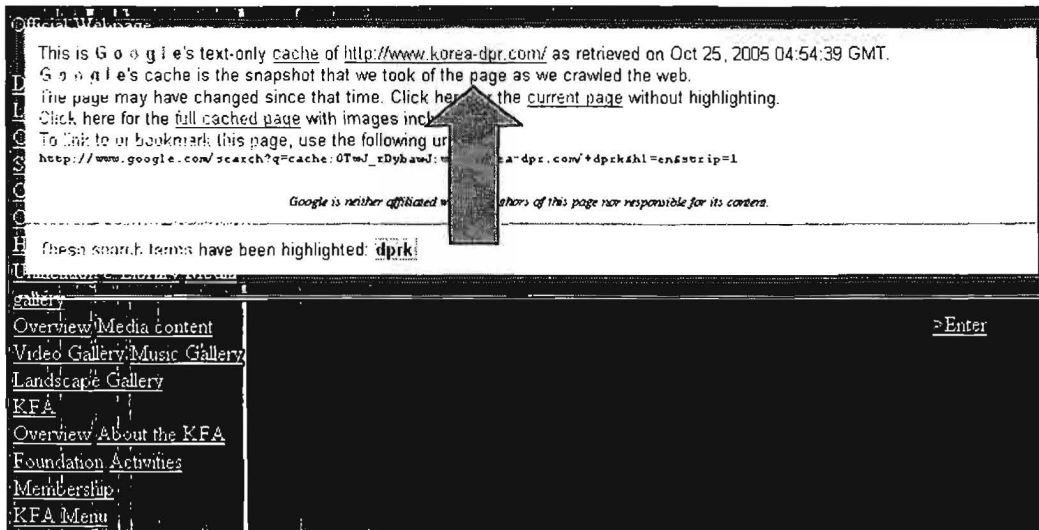
Official Homepage - Democratic Peoples Republic of Korea (DPRK ...
Official Homepage of the Democratic Peoples Republic of Korea (DPRK, North Korea)
www.korea-dpr.com/ - 10k - Cached - Similar pages

Context menu for the 'Official Homepage - Democratic Peoples Republic of Korea' link, showing options like 'Open Link in New Window', 'Copy Link Location', and 'Bookmark This Page'. The 'Copy Link Location' option is highlighted.

UNCLASSIFIED//FOR OFFICIAL USE ONLY

UNCLASSIFIED//FOR OFFICIAL USE ONLY

This will show the cached version of the page that contains only text:



Getting around the 32-word limit. For years Google had a 10-word limit for search queries, meaning that anything more than that, and Google would drop those terms from your query. However, Google expanded the number of terms searched to a 32-word limit. While the casual Google searcher will probably never notice the difference, professional researchers certainly will. There are many times when researchers need to search for long phrases (error codes, for example), exclude large numbers of terms to avoid unwanted results, run complex Google API searches, run queries of multiple sites, etc., and that darned 10-word interfered with the search. While there are a number of work-arounds all were unsatisfactory. Allowing more search terms is a big improvement, but I am sorry to report that the new 32-word limit only applies at present to main Google search, Google Images, Froogle and the Google Web API, while the 10-word limit is still in effect for Google Groups and Google News. This is especially disappointing vis-à-vis Google Groups because it has long been one of the best sources of information about complicated computer error codes and other computer arcana. Perhaps the folks at Google will see fit to expand the 32-word limit to include Google Groups.

You can, however, still *use the wildcard to trick Google Groups into searching more than 10 keywords*. Google will not count wildcards as search terms, so inserting a wildcard into a phrase will let you search for more than 10 terms. I have found this most useful when searching for a long phrase such as a computer error message, which may frequently run well over 10 words. By simply removing the "little words" such as *an*, you can easily search for the entire error message.

Here's an example of an error message containing more than 10 terms: Windows Socket Error: An Invalid Argument was supplied (10022), on API 'connect'

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

It can be written using wildcards to run the complete message as a Google Groups query:

[“Windows Socket Error: * Invalid Argument * supplied (10022), * API 'connect'”]

Undocumented Google Filetype Searches. Google can search for many more file types than those documented on the Google FAQ page. Here are some—but not all—of the file types users can search for using **filetype: plus keyword or another special syntax**, e.g., *site:* (try [filetype:cgi bin] or [filetype:js inurl:login]).

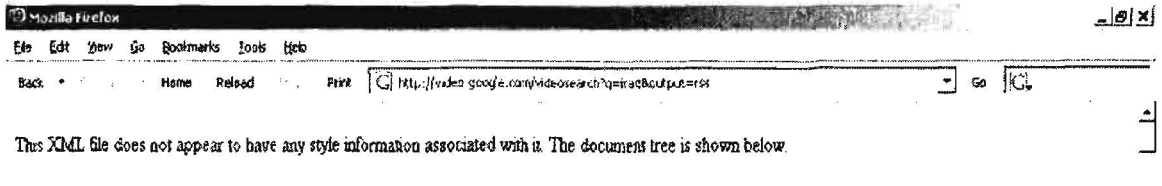
| Undocumented | Filetype | Searches |
|--------------|----------|---------------------------------|
| bak | system | backup file |
| back | system | backupfile |
| bat | system | batch file |
| bin | system | binary file |
| gz | UNIX zip | binary |
| hlp | text | help files |
| ico | graphic | icon |
| ini | system | initialization file |
| js | script | Javascript |
| log | text | log files |
| php | script | HTML |
| pls | script | PERL script |
| sql | language | database |
| tmp | system | Windows temporary file |
| uu | script | encoding |
| vbs | script | Microsoft's Visual Basic Script |

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Google Hack: Create Your Own Google Video RSS Feed. Users can easily create their own personal RSS feed of Google videos using a simple Google hack. Let's say you want a feed of Google videos about Iraq. Here's how to create it:

1. Go to Google Video<<http://video.google.com/>>and enter a search term, e.g., [iraq]
2. The result will be <http://video.google.com/videosearch?q=iraq>



The screenshot shows a Mozilla Firefox browser window with the address bar containing the URL <http://video.google.com/videosearch?q=iraq&output=rss>. Below the browser window, the XML document tree is displayed, showing the following structure:

```

- <?xml version="1.0"?>
- <channel>
  <title>Google Video - Video Search [iraq]</title>
  <link>http://video.google.com/videosearch?q=iraq</link>
  <description>Search results for query [iraq]</description>
  <generator>Google Video</generator>
- <image>
  <title>Video Search [iraq]</title>
  <link>http://video.google.com/videosearch?q=iraq</link>
- <ml>
  http://video.google.com/common/google_logo_small.jpg
  <url>
  <width>100</width>
  <height>37</height>
</image>
- <item>
  <title>Iraq Veteran Speaks Out On War Crimes</title>
  <link>
  http://video.google.com/videoplay?docid=-5587990522549547050&sourceid=searchfeed
  </link>
  <guid>
  http://video.google.com/videoplay?docid=-5587990522549547050
  </guid>
  <pubDate>Sun, 21 May 2006 13:16:21 PDT</pubDate>
  <description>
  <div style="position:relative; width:322px; height:268px; float:right; margin-left:1em; margin-bottom:5em;"><div style="position:absolute; top:0; left:0; border:1px solid #ccc; padding-top:12px; padding-left:12px; border-bottom:1px solid #ccc;">http://video.google.com/videoplay?docid=-5587990522549547050&sourceid=searchfeed#v=one&v=
  
```

3. Now add &output=rss so that your new query string <http://video.google.com/videosearch?q=iraq&output=rss>
4. Now you can add this XML output to your favorite newsreader, e.g., Bloglines to create a Google Video feed:

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Available Feeds

Google Video - Video Search [Iraq]

<http://www.google.com/feeds/videosearch?q=iraq>

Search results for query: [Iraq]

<http://www.google.com/feeds/videosearch?q=iraq&use=feed>

[Preview This Feed](#)

Options

| | |
|--|--|
| Folder | TopLevel ▾ |
| Updated Items | <input checked="" type="radio"/> Display As New <input type="radio"/> Ignore |
| Display Preferences | <input checked="" type="radio"/> Default <input type="radio"/> Complete Entries <input type="radio"/> Summaries if Available <input type="radio"/> Titles |
| Monitored By Bloglines Notifier | <input checked="" type="checkbox"/> |

Every time a new video containing the tag "iraq" is added to Google video, it will be automatically added to your blog feed. Unfortunately, this hack does not work with Google search; that would be a very nice feature. Google Blogsearch and News have RSS/Atom feed options already built into them.⁴⁷

Getting around Google's keyword restriction for *filetype*: searches. The Google *filetype*: syntax requires a keyword. To get around the requirement for a keyword, use the filetype extension as the keyword, e.g., [filetype:pdf pdf].

Google's "synonym" (related term) search. If you place a *tilde* (~) in front of a keyword, Google will search for the keyword and for its synonyms. For example, a search for ["computer ~security"] will find not only *security* but also *vulnerability*, *encryption*, *secure*, *firewall*. As you can see, this is not a search for synonyms but for related terms.

⁴⁷ Ionut Alex. Chitu, "Secret Feeds in Google Video," Google Operating System, 12 September 2006, <<http://googlesystem.blogspot.com/2006/09/secret-feeds-in-google-video.html>> (5 October 2006).

UNCLASSIFIED//FOR OFFICIAL USE ONLY

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Google Search Sinker. This query option comes from Tara Calishain's website, where she explains how it works: "Search Sinker asks for two things: a query, and then a 'sinker"—a word that you want to emphasize as much as possible in the search, e.g., [java] as the query and [coffee] as the "sinker." The script counts your query words, then fills up any remaining space in the query—remember Google's query limit is ten words—with your 'sinker.'" No API required.

http://www.researchbuzz.org/2004/01/happy_google_hacks_week_2004_2.shtml

Use Google to "Search This Site." Have you ever noticed how bad many internal site searches are? Maybe it's just me, but I can't ever seem to find what I want using a site's internal search. However, Google can perform internal site searches for you, and generally my experience is that Google will do it better. All you have to do is use the *site:* search syntax. So, if I want to search the French Ministry of Foreign Affairs' website for [afrique], all I need to do at Google is search on:

[site:www.france.diplomatie.fr afrique]

Number Range Search Option. Google has a powerful *numrange* search, which uses two number separated by two periods (dots) and no spaces. Numrange has proven invaluable to malicious types who use it to harvest credit card numbers. But it has other legitimate uses as well. It may be important to indicate what the numbers mean, e.g., weight, money, pixels, etc. Google does recognize the almighty dollar sign (but results searching on the Euro symbol are inconsistent). I'm sure you can think of many uses for the Google *numrange* search, such as searching on phone number ranges, dates, address ranges, etc. Try a search such as [amman telephone 617..680] to see how this search works. *Numrange* can be used with other Google search options, such as [site:www.jordanislamicbank.com 617..780]. Also, you don't have to use two numbers: this search finds DVD players up to \$150: ["dvd players" .. \$150].

Numrange Searches

Numrange can be used to specify that results contain numbers in a range you set. You can conduct a numrange search by specifying two numbers, separated by two periods, with no spaces. Be sure to specify a unit of measure or some other indicator of what the number range represents.

For example, you might conduct a search for *DVD player \$250..300* or *3..5 megapixel digital camera*. Numrange can be used to set a range for everything from dates (*Wilkie Mays 1950..1960*) to weights (*5000..10000 kg truck*).

DVD player \$250..350

Google Search

Numrange has other uses that I discuss under Google hacking.

Yahoo Search

In February 2004 Yahoo did what has been expected for about a year: it dropped Google. However, what had not been correctly predicted was the new search engine used by Yahoo in place of Google. Pretty much everyone expected Yahoo to go with the Inktomi search engine it purchased the previous year. Instead, **Yahoo introduced a new search engine** that “draws on” the technology of Inktomi and other search engines such as AlltheWeb and AltaVista, which are owned by Overture, a company Yahoo also acquired in 2003. After many years, **Yahoo is a legitimate search engine** and not just a directory or a pale copy of Google.

Yahoo has been and remains one of the most visited and most popular websites on the Internet. It was one of the first web portals, debuting in 1994 as a relatively simple web guide and expanding rapidly to over 25 countries in 13 languages. Along with Google, Yahoo is one of the most recognizable brand names on the Internet. At the core of Yahoo was its directory of websites, a hierarchical list of sites organized by subject. Until October 2002, that list was compiled and maintained by humans. At that point, Yahoo switched to Google search results, thereby undermining, in my opinion, its uniqueness and value.

Yahoo announced in mid-2006 what it claims is the “most significant redesign of the www.yahoo.com home page ever...[because] we’re also on a mission to empower people to find information and turn it into knowledge, play, and meaningful communication.”⁴⁸ In my opinion, this was sorely needed. What are you going to find on the new Yahoo homepage? A great deal of personalization choices: links to Yahoo email and Yahoo Messenger; local weather, traffic, events; and the most popular trends, primarily in entertainment and pop culture. The most significant addition to search is Yahoo Answers, Yahoo’s version of “ask a question.”

The current Yahoo Search Technology (YST) combines the technologies of the various Yahoo search properties—Inktomi, AlltheWeb, and AltaVista—as well as certain Google-like features to create a very powerful search tool. Among these features are:

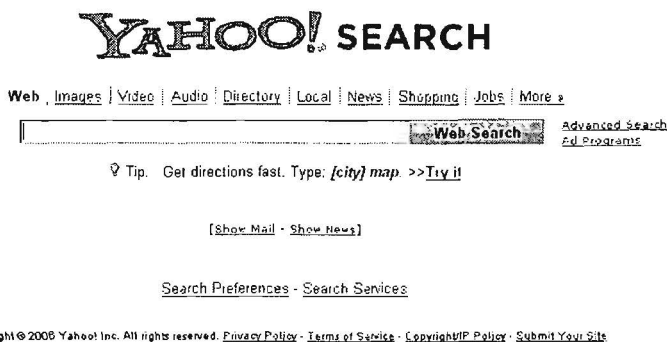
- cached versions of webpages
- automatic conversion of non-HTML filetypes to HTML is available

⁴⁸ Yahoo! Search Blog, 15 May 2006, <<http://www.ysearchblog.com/archives/000304.html>> (31 October 2006).

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

- all backlinks are shown (Google has started limiting the number of backlinks it shows)
- Yahoo refreshes its index continuously, not on a schedule (this is a good thing)
- Yahoo limits the size of indexed pages to the about the first 500KB
- Yahoo claims its database contains over 20 billion "items" (webpages, images, audio and video files); if true, it would be the largest search engine database
- default operator is AND; searchers may use OR
- search by filetype, but can only be used via the Advanced Search option (there is no special syntax)
- it appears there is no upper limit on the number of search terms (Google ignores anything beyond 32)
- an option to open link in new window (small double-window icon at end of url)

At the same time it was introducing its proprietary search engine, Yahoo streamlined its search page interface so that it resembles the Google homepage.



One feature Yahoo Search offers that Google would do well to imitate is the ability to edit the search options. Here are the options Yahoo offers; you can select the ones you want to appear on your Yahoo Search page. With Google, you get what they give you.

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Customize Yahoo! Search Homepage Tabs

Select the tabs you want to appear on your Search Homepage and click "Save".

| <input type="checkbox"/> | Tab | Description |
|-------------------------------------|----------------------|--|
| <input type="checkbox"/> | Web | The most relevant results from across the Web. This tab cannot be removed. |
| <input checked="" type="checkbox"/> | Images | Find photos and illustrations from all over the Web. |
| <input checked="" type="checkbox"/> | Directory | Search Yahoo!'s categorized guide to the web. |
| <input checked="" type="checkbox"/> | Local | Find local businesses serving your area. |
| <input checked="" type="checkbox"/> | News | Search for news stories, pictures, and audio/video. |
| <input checked="" type="checkbox"/> | Products | Find millions of product reviews and prices on the Web. |
| <input type="checkbox"/> | Maps | Find maps and directions to anywhere you'd like to go. |
| <input type="checkbox"/> | People Search | Search for people using Yahoo!'s telephone and email directories. |
| <input type="checkbox"/> | Travel | Search for airplane tickets, hotel reservations, car rentals, and more. |

Yahoo also introduced a number of international versions as well as "local" Yahoos for the US. These international versions can be very useful for locating information about a specific region or country. The pages are in the native language, so if Catalan is unfamiliar, you are probably better off sticking with the main Yahoo search page.

YAHOO! Sign In Help
New User? Sign Up

International Yahoo!

| | | |
|--|--|---|
| Americas | Asia Pacific | Europe |
| Argentina U.S. Brazil U.S. in Chinese Canada U.S. in Russian Canada en français U.S. en español Mexico | Asia Australia & NZ China Hong Kong India Indonesia Japan Korea Malaysia Philippines Singapore Taiwan Thailand | Catalan Denmark France Germany Greece Italy Norway Spain Sweden UK & Ireland |

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

Yahoo is systematically adding "local" searches outside the US. During 2006, Yahoo added Germany and UK/Ireland to its "local" page. See the [map section](#) for more information.

Yahoo International

<http://world.yahoo.com/>

Customizing Yahoo Preferences

Yahoo currently offers four basic **Search Preferences** settings:

1. SafeSearch Filter: choose among Strict, Moderate, Off. The Filter lock option requires registration.
2. Languages: search in any language or one or more of 35+ languages, including Arabic, Chinese, Greek, Hebrew, Korean, etc.
3. Display and Layout:

Open search results in a new browser window when you click on the link. Note: Yahoo also offers an "Open this result in a new window" option following the link for each result if you do not select this option.

Results per page: choose to display 10, 15, 20, 30, 40, or 100 results at a time.

Show Save and Block links (to easily save or block pages to Yahoo's My Web service directly from search results)

Show Instant Search results

4. Subscriptions: search Yahoo's premium content sources such as LexisNexis, Factiva, *The Wall Street Journal*, *Consumer Reports*. Access to these sources requires a paid subscription.
5. Multimedia Search: allows you to search and access songs already available from existing audio service provider.

The Yahoo Results Page

Once you have entered your search term(s) and clicked the Yahoo Search button, Yahoo will present you with a list of results or hits. Depending on the search you are running, you will see some or all of the following for a web search:

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Web Images Video Audio Directory Local News Shopping More >

YAHOO! SEARCH kenya

Answers Search Services Advanced Search Preferences

Search Results 1 - 10 of about 69,330,000 for kenya - 0.13 sec (All words exact)

Also try [kenya newspapers](#), [daily nation kenya](#), [kenya news](#), [kenya airways](#) More...

A

Travelzoo. Handpicked Kenya Deals
[www.travelzoo.com](#) Travelzoo handpicks outstanding Kenya vacation deals and guides you directly to the quality companies providing them.

Visiting Kenya?
[www.kayak.com](#) Find cheap flights and hotel rates for Kenya from over 100 top travel sites at Kayak.com.

Kenya Visitor Guide - World Factbook - Get Travel Tips
D **Desally says Kenya to blame for its woes** AFP via Yahoo! News - 2 hours, 21 minutes ago
FIFA suspends Kenya from international football AFP via Yahoo! News - Oct 26 4:34 AM
 More [Kenya](#) headlines - For information on [Kenya](#) visit [Shortcuts](#) - [Sign Up](#)

1 Kenya Web **B**
 Guide to all aspects of Kenyan life and links to companies in the East African region.
H Category: [Kenya](#) [Home](#) [About Us](#) [Contact Us](#)
[www.kenyawebsite.com](#) - [www.kenyawebsite.com](#)

2. http://educ.queensu.ca/~idea/downloads_manu/ProjetKenyaDenmarkFrancais.doc (MICROSOFT WORD) **E**
Résumé du Programme Kenya-Danemark programme de formation pour des Danois au Kenya et vice versa a été soulevée au ... vont au Kenya chaque année en ...
F [rdouc/queensu.ca/idea/downloads_manu/ProjetKenyaDenmarkFrancais.doc](#) - [www.kenyawebsite.com](#) **G**


Safari Magic
D Private African safaris in Kenya and Tanzania, East Africa. Travel...
[www.safarimagic.com](#)

Kenya Tour with G. A. P Adventures
 G. A. P Adventures invites you to truly experience Kenya with one...
[www.gapadventures.com](#)

Kenya Hotel
 Book your hotel in Kenya online with up to 70% off regular rates.
[www.justbookmya.com](#)

Feed a Child in Kenya - Help Mary
 Feed starving children in Africa. \$50

Kenya Hotel
 Find out about the Hotels in Kenya before you book. Discounted rates.
[hotelsbraxset.org](#)

- **A Also try:** queries to help you refine your search.
- **B Results:** 1-x out of total number of results.
- **C YAHOO Shortcuts** Indicated by  next to the first result, Yahoo Shortcuts provide "links to useful content from Yahoo, its partners, or across the web."
- **D Sponsor Results** are pay for placement results provided by Overture, a Yahoo company.
- **E File Type:** if the document found is not HTML, the file type will follow the link, e.g., (MICROSOFT WORD).
- **F View as HTML:** for non-HTML file types, select this link to see the page in HTML format. *Microsoft file types are potentially dangerous* to open in their native formats. Use view as HTML to open the document as a webpage instead of downloading it. This is a safe procedure. For more information, please review these instructions for [handling Microsoft files on the Internet safely](#).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

- **G More pages from this site**: select this link to search for other pages within this particular website that *include the search term(s)*.
- **H Category**: links to the appropriate category in the Yahoo Directory.
- **I Cached**: links to a copy of the page as saved by Yahoo's search engine; search terms are highlighted on the cached page.
- **J Translation**: if Yahoo offers a translation, you will see [Translate this page] following the link.

Yahoo Basic Search

<http://search.yahoo.com/>

Yahoo assumes as its default that multiple search terms are joined by the **AND** operator, so that a search on the keywords [windows explorer] will find all the webpages that contain **both** search terms.

Yahoo **will not return any results** if there is no webpage containing all the search terms. Try this query to see what I mean:

rollerskate handshake buckyball

Unlike Google, Yahoo **does not limit the number of search terms to 32 keywords**. Yahoo will try to match all the keywords you enter.

Yahoo is **not case sensitive**.

Yahoo does not have true **word stemming** or **truncation**, i.e., searching for variations of search terms. Normally, word stemming involves searching for plurals and verb conjugations such as *drink, drank, drunk*. However, Yahoo's word stemming is not consistent and somewhat confusing. For example, stemming works with some words but not with others, i.e., a search on [child] will not find *children*; a search on [drink] will find *drinks* but not *drinking* or *drunk*. Users should search on all variations of a term, including plurals, and not rely upon the automatic but inconsistent stemming feature. However, there is a Yahoo hack to disable word stemming.

Yahoo automatically **clusters search results**. If you want to **see more pages** from a specific site, simply select that link following the url of the result.

Undocumented Yahoo Search Feature:

Yahoo **permits the use of boolean operators** in simple search. The operators must be capitalized. Yahoo Search will run full nested boolean queries (those using parentheses), such as:

[cardinals AND (bird OR catholic) AND NOT (baseball OR football)]

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Yahoo recognizes *double-quotes* as enclosing a phrase.

Yahoo does not have any *stop words*, i.e., commonplace words. In fact, you can search on any single letter or number.

It is unnecessary to use the plus sign (+) with any terms because by default Yahoo searches for all keywords. However, there are many times when searchers need to exclude certain terms that are commonly associated with a keyword but irrelevant to their search. That's where the *minus sign* (-) comes in. Using the minus sign in front of a keyword ensures that Yahoo excludes that term from the search. For example, the results for the search ["pearl harbor" -movie] are very different from the results for ["pearl harbor"].

Yahoo will search for the *ampersand* [&]. Yahoo will search for [barnes&noble], but if you insert spaces—[barnes & noble]—Yahoo ignores the ampersand. Also, while Yahoo will not actually search on a plus sign, the search engine will search for [c+] and [c++], although it does not recognize the difference between one and two pluses.

Yahoo Advanced Search

Yahoo has many advanced search features that can be accessed from the Advanced Search page or, in many cases, employed in the simple search screen by using the correct syntax.

Yahoo has incorporated from AlltheWeb and AltaVista most of the *languages* in which users may search; like its predecessors, Yahoo is superb at searching for non-English and non-Latin languages. Using either the language preference settings or the advanced search page, users can select from over 35 languages and encodings in which to search and see results.

Keep in mind that if you search on a word using diacritical marks such as accents or umlauts, **Yahoo will only search for terms matching those with the diacritic.** However, if you search for the unaccented term, Yahoo will find the term with and without the diacritic. For example, a search on [façade] will not return pages containing only the term *façade*, but a search on [facade] will find both *facade* and *façade*.

Yahoo offers many "search meta words," i.e., special search terms to restrict searches and make them more effective. These special operators can be used in both simple and advanced search.

- **site/domain:** restricts results to a specific website or domain, including a specific top-level domain. May be used with or without keywords.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

Advanced Web Search > Site/Domain > Filter results from specific domains (com, gov, dell.com, etc.)

Examples of how to use the **site/domain:** command:

[site:amazon.com] finds *www.amazon.com*, *cards.amazon.com*, *www.amazon.com/dvd/*. However, it will not find *www.amazon.com.br*.

[domain:ir] finds all the pages from the Iranian (.ir) top-level domain indexed by Yahoo.

- **url:** use to find a specific document in the Yahoo index. This command is very limited in its usefulness and requires the full url (address), including the *http://* to work. This command works best using the Yahoo Advanced Search to find all the words in the url. May be used with or without additional keywords.

Advanced Web Search > Show Results > all of these words > In the URL of the page

Examples of how to use the **url:** command:

[url:http://impact.arc.nasa.gov/intro.cfm] finds this specific page as it is stored in the Yahoo index.

- **inurl:** restricts results to any page with a term in its url (address). May be used with or without additional keywords.

Advanced Web Search > Show Results > In the URL of the page

Examples of how to use the **inurl:** command:

[inurl:nasa] finds any site containing *nasa* anywhere in the url. Will find webpages at *www.nasa.gov* as well as pages at *www.beeville.net/NASA/*

[inurl:nasa.gov columbia] finds any site at *nasa.gov* that contains the keyword *columbia* anywhere on the website.

- **title or intitle:** restricts results to pages containing a specific word or phrase anywhere in the webpage's title, which usually appears in the browser's title bar and is the HTML *<title>* tag. *Title and intitle appear to work identically.* May be used with or without additional keywords.

Advanced Web Search > Show Results > In the title of the page

Examples of how to use the **title:** or **intitle:** command:

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

[title:amazon] finds all pages that include the word *amazon* in their title

[title:amazon "rain forest"] finds all pages that include the word *amazon* in their title and mention the phrase "*rain forest*" anywhere in the document (title or text or anywhere in the document)

- **link:** restricts results to pages containing links to a specific url. Yahoo has the unfortunate *requirement to enter the full url including the http://* to use the link command, which adversely affects its usefulness. May be used with or without additional keywords.

No Advanced Web Search Option

Examples of how to use the **link:** command:

[link:http://jpl.nasa.gov] finds all pages containing links to any page at *jpl.nasa.gov*

[link:http://jpl.nasa.gov asteroid] finds all pages containing links to any page on the *jpl.nasa.gov* site and the keyword *asteroid* anywhere on the page.

- **linkdomain:** the Yahoo *link* command finds every other page that links to a specific webpage, but *linkdomain* finds every page that links to an entire domain. Simply put, the *linkdomain* command should not be used with the full address while the *link* command requires the full address to work properly. May be used with or without additional keywords.

No Advanced Web Search Option

Examples of how to use the **linkdomain:** command vs. the **link:** command:

[linkdomain:amazon.com]
use to find links to an entire domain

[link:http://www.amazon.com] or [link:http://help.yahoo.com/help/us/ysearch]
use to find links to a specific webpage

Here is an interesting twist on link searching, that is, finding sites that link to a specific address. This search, which works with Yahoo and to a lesser extent Live Search, finds pages that link to a specific domain or domains but not to another specific domain or domains. An example would help. Let's say I start by finding the sites that link to the Iranian Ministry of Defense. Here is the query I would use:

[linkdomain:mod.ir]

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

This query returns 545 hits. Now, suppose I want to see which sites link to both the Iranian MOD and the Iranian Electronics Industries. I can do that easily with this query:

[linkdomain:mod.ir linkdomain:ieimil.com]

However, I see lots of sites that also link to the ever-present CIA World Factbook, which, while a wonderful resource, isn't what I want. I would really like to see the sites that link to both the Iranian MOD and IEI sites but *not* to the CIA Factbook. Can I do this? Sure:

[linkdomain:mod.ir linkdomain:ieimil.com -linkdomain:cia.gov]

While this technique has obvious applicability for search engine optimization ("who is linking to my competitors but not linking to me?"), I think it is worth knowing about because you may come up with some creative ways to use it. Just as an interesting example, try these two queries in both Live Search and Yahoo. It's interesting to see what drops from the results' list on the second query.

[linkdomain:cia.gov linkdomain:nsa.gov]

[linkdomain:cia.gov linkdomain:nsa.gov -linkdomain:fbi.gov]

I believe you will consistently find that Yahoo provides more results than Live Search for the linkdomain: searches. However, the results will vary, so it's worth using both search engines. Google does not offer a linkdomain: search, and its link: search has been hobbled.

- **search by file type:** restricts results to PDF, MS Word, XML, and other filetypes.

Yahoo does not offer a filetype: syntax. However, you can search by file type using the advanced search option:

Advanced Web Search > File Format > Only find results that are: > All formats or one format

There is also a Yahoo Hack that lets you search by file type.

- Yahoo's video search now includes an option to search for **Macromedia Flash** files. To limit your Yahoo video search to Flash files, go to Yahoo Video Search, select Advanced Search, uncheck all formats except Flash and run your query. A simpler approach is to enter the query [filetype:swf keyword]. ***This query only works in Yahoo Video Search***, not Web Search.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Yahoo Video Search

<http://video.search.yahoo.com/>

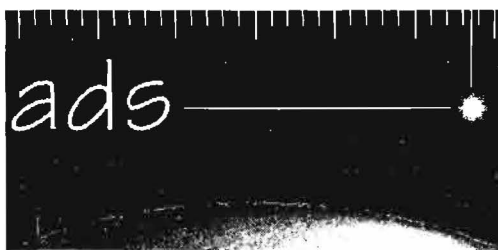
- Yahoo now links its cached copies of webpages to the **Internet Archive's Wayback Machine**. To use the link, select the Cached copy of a result, then click on "check for previous versions at the Internet Archive." That link takes you to the results of a Wayback Machine search of that precise url.

YAHOO! SEARCH[Help](#) - [Help for Webmasters](#)[← back to results for "nasa astrophysics"](#)Below is a cache of <http://cdsads.u-strasbg.fr/>. It's a snapshot of the page taken as our search engine crawled the Web.We've highlighted the words: **nasa astrophysics**The web site itself may have changed. You can check the [current page](#) (without highlighting) or check for previous versions at [the Internet Archive](#).

Yahoo is not affiliated with the authors of this page or responsible for its content.



The NASA Astrophysics Data System



The Digital Library for Physics, Astrophysics, and Instrumentation

Using Yahoo's [More from this site](#) option, you can select other specific pages to view using the Wayback Machine's access to the huge Internet Archive database.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//FOR OFFICIAL USE ONLY



Enter Web Address: [Adv. Search](#) [Compare Archive Pages](#)

Searched for <http://cdsads.u-strasbg.fr>:

89 Results

Note some duplicates are not shown. [See all](#).
* denotes when site was updated.

Search Results for Jan 01, 1996 - Oct 17, 2005

| 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2 |
|------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|----------|---|
| 0 pages | 2 pages | 4 pages | 9 pages | 5 pages | 9 pages | 10 pages | 14 pages | 20 pages | p |
| Dec 10, 1997 | * Jan 19, 1998 | * Jan 25, 1999 | * May 20, 2000 | * Feb 01, 2001 | * Jan 21, 2002 | * Jan 30, 2003 | * Feb 25, 2004 | | |
| Dec 22, 1997 | * Apr 29, 1998 | * Feb 08, 1999 | * Jun 09, 2000 | * Feb 24, 2001 | * Mar 19, 2002 | * Feb 15, 2003 | * Mar 24, 2004 | | |
| | Jul 14, 1998 | * Feb 09, 1999 | * Jul 06, 2000 | * Mar 01, 2001 | * May 31, 2002 | * Mar 23, 2003 | * Mar 25, 2004 | | |
| | Dec 12, 1998 | * Feb 21, 1999 | * Aug 17, 2000 | * Mar 08, 2001 | * Jun 04, 2002 | * Apr 24, 2003 | * May 18, 2004 | | |
| | | * Mar 02, 1999 | * Oct 17, 2000 | * Mar 31, 2001 | * Jul 21, 2002 | * Jun 02, 2003 | * May 22, 2004 | | |
| | | * Mar 02, 1999 | | * Apr 05, 2001 | * Aug 02, 2002 | * Jun 09, 2003 | * Jun 06, 2004 | | |
| | | * Apr 22, 1999 | | * May 16, 2001 | * Sep 21, 2002 | * Jun 19, 2003 | * Jun 09, 2004 | | |
| | | * Apr 29, 1999 | | * Jul 22, 2001 | * Sep 24, 2002 | * Jul 28, 2003 | * Jun 12, 2004 | | |
| | | * Oct 12, 1999 | | * Sep 25, 2001 | * Nov 19, 2002 | * Aug 06, 2003 | * Jun 19, 2004 | | |
| | | | | | * Nov 28, 2002 | * Sep 20, 2003 | * Jun 27, 2004 | | |
| | | | | | | * Sep 25, 2003 | * Jul 01, 2004 | | |
| | | | | | | * Nov 21, 2003 | * Sep 20, 2004 | | |
| | | | | | | * Dec 04, 2003 | * Oct 14, 2004 | | |
| | | | | | | * Dec 19, 2003 | * Oct 21, 2004 | | |
| | | | | | | | * Nov 12, 2004 | | |
| | | | | | | | * Nov 22, 2004 | | |
| | | | | | | | * Nov 25, 2004 | | |
| | | | | | | | * Nov 27, 2004 | | |
| | | | | | | | * Nov 28, 2004 | | |

This is an excellent way to use the Wayback Machine to its fullest because it integrates the power and speed of a huge search engine such as Yahoo with the powerful, though sometimes lumbering and hard to use, Internet Archive query tool.

Yahoo Shortcuts

Yahoo Shortcuts are special features and syntax designed to help users find specific kinds of information faster and more easily. The Shortcuts, indicated by the



include:


Local Shortcuts, including zip and area codes, weather, traffic reports, gas prices, and

Maps: to map any US location, search on the street address, city and state or the word *map* and a location. Some international maps are now available.

News & Information links when search term matches current news stories to include video and images.

Travel Shortcuts: airport information, hotel reservations, exchange rates, and flight tracker. To find directions, terminal maps, flight delays, and weather conditions at a

UNCLASSIFIED//FOR OFFICIAL USE ONLY

US airport, enter the airport's three-letter code and the word *airport*. For example, to information about Baltimore-Washington International, enter [bwi airport]. At the top of the Yahoo results page you will see a Yahoo Shortcut link marked by the . Clicking on the Shortcut link takes you to the Yahoo BWI resource page:



YAHOO! FARECHASE [Sign In](#)
New User? [Sign Up](#)

Baltimore Washington Intl. Airport (BWI), Baltimore, MD

Phone: (410) 859 7111
Address [map](#)
BWI Terminal Building
PO Box 8766
Baltimore, MD 21240-0766
Directions: [To / From](#) Airport
[Ground Transportation](#)

Terminal Information

- [Parking](#)
- [Terminal Map](#)
- [Airlines and Terminals](#)
- [Security, Customs, Immigration and Luggage](#)

Current Airport Status

Departures:
No delays (15 mins. or less)

Arrivals:
No delays (15 mins. or less)

[Security Checkpoint Wait Times](#)

Weather
47° / 69° F
[Extended Forecast](#)
[Travel Tips](#)

Find BWI Airport Hotels: Nearby Airports.

- Washington, DC (IAD)
- Washington, DC (DCA)

Airport Directory

- Atlanta, GA (ATL)
- Boston, MA (BOS)
- Chicago, IL (MDW)
- Chicago, IL (ORD)
- Dallas, TX (DFW)
- Denver, CO (DEN)
- Detroit, MI (DTW)
- Fort Lauderdale, FL (FLL)
- Las Vegas, NV (LAS)
- Los Angeles, CA (LAX)
- Miami, FL (MIA)
- New York, NY (JFK)
- New York, NY (LGA)
- Newark, NJ (EWR)

Flights Hotels

Search by: City Airport

Airport
BWI [airport locator](#)

Check-in
11/24/05

Check-out
11/26/05

| Rooms | Adults | Children |
|-------|--------|----------|
| 1 | 1 | 0 |

What is Yahoo! FareChase?
This Yahoo! FareChase Airport Directory page provides you with information about BWI Airport, including address, map and driving directions, terminal

SPONSOR RESULTS

BWI Airport Limo Service
Star Limousine
Star Limousine offers shuttle services to Baltimore International Airport (BWI) in Maryland...
[www.starsedan.com](#)

BWI Airport Shuttle and Transportation
BWI airport shuttle, BWI door-to-door transportation. We serve BWI airport to Maryland, Vi...
[www.goairports.com](#)

BWI Airport Limousine Service
Corporate clients seeking optimal efficiency can reserve professional limo service from BW...
[www.righthandsun.com](#)

Ground Transportation for BWI
Low Internet fare for

Reference Shortcuts such as:

Dictionary Definitions: as with Google, Yahoo offers the *define* option. To use it, enter *define* then a word or phrase. Yahoo only refers to the *American Heritage Dictionary*; for definitions, so Yahoo's *define* option is more limited than Google's.

Synonym Finder: similarly, *Roget's Thesaurus* provides synonyms for Yahoo Search. They syntax is [synonym keyword].

Encyclopedia: the *Columbia Encyclopedia* supplies the facts about a topic. The syntax is: ["facts about" keyword]

Number Search: Yahoo offers many types of number searches. The numbers Yahoo will search for are:

➤ **US Patent numbers:** syntax is [patent 5521308].

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Yahoo Services and Specialty Searches

News: Yahoo News headlines are entirely generated using a computer algorithm that scours approximately 4500 worldwide news sources. Yahoo News also offers *international editions* for France, Germany, India, Italy, Spain, and several other countries. For details on Yahoo News, see the news search engine section below.

Images: the Yahoo image database contains more than a billion images and includes content from Yahoo news and movies. The advanced search options for images permit users to limit the search by size, color, and image type. Users of Yahoo Search can add the *Image* tab to the left-hand side of the main search page.

Video: Yahoo Video search is extremely powerful. I address this option below in the video search section.

Translations: Yahoo absorbed AltaVista's *Systran* translation page (Babelfish) but without the excellent virtual keyboard. Systran provides reasonably decent *machine translations* of web pages from many European and Asian languages. The translation page lets you automatically translate a search result, enter any url you like, or copy/type text directly onto the translation page. If a page appears in the results list in one of the languages Yahoo translation supports, you will see [Translate this page] after the page title.

Yahoo Babelfish

<http://babelfish.yahoo.com/>

Yahoo Language Tools

<http://tools.search.yahoo.com/language/>

Mobile SMS Search: This is a new service that is different from the SMS text messaging that has been available at Yahoo for some time, i.e., Yahoo Messenger. The new SMS at Yahoo permits queries using mobile technology. Yahoo's SMS number is 92466 (which spells Yahoo on most phones). Here is Yahoo's explanation of their service, which is presently only available to US Cingular, Sprint and Verizon subscribers:

Right now you can search for any local information by sending a query with your location or zip code like: "pizza 94025," you can get a stock quote with: "s yho," weather information: "w 94025," dictionary definitions: "d garrulous," horoscopes: "h aquarius," WiFi hotspots: "wifi 94123," and more are coming. Hence typing "w" and the ZIP code get you a short weather forecast.'

Yahoo Mobile Search

<http://mobile.yahoo.com/search>

Directory: As with Google, Yahoo's web directory uses the Open Directory Project's collection but the two versions are not identical. For example, if I search the directory for the keyword *java*, Yahoo presents multiple directory categories:

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//FOR OFFICIAL USE ONLY

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

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

RELATED DIRECTORY CATEGORIES

[Programming Languages > Java](#)
[Indonesia > Java](#)
[More ..](#)

Sponsor Results
[Hands-On Java Training Nationwide](#)
 Learn hands-on from experts in more

Then I can select an appropriate category where Yahoo Directory lets me limit my search to a specific directory category.

YAHOO! search [directory](#) [Dire](#)

Indonesia > Java
[Directory](#) > [Regional](#) > [Countries](#) > [Indonesia](#) > [Islands](#) > [Java](#)

Search the Web just this category [Advance](#)

CATEGORIES

Places in Java

- ◆ [Cities](#) (16)
- ◆ [Provinces](#) (3)

Java Categories

- ◆ [Business and Shopping](#) (453)
- ◆ [Community](#) (66)
- ◆ [Education](#) (73)
- ◆ [Entertainment and Arts](#) (30)
- ◆ [Health](#) (11)
- ◆ [News and Media](#) (7)
- ◆ [Real Estate](#) (10)
- ◆ [Recreation and Sports](#) (14)
- ◆ [Travel and Transportation](#) (59)

Yahoo Answers: the biggest change to Yahoo search during 2006 was the integration of “ask a question” or Yahoo Answers into core search. Yahoo began integrating responses from Answers into its main search results in mid-2006. Although you can read answers at the Yahoo Answers site, in order to ask and answer questions users must register with Yahoo. If a question has already appeared at the Answers site, it will now turn up in the results from the main Yahoo search. Don’t expect to get answers to esoteric or difficult questions (you are much more likely to get opinions about the best DVD player than an answer to a question about which languages are spoken in Afghanistan). But the reservoir of questions and answers will continue to grow, so perhaps we will see a richer set of answers than exist thus far.

Yahoo Answers

<http://answers.yahoo.com/>

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Yahoo Site Explorer: The Yahoo Site Explorer website is still in beta; its goal is to help users learn detailed information about a specific website:

The Yahoo search database contains detailed information about the structure of the web. In addition to the web pages themselves, the database stores information about links among pages, and uses that information (as well as additional algorithms) to gauge the popularity of a given page.

Site Explorer gives you access to this information so you can learn about a site. To explore a site, you submit a URL using a search box, just as you would for a normal web search. You can then click links on the results page to see detailed information.

The Yahoo Site Explorer will reveal all the pages in a specific domain, all the pages in any subdirectory of a domain, and all the links to a domain. The main purpose of the Site Explorer is to help webmasters improve the rankings of their sites, as evidenced by the capability for sites to submit missing urls, the fact that Site Explorer provides 50 results by default, its web services APIs, and its ability to export the data to a tab separated (TSV) file for further analysis. The initial response from the search community has been lukewarm, but I like this new tool because it simplifies learning about a site and, unlike Google, Site Explorer provides all the links to a site (which Yahoo calls "inlinks") instead of a limited subset of links.

Let's examine Site Explorer from the point of view of a researcher instead of a developer. Here's the Site Explorer result page for the url [http://www.who.int] showing all the pages in all subdomains of that website. The order is by the most visited pages at the domain according to Yahoo's records about the page:

YAHOO! SEARCH
Site Explorer BETA

http://www.who.int

[Web Services API](#) [Site Explorer Help](#)
Pages: 1 - 50 of about 238,132 - 0.96 sec.

Results

Pages: (238,132) | Inlinks: (173,647)

Show pages from: All subdomains | Only this domain

1. [World Health Organization \(WHO/OMS\)](#)
www.who.int -
2. [World Health Organization \(WHO\)](#)
www.who.int/en -
3. [WHO | Severe Acute Respiratory Syndrome \(SARS\)](#)
www.who.int/es/sars/en - 19K - -
4. [WHO | World Health Organization](#)
www.who.int/home/page - 22K - -
5. [WHO | International Travel and Health Publication](#)
www.who.int/ih/en - 20K - -
6. [WHO Statistical Information System](#)
www.who.int/stats - 2K - -
7. [OMS | Organización Mundial de la Salud](#)
www.who.int/es - 23K - -

UNCLASSIFIED//FOR OFFICIAL USE ONLY

Now look at the "inlinks" (links to the WHO website) Site Explorer found that show the link to this url (<http://www.who.int>) only:

YAHOO! SEARCH
Site Explorer beta

[Web Services API](#) [Site Explorer Help](#)
Inlinks: 1: 50 of about 173,647 (1.07%)

Results
Pages (238,134) | Inlinks (173,647)
Show Inlinks to: [URL](#) | [Entire Site](#)

1. [UNAIDS: The Joint United Nations Programme on HIV/AIDS](#)
www.unaids.org/ - 50k - [Cached](#) - [Explore URL](#)
2. [Tobacco Free Initiative Geneva Hearings](#)
www.who.org/ - 6k - [Cached](#) - [Explore URL](#)
3. [Welcome to the Official Web Site Locator for the UN System of Organizations](#)
www.unsystem.org/ - 80k - [Cached](#) - [Explore URL](#)
4. [AIDS in Africa: Three scenarios to 2025](#)
www.unaids.org/en/ - 25k - [Cached](#) - [Explore URL](#)
5. [UNESCO Water, sustainable development and conservation of freshwater resources in the world](#)
www.unesco.org/water/ - 25k - [Cached](#) - [Explore URL](#)
6. [Medical Source | Medical School | Medicine - MedicalStudent.com: A digital library of authoritative medical](#)
www.medicalstudent.com/ - 80k - [Cached](#) - [Explore URL](#)
7. [Understanding the Web - Discipline Links](#)
www.icsi.wa.edu/Others.htm - 15k - [Cached](#) - [Explore URL](#)
8. [CNN.com - Aid groups accepting donations for victims - Jan 28, 2005](#)
www.cnn.com/2004/WORLD/asiapcf/01/28/05.usma.aid.sites/ - 84k - [Cached](#) - [Explore URL](#)
9. [Pediatrics, Perinatal, Paediatrics, Paediatric, General Pediatrics, Pediatric Education, General Pediatrics.com](#)
www.generalpediatrics.com/ - 51k - [Cached](#) - [Explore URL](#)
10. [Abbreviations and Acronyms of the U.S. Government](#)
www.clib.iupui.edu/subjectareas/govdocs_abbrev.html - 45k - [Cached](#) - [Explore URL](#)

By way of comparison, the **site:** search in Google—[site:www.who.int]—returned a whopping 2.2 million pages for this domain but only about 29,000 inlinks to the url, whereas Yahoo Site Search returned nearly 174,000 inlinks. However, I find it easier to explore a domain or subdomain using Site Explorer. Note the [Explore URL](#) link for every page on the results' list. This effectively lets you "dig" deep into specific pages, directories, or subdirectories in a very orderly way. Keep in mind that you cannot use any keywords or other special syntax with the Yahoo Site Explorer whereas both the Google and Yahoo **site:** command lets users include both other special syntax and keywords, e.g., [site:www.who.int inurl:sars vaccine].

The main value of Site Explorer is the powerful **inlinks** command. Look at the big increase in the number of inlinks (917,704) when you look at inlinks to the Entire Site:

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~
YAHOO! SEARCH
 Site Explorer beta

Results

[Web Services API](#) [Site Explorer Help](#)
 Inlinks 1 - 50 of about 917,704 - 1.67 sec.

Pages (236,143) | Inlinks (917,704)

[Submit URL](#)Show inlinks to: [URL](#) | [Entire Site](#)Export results to: [TSV](#)

1. [WHO | World Health Organization](#)
www.who.int - 22k - [Cached](#) - [Explore URL](#)
2. [Nature Publishing Group - science journals, jobs, and information](#)
www.nature.com/ - 74k - [Cached](#) - [Explore URL](#)
3. [UNAIDS: The Joint United Nations Programme on HIV/AIDS](#)
www.unaids.org/ - 608k - [Cached](#) - [Explore URL](#)
4. [Welcome to the Official Web Site Locator for the UN System of Organizations](#)
www.unsystem.org/ - 80k - [Cached](#) - [Explore URL](#)
5. [OneWorld.net](#)
www.oneworld.net/ - 36k - [Cached](#) - [Explore URL](#)
6. [WHO | Severe Acute Respiratory Syndrome \(SARS\)](#)
www.who.int/bsr/sars/en - 10k - [Cached](#) - [Explore URL](#)
7. [MedicalWorldSearch](#)
www.mwsearch.com/ - 15k - [Cached](#) - [Explore URL](#)
8. [Official Google Blog](#)
googleblog.blogspot.com/ - 54k - [Cached](#) - [Explore URL](#)
9. [IARC | INTERNATIONAL AGENCY FOR RESEARCH ON CANCER](#)
www.iarc.fr - 19k - [Cached](#) - [Explore URL](#)
10. [CDC - Influenza \(Flu\)](#)
www.cdc.gov/flu - 40k - [Cached](#) - [Explore URL](#)

Remember: Google will not let you use any keywords or other syntax with its **link:** command, and it purposely limits the number of inlinks as a way of trying to control webspam. However, Yahoo's **linkdomain:** syntax will let you use keywords:, e.g., [linkdomain:www.who.int sars]. The bottom line is that the Yahoo Site Explorer does not add any genuinely new functionality for researchers, while it does offer new capabilities for developers. You can do everything (and in some cases more) with the old Yahoo search syntax. So why use Yahoo Site Explorer? Because:

- it provides an extremely orderly and easy to use way of digging deep into a site,
- it provides if not a complete then a huge set of inlinks to a specific url, and
- it ranks the pages of a site by their popularity in Yahoo's statistical records of the site.

Yahoo Site Explorer (beta)

<http://siteexplorer.search.yahoo.com/>

Yahoo Mindset: This new search tool from Yahoo Research Labs is worth a look, especially for those queries that turn up a lot of commercial/shopping hits at the top of the list when you are trying to find "academic, non-commercial, or research-

UNCLASSIFIED//~~FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//FOR OFFICIAL USE ONLY

oriented sources." Mindset uses a single slider with two options: at the far left is *Shopping* and at the far right is *Researching*. You can move the slider anywhere along the continuum to minimize commercial results and maximize research results or vice versa. Here's a good example of the two different sets of results you'll get for the query ["windows xp"]:

"Researching" Windows XP (ignore the sponsored results)

MINDSET Internet - web site search

Search: windows xp

Search Results: 1 - 10. Ordering Results 1 - 100 of about 16080367 for "windows xp" (About this page...)

shopping researching

SPONSOR RESULTS

XP Windows - Cheaper Prices Before you buy, compare prices at Calixet. We have a complete selection of computers, electronics, video games and office products from online stores.
www.calixet.com

MS Windows XP - Save 30-75% Sale, this week only, includes free shipping. Windows XP Home \$92.99, Windows XP Pro \$139.49. Save more when purchasing additional licenses.
www.thefirstwarebuy.com

Windows XP OS Software: 4AllMemory.com In addition to PC and notebook memory, 4AllMemory also carries extensive lines of CD-RW and DVD drives, internal and external HDs, Flash and digital camera memory, and Windows OS software.
www.4allmemory.com

Save Now on Windows XP Save 70% when you buy Windows XP at TheSoftwareKing.com. Includes free shipping.
www.theoftwareking.com

Compare Prices on "Windows XP" and Save At Nextag.com, the search engine for shopping, compare prices, buy, shipping, ratings and reviews from some home stores all over the Web. Believe. Put your computer at Nextag today.
www.nextag.com

Windows XP Shop eBay for everything and everything - from specialized gifts to custom clothing. It's all on eBay.
www.ebay.com

Microsoft Windows XP Training - Hands On Individual programs emphasize hands-on practice in small class settings. Windows XP classes: programming, system administration and certification track. Public classes in over 75 cities.
www.traininghut.com

Microsoft Windows XP Pro - Student Discounts for students and education. Over 50,000 products from Autodesk, Quark, Apple, Microsoft,

SPONSOR RESULTS

1. (92) **TweakGuides.com - The Windows XP Tweaking Companion (XPTC)**
The source of the best Tweak Guides on the web - Extra Performance for Free! ... The Windows XP Tweaking Companion (XPTC) is the complete Windows XP and system optimization guide ...
www.tweakguides.com/XPTC.html

2. (15) **Microsoft Windows XP**
Search Microsoft.com for: Microsoft Windows XP. Latest News. Learn About SP2. Windows XP Service Pack 2 with Advanced Security Technologies helps you protect your PC against viruses, hackers, and worms. Learn about the benefits of Windows XP SP2 here. ... Microsoft Windows XP Professional x64 Edition Solution Center. Windows XP Professional x64 Edition released on April ...
support.microsoft.com/default.aspx?acid=11%3Ben-us%3Bwinxp

3. (56) **Windows XP Shows the Direction Microsoft is Going**
... Windows XP Shows the Direction Microsoft is Going ... of the major security/ privacy/ stability concerns of Windows XP. Mr. Schneier wrote the books Applied Cryptography ...
www.hevansL.com/peace/microsoft.htm

4. (16) **Doud's Windows 95/98/Me/XP Tweaks and Tips**
Tweaks tips and fixes for Windows 95, 98, Me, XP. Visual Basic Scripts: Utility, security, repair image preview, thumbnail, control, secure, password protect. ... The exceptions are for Windows XP. Some utilities have been converted to Visual

UNCLASSIFIED//FOR OFFICIAL USE ONLY

"Shopping for" Windows XP

The screenshot shows a search results page from Yahoo! Mindset. At the top, there is a search bar with the text 'windows xp' and a 'Search' button. Below the search bar, it indicates 'Search Results: 1 - 10' and 'Ordering Results: 1 - 100 of about 16,000,387 for "windows xp"'. There are two tabs: 'shopping' (selected) and 'researching'. The results are divided into 'SPONSOR RESULTS' and 'NON-SPONSOR RESULTS'. The 'SPONSOR RESULTS' section contains several links to websites like Calibex, thesoftwareguy.com, and 4AllMemory.com, all advertising Windows XP software and hardware. The 'NON-SPONSOR RESULTS' section contains four numbered links: 1. Microsoft Windows Family, 2. Amazon.com: Software: Microsoft Windows XP Home Edition Upgrade, 3. Amazon.com: Software: Microsoft Windows XP Professional Upgrade, and 4. RJ Campbell is the Software Price Beater specialis. On the right side of the page, there are several vertical advertisements for 'Save Now on Windows XP', 'Compare Prices on Windows XP and Save', 'Windows XP', 'Microsoft Windows XP Training - Hands On', and 'Microsoft Windows XP Pro - Student'.

Yahoo Mindset

<http://mindset.research.yahoo.com/>

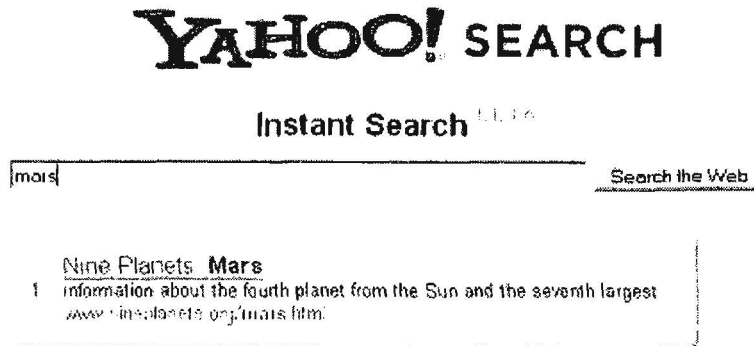
Yahoo Instant Search: This is yet another attempt by a major search engine to provide answers instead of lists of search results, Yahoo's Instant Search tries to give you a single, relevant response to a query. As their blog explains,

If search engines are so smart, why do they give you millions of results when you type in "boston weather"? Why even ten, for that matter? Why not just **one**? Or better yet, why not just tell me what the weather is?...

Of course, the answer is that often there isn't one obvious "answer" to a query and you may want a number of possible sites to explore. However, sometimes you just want a straightforward response, and that is what Instant Search is trying to achieve. Instant Search works best (in fact, only works) with simple requests, e.g., [weather london] or [bwi] or [artificial intelligence] or [convert 150 dollars to pounds]. Keep in mind, Instant Search is not trying to be another Answers.com, which "researches" answers to specific questions. Rather, Instant Search tries to "guess" what *most users* would want to know if they entered a query such as [ravens]; Instant Search guesses most users are looking for information about the Baltimore Ravens and not about birds. Sometimes the guesses are good, sometimes not.

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Instant Search employs the AJAX⁴⁹ web development technique that is starting to make a big splash. **AJAX** uses asynchronous JavaScript and XML (thus the name "AJaX") to allow interactive web browsing. You can see AJAX in action with one of the most intriguing features of Instant Search. You don't need to hit the Search button to get your answer. As you type, the answer (or best guess) will appear in a "speech bubble" below the query box:



You can also add Instant Search to your Yahoo Search page by clicking on a link on the Instant Search webpage. I actually think Instant Search is closer to Google's "I Feel Lucky" option than to Answers.com, something the Yahoo press release announcing Instant Search alludes to: "Why feel lucky when you can be right?" Instant Search doesn't just link to a webpage but actually tries to figure out what you want and give it to you (fast). It will be interesting to see if this tool catches on. I suspect that people will find the magically appearing "speech bubble" irresistible for a while, but whether or not Instant Search has staying power will depend on the quantity, quality, and reliability of the responses it provides.

Yahoo Instant Search

<http://instant.search.yahoo.com/>

Yahoo Podcasts: Yahoo's new search site is designed not only to find podcasts on topics of interest but also let users search podcasts by keywords, categories or user-generated topic tags. The new site is a variation on the traditional Yahoo directory, offering a category list by topics, lists of "what other people like" and "what we like," and a search box that lets users choose to search either series, episodes or both. A search on [spyware] returned 5 series results and 202 episode results. It is clear from the highlighted terms in the results that Yahoo's Podcast search looks not only at tags but at the content as well:

⁴⁹ "AJAX." *Computer Desktop Encyclopedia*. Computer Language Company Inc., 2005. Answers.com, <<http://www.answers.com/ajax>> (15 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

UNCLASSIFIED//FOR OFFICIAL USE ONLY

The screenshot shows the Yahoo! Podcasts search results for the keyword 'spyware'. The page is titled 'YAHOO! PODCASTS' and has a search bar containing 'spyware'. Below the search bar, there are navigation links for 'Home', 'Getting Started', 'Publish a Podcast', and 'Help'. The main content area is divided into two sections: 'Search Results' and 'Social Series results'.

Search Results
 Series containing "spyware" (3) Results: 1 - 3 of 5 See all Series results

- 1. Free Utilities and Software Announcements**
 Get Rid Of That Expensive Software Once And For All! **★★★★★ (1)**
 Listen | Subscribe | 11 Subscribers | 6 Episodes | Published: Tue, 19 Oct 2005 09:12:10 GMT
- 2. Security Podcast**
 ...vulnerabilities, Microsoft, Apple, and all kinds of other cool stuff that we think about. This week we are talking about Adware and Spyware, and a ton of vulnerabilities. **★★★★★ (3)**
 Listen | Subscribe | 3 Subscribers | 1 Episodes | Published: Sat 04/08/05
- 3. Network Corrupting Services**
 ...covered to keep your network safe and running smoothly. Spam, Internet Management, Vulnerability Assessment, Security Profiling, Spyware, Virus, Trojan Horse, Malware, MMC, Patch and Workstation Management are all critical issues that are addressed on this podcast. **★★★★★ (1)**
 Listen | Subscribe | 1 Subscriber | 1 Episodes | Published: Thu, 16 Sep 2005 10:29:00 GMT

Social Series results
 Episodes containing "spyware" (22) Results: 1 - 2 of 202,500 all-episode results

- 1. Security Now Episode 7: Spyware** **★★★★★ (4)**
 Series: Security Now
 It had to happen eventually, the man who coined the term **Spyware**, and wrote the first spyware removal program, addresses it. This is the ultimate spyware primer, what it is, how it works, and...
 Listen | Download | Duration: 8:00. Published: Fri, 30 Sep 2005 14:45:26 GMT
- 2. SSR #34 - Kispy Spyware, Miranda, OllieZero iPod, Chris Palmer Rocks, Nano, Waas, Security 360, Acrylic, Vista, 64 vs dual 32, Security 360, Acrylic, Vista, 64 vs dual 32, Super TV, Tivo DRM, New Mics, Bonus Elevator, Nintendo Controller, 360 in Japan, 360 Hacking, OE Mitsu, E-Spy, PeerFit** **★★★★★ (3)**
 Series: OllieZero iPod, Chris Palmer Rocks
 Kispy Spyware, Miranda, OllieZero iPod, Chris Palmer Rocks, Nano, Waas, Security 360, Acrylic, Vista, 64 vs dual 32,

On the right side of the page, there is a vertical advertisement for 'SUPERCHARGE YOUR PODCAST EXPERIENCE' with the text 'Get the FREE Yahoo! Music Engine' and 'Get It Now!'.

Yahoo Podcasts includes a player that does not require any installation so that users can simply click on the **Listen** button to hear the podcast in MP3 format. If you prefer, Yahoo Podcasts even lets users download a copy of a podcast for later listening. All this is free and does not require registration. The site is in beta as of now.

Why is Yahoo breaking into podcasting in such a big way? Perhaps this is in response to the new study from Yahoo and Ipsos that shows that while **RSS feeds** are gaining ground among the technology elite, they have made little headway with most Internet users. At least that is what most users believe, but the reality is somewhat different. How's that? "The survey found that 31 percent of respondents used RSS, but only 4 percent were aware of it. A full 96 percent of people participating in the survey told Ipsos Insight that they did not use RSS. Their obliviousness came from the fact that most people used browser-based feed-reading software." In short, folks are using RSS and don't know it." When users go to My Yahoo or Firefox's live bookmarks, they are using RSS technology. Moreover, the study found that even the most tech-savvy users prefer to use browser or web-based RSS readers than stand-alone software. That's easy to understand because people are so deluged with technology that unless it is easy to use or brings some indispensable new capability, people simply won't use it. The study concluded:

UNCLASSIFIED//FOR OFFICIAL USE ONLY

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~

Internet users do not understand how to use the XML button, how to actively seek out RSS feeds, or even what the term RSS means. Instead, they need a simple interface where they can choose the information and content that interests them. This is where personalized start pages and browser-based experiences can help move RSS into the mainstream.

Yahoo's new Podcasts Search tries to simplify and demystify not only podcasting but also RSS feeds so that users will feel comfortable enough to try a new technology and, Yahoo hopes, get hooked on it. I know I have pretty much given up on email newsletters in favor of RSS feeds at Bloglines because it is just so much more convenient.⁵⁰

Yahoo Podcasts Search

<http://podcasts.yahoo.com/>

RSS: Crossing into the Mainstream, by Yahoo and Ipsos Insight, October 2005

[PDF]

http://publisher.yahoo.com/rss/RSS_whitePaper1004.pdf

⁵⁰ Chris Sherman reviews eight RSS readers (for parsing primarily XML formatted news and blogs); some are integrated into a browser while others are standalone products that must be downloaded and installed. Chris Sherman, "Choosing an RSS Reader," SearchDay, 1 September 2005, <<http://searchenginewatch.com/searchday/article.php/3531486>> (14 November 2006).

~~UNCLASSIFIED//FOR OFFICIAL USE ONLY~~